Evaluating Economic Impacts of Agricultural Research:

What Have We learned?

George W. Norton
Department of Agricultural and Applied Economics
Virginia Tech
Blacksburg, VA 24060

DRAFT

Seminar paper, Department of Agricultural and Resource Economics, North Carolina State University, March 3, 2015
Abstract

The net benefits of agricultural research investments are of significant interest to research administrators and funding agencies, and the demand for benefit estimates has grown over time. The body of evidence on the benefits of agricultural research continues to mount, even if the quality of some estimates is suspect. Data availability and tools for research evaluation have improved, and it is important to match appropriate tools to the type of assessments needed. Research evaluation presents challenges that other types of impact assessments do not. The length of time needed to complete many types of research, mid-stream adjustments in agricultural research protocols for the research being evaluated, and dynamic biological environments add to the normal impact assessment challenges such as multiple goals for interventions, aggregation issues, retrospective versus prospective estimation, and identifying appropriate counterfactuals. This paper provides an overview of key issues that need addressing in agricultural research evaluations, identifies some assessment options/methods, and concludes with a few results. The paper is a brief summary of what will be a more detailed handbook of agricultural research evaluation issues and methods. However, a continuing concern is the limited progress made in some evaluation areas such as measuring impacts of systems research or social science research, and in identifying impacts on broad objectives such as food security and livelihood improvement.
Evaluating Economic Impacts of Agricultural Research: What Have We learned?

1. Twenty years since *Science under Scarcity*

The net benefits of agricultural research investments are of significant interest to research administrators and funding agencies, and the demand for agricultural research evaluation to estimate those benefits has grown over time. Twenty years ago, *Science under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting* (SUS) by Alston, Norton, and Pardey (1995) provided a summary of agricultural research evaluation issues and methods for assessing economic impacts of agricultural research programs and resulting technologies. Since then, the types of impacts being assessed have expanded and become more complex (e.g., impacts on livelihoods), and the methods used for impact assessment have been refined. Issues identified in SUS remain important, new ones have emerged, and research impact assessment has become a growth area of research. The body of evidence on the benefits of agricultural research continues to mount, even if the quality of agricultural research evaluations is a bit uneven. Data availability and tools for research evaluation have improved, but some studies struggle to match the appropriate method to the type of assessment needed and others cut corners with data or assumptions. Such problems are not unique to agricultural research evaluation, but research evaluation presents challenges that other types of impact assessments do not. The length of time needed to complete many types of research, mid-stream adjustments in agricultural research protocols for the research being evaluated, and dynamic biological environments add to the normal impact assessment challenges such as multiple goals for interventions, aggregation issues, retrospective versus prospective estimation, and identifying appropriate counterfactuals. This paper provides an overview of key issues that need addressing in agricultural research evaluations, identifies assessment options/methods, and concludes with a sampling of some results.

2. Issues

As public sector budgets tighten, many governments and donors demand increased accountability and greater care in measuring impacts of research interventions against relevant counterfactuals. Impact analyses must confront the simultaneous changing of many factors, non-random roll-out of technologies, and non-random selection of technologies by individuals. Assessments of agricultural research interventions are undertaken in a dynamic environment in which structural change occurs rapidly across economic sectors, demographic and income changes affect demand for food and plant based energy, and climate change creates increased uncertainty. To complicate the analysis further, impact assessments are needed for R&D investments for which effects have already been realized, have only partially been realized, or may occur completely in the future.

Agricultural research investments can result in distinct or multidimensional impacts on (1) farmer income and risk, (2) poverty rates, (3) the environment, (4) health and nutrition, and (5) differential benefits by gender. Sometimes these impacts are phrased in more nebulous terms such as food security or livelihood improvement. Impacts occur at the field, household, regional, national, and International levels. Evaluations are conducted at the project (individual technology or policy) level or program level. Some research investments are completed and pay off quickly, while others require many years before benefits are realized and the benefits may last for decades. Some research results depreciate quickly and require maintenance research to forestall a decline in their impacts. The temporal distribution of benefits and costs requires careful assessment of their distribution over time and discounting at appropriate rates.
Agricultural research results in both technology and institutional development, and the effects of technologies and institutions spill over geographically. They may be picked up by public and private entities for distribution. They may result in general equilibrium effects on employment and incomes (de Janvry, A. and E. Sadoulet, 2002).

3. Assessment Options

Given the complexity and sequencing of research and non-research interactions, mapping out the actual or potential impact pathway of a research intervention is a logical starting point for almost any agricultural research evaluation. The heart of the impact assessment of the agricultural research investment is then to identify the appropriate counterfactual, measure the per unit effect(s) of the intervention, and add up the effect(s) over the target population and over time. First-round effects of a productivity- or efficiency-enhancing technology can often be conceptualized as a cost reduction per unit of output that, at the market level, shifts out a supply curve and results in economic surplus benefits to producers and consumers. These benefits usually occur over several years with future benefits worth less than current benefits by some discount factor. For a closed economy with a parallel supply shift that results from an improved technology, the annual change in market-level total economic surplus ($\Delta TS$) can be measured as:

$$\Delta TS = P_0 Q_0 K(1+0.5Z\eta),$$

where $P_0$ is the base price of the commodity, $Q_0$ is the base quantity, $\eta$ is the absolute value of the price elasticity of demand. $Z = K \varepsilon / (\varepsilon + \eta)$ or the proportionate price reduction in the market, where $\varepsilon$ is the elasticity of supply (Alston, Norton, and Pardey). $K_\eta$ is the proportionate reduction in cost per ton of production in time t and for use in market-level impact assessments can be calculated as:

$$K_\eta = ((E(Y)/\varepsilon) - (E(C)/(1+\varepsilon E(Y)))A_t(1-d)^t$$

where $E(Y)$ is the proportionate yield increase per hectare for technology adopters, $\varepsilon$ is the price elasticity of supply, $E(C)$ is the proportionate variable input cost change per hectare, $A$ is the proportion of the area affected by the technology, and $d$ is the depreciation rate of the technology (Alston, et al., 1995). For a small open economy, the equivalent to equation 1 is:

$$\Delta TS = P_w Q_0 K(1+0.5K \varepsilon),$$

where $P_w$ is the world price.

These and related economic surplus formulae for other market scenarios are widely used in economic evaluations of agricultural research. A crucial variable in each of them is $K$; how it is estimated can significantly affect the results.

**Estimating K** -- Approaches for estimating $K$ for specific technologies include (1) use of expert opinions of scientists and others, (2) incorporation of input and yield data from biological field experiments in budgets combined with adoption data from surveys, (3) use of survey data at the farm-household or plot level in regression-based analyses (e.g., instrumental variables, propensity score matching, double
difference), and (4) randomized controlled trials (RCTs) in which villages and farmers are randomized with treated (receives research or technology treatment) and untreated groups. Approaches for estimating K for broad research programs at an aggregate level typically include estimation of production, productivity, cost, and production functions with lagged variables for research expenditures or patents using secondary data.

**Expert Opinion** -- Many analyses use expert opinion, often combined with cost and yield data from biological field experiments and with technology adoption surveys when adoption has already occurred (e.g., Rudi et al., 2010; Alene et al., 2009). The popularity of using expert opinion is due to (1) the lower cost and time required as compared to estimating per-unit cost reductions using data from surveys and regression approaches and (2) the need to project future benefits for technologies that may not have been adopted yet. However, any approach other than an RCT will suffer from potential selection bias unless it is carefully implemented and its basic assumptions hold. For studies employing expert opinion, interviews with the most knowledgeable research experts must be combined with interviews with experts who are most familiar with the commodity in its overall farming system.

**Randomized Controlled Trials** -- RCTs are recommended for evaluating the impacts of interventions including improved agricultural technologies because they hold constant many confounding factors (Khandker, Koowal, and Samad, 2010). With RCTs, the interventions are randomized across the population of interest in an effort to guarantee the statistical independence of the intervention from observable factors related to individuals, households, communities, and broader economy, and from unobservable factors such as ambition and time and risk preferences. Partial randomization may also be suitable where treated and control samples are chosen randomly, conditional on some observable characteristics such as income levels or land-holdings. If the dissemination of the research result or technology is exogenous to the observable characteristics, estimates of impacts will still be unbiased.

Unfortunately, several issues complicate the application or desirability of RCTs for research impact assessment including (1) spillovers from the treated to the untreated group, (2) difficulty in convincing subjects to participate, and then if they do, keeping them in full compliance or in the trial at all as on-farm research trials can often take years, (3) validity of the results beyond the area where the trial takes place, (4) ethical considerations such as those associated with withholding a potentially valuable research intervention from part of the population during the trial, (5) the high cost that tends to accompany careful RCTs, and (6) the difficulty of running RCTs when multiple interventions are sequenced into the population during the assessment period. Some of these issues can be addressed by using pilot programs or phasing in an intervention, randomizing villages rather than individuals to reduce spillovers, or randomly assigning subjects who receive an announcement or incentive to encourage participation (Duflo, Glennerster, and Kremer et al, 2008). The severity of the problems differs by type of intervention.

RCTs are most suited for assessing a single research output or technology that can be completed within a short number of years. Multiple overlapping research interventions complicate such assessments. RCTs can be used for replicated experiments in farmer fields with the farmers randomized (to assess yield and cost effects of the technology treatments). Currently, biological scientists conduct experiments
in farmers’ fields in which treatments are randomized in the fields of farmers who are willing cooperators and located in areas that represent the diversity of soil types and climates found in the target region. The treatments, except the typical “farmer practice,” are defined by the scientists in these experiments. Current farmer practice may differ little among farmers, although the degree of difference is an empirical question. If farmer practice varies significantly, it is difficult to extrapolate beyond the local area of the trial, calling the “external validity” of the results into question. An RCT could be used to test whether (1) the new technology is more profitable than the old one and (2) the differences in profitability are significant between willing cooperators and randomly selected ones. One difficulty with using randomly selected farmers is that the less willing collaborators are more likely to drop out of the experiment over the multiple years it takes for technology trials. It would be useful to know how much bias is being imparted by the non-random selection of farmers. If it is small, the added cost of randomizing farmers in addition to fields and plots may not be worth the costs. In cases in which technologies are being sequenced and scientists are juggling experiments over multiple technologies and crops, they may be quite hesitant to add additional trials, but they might be willing to randomize farmers if it does not add too much to their monitoring costs.

One use of RCTs for research evaluation is when a technology or technology package has already been developed and will be tested in a new location. An RCT must be set up before any intervention occurs in a location, and because of the potential for spillovers across farmers, the initial randomization is often at the village level. An RCT can also be useful for assessing the most cost-effective means for scaling up adoption of technologies.

Unfortunately, research evaluations often are conducted after the on-farm trials have been set up by scientists and perhaps have already been completed and partially adopted. In those cases, the primary options are quasi-experimental approaches such as instrumental variables (IV), propensity score matching (PSM), and double difference (DD).

**Instrumental Variables** -- The most common statistical method used in research evaluations to account for selectivity in adoption and/or in the location where technologies are promoted is instrumental variables (Larochelle et al., 2013; Yarobe, Rejesus, and Hammig, 2011). The method involves finding a variable (instrument) that is correlated with technology adoption but not with unobserved characteristics that affect research outcomes such as higher yield, lower input costs, or profits. Valid instruments can be difficult to find. For example, a variable such as distance to a paved road or distance to market might work as an instrument in a model where the outcome is yield (because it affects adoption but not yield), but not for one where the outcome is input costs.\(^1\) Availability of improved seeds might work for both situations as it affects adoption but only affects yield if adopted.

An important limitation of IV approaches is that they can only be used to identify an average treatment effect, but this effect might not be interesting. When treatment effects vary substantially over a

---

\(^1\) Regression discontinuity is sometimes used for impact assessment in which a program is targeted at a certain group, and households that fall just outside the cutoff for program eligibility can be compared with ones that just fall within the cutoff, essentially creating an instrumental variable based on eligibility. However, eligibility guidelines are seldom considered for agricultural technology recipients.
heterogeneous population, alternative techniques are needed and the econometrics of such analyses are still in development. This is one reason why some analysts favor PSM as heterogeneous effects can be more easily accommodated.

Propensity Score Matching -- The second most popular statistical approach to account for selectivity in evaluation of agricultural technologies is PSM which constructs a comparison group based on a model of the probability of adopting the technology, using observed characteristics. Adopters are matched with non-adopters based on this probability or “propensity score” (Gotland et al., 2004; Kumar and Quisumbing, 2010) A baseline survey can be helpful to provide the data for the matching before the technology intervention. The problem with PSM method is that unobserved characteristics may drive the decision to participate in technology adoption. The K factor is derived as the mean difference in outcomes between the two groups. To be valid, there must be few if any unobservable factors affecting adoption and a significant proportion adopters and non-adopters with overlapping propensity scores. Unfortunately, there may be uncertainty over the presence of unobservable factors such as targeting of the technology or differences in innate ability or entrepreneurial spirit of farmers. The fear is that evaluators may jump too quickly to the assumption that there are few important unobserved factors because of the difficulty of testing for those factors.

Double Difference – If there is a strong concern that unobserved factors are important but a belief that the factors are constant over time, a DD approach can be applied if a baseline survey and a post adoption survey are both possible (Walker and Kshirsagar, 1985). DD assumes this unobserved heterogeneity is time invariant, so bias is canceled out through differencing. The main drawback of the model is the assumption of time-invariant selection bias. Factors such as changes in pest pressure or sub-soil moisture can differ can cause heterogeneity that differs over time. In addition, before and after data are needed with a double difference approach, which eliminates going into a site and applying it if earlier surveys were not completed. However, if results of a baseline survey are not be available, a triple-difference method is another possible method in which a separate controlled experiment is run after adoption has occurred but in a similar yet untreated area.

Partial Adoption – Unlike many other types of development or health interventions, agricultural technologies are often partially adopted by individual households. For example, new bean or cassava varieties seldom all existing varieties on a specific farm or even individual plot. Using plot-level rather than farm-level data reduces the problem but does not eliminate it (Larochelle et al., 2103), and thus decisions must be made on what constitutes adoption for the impact analysis. In some cases, it may be possible to exploit plot-level differences in adoption on individual farms to identify the counterfactual.

Adoption of systems such as integrated pest management usually involves partial adoption of a set of practices and assumptions are needed on levels (high, medium, low) adoption (e.g., Ricker-Gilbert et al, 2008). Categorical assumptions may perform better than scales that are based on weighted individual practices, due to difficulty in assigning weights. However, it may also be useful to empirically examine why certain individual practice are adopted and others are not.
**Attribution to multiple sources of support** – Research impact assessments must consider the cost of the agricultural research being evaluated and attribution of estimated benefits to various cost centers. For example, a typical agricultural research program is supported by two or more sources and the aggregate benefits are a return to all costs. However, in many cases certain sources of support can be considered fixed costs incurred by other entities and the returns to a specific funding source need not consider the other sources if the benefits result from their marginal investment (Norton, Ganoza, and Pomareda, 1987). If estimated returns to all sources are desired, the analysis is relatively straightforward, but if apportioning those benefits to each source is desired, arbitrary decisions are needed, such as apportioning benefits in proportion to funding support as to avoid addressing the difficult issue of complementary or substitute sources of support. Spillovers of research benefits across geographic areas also create attribution problems, especially for models that use secondary data obtained for political regions. Research knowledge created within one geopolitical area can impact technologies in other areas. For example, Huffman and Evenson (1993) found that up to 45 percent of benefits from research conducted in U.S. agricultural experiment stations resulted from interstate spillovers. Models have used approaches such as geographic area proximity (Huffman and Evenson, 1993), product mix similarity (Alston et al, 2010), or agro-climatic factors important for key crop of interest (Norton, 1981).

**Multiple Objectives** – Usually, those who fund research interventions are interested in one or two key impacts. Productivity or income is often one of them, but evidence on many other types of impacts are requested as well, including poverty, environmental, health, nutrition, gender, and risk. The non-income impacts may be measured in physical or biological terms but may have economic values attached to them when appropriate. Sometimes the tradeoffs across objectives can be meaningfully calculated between types of impacts one of which is income (Antle, 2011). Tradeoff analysis is generally preferable to scoring approaches for exploring impacts of research on multiple objectives. If weighting schemes are employed, it can be preferable to base the weights on economic values derived from preferences of economic agents. Otherwise the weights may be completely arbitrary or nonsensical.

**Poverty** – Donors are often interested in impacts of a research intervention on poverty. Various methods have been used to measure changes in poverty. The Foster-Greer-Thorbecke (FGT) (1984) approach has been commonly applied because effects of an intervention can be added across population sub-groups and it can easily provide a headcount estimate of poverty as well as indicate changes in the depth and severity of poverty.

\[
P_\alpha = \frac{1}{n} \sum_{i=1}^{q} \left[ \frac{z - y_i}{z} \right]^\alpha
\]

where \(n\) = total number of households; \(q\) = number of poor households; \(y_i\) = income or expenditure of the \(i\)th poor household; \(z\) = poverty line and \(\alpha\) = poverty proportion, depth, severity (\(\alpha = 0\), \(\alpha = 1\), \(\alpha = 2\))

For example, income gains can be estimated from the new technology, and then who is adopting the technology can be assessed and the poverty rate calculated with and without the technology. FGT is complementary with RCTs, IVs, expert opinions, projections of changes in economic surplus, and other impact assessment tools. For example, Larochelle et al (2013) used an IV model to estimate income changes for farm-households that resulted from adoption of improved bean varieties in Uganda and
Rwanda. They estimated the income that adopters would have had had they not adopted the improved varieties. The counterfactual income was computed by subtracting from the observed income (obtained from a household consumption survey) the additional farm profit resulting from adoption.\(^2\) FGT poverty indices were computed for both the observed and counterfactual income and their differences were the poverty impacts. In another example, Moyo et al (2007) calculated economic surplus changes resulting from virus-resistant groundnut varieties, disaggregating the income and poverty rate changes to adopters who were also groundnut consumers, adopters who were not groundnut consumers, and consumers who were not groundnut producers.

**Environmental benefits** – Environmental impacts can be assessed with multiple methods. For example, RCT or IV analysis of farm-survey data can be used to assess the impacts of a pesticide reducing technology on pesticide use. Impacts of a change in pesticide use could be measured in physical terms alone or they could subsequently be valued using market or non-market valuation methods. Pesticides can affect many environmental categories (water, biodiversity, mammals, birds, fish, and so forth) in the short and the long term. Therefore, assessing these impacts would require projections on specific physical changes resulting from the intervention that altered pesticide use. If those changes are significant, other tools then could be applied to value them. Or, without valuation, the decision maker could be presented trade-offs in terms of physical changes.

If an economic value were to be placed on the projected pesticide reduction, values would need to be obtained on willingness-to-pay for the risk reductions to various categories of the environment. Contingent valuation could be used to collect data on people’s stated willingness to pay (WTP) for reduced pesticide risk (Mullen, Norton, and Reaves, 1997; Swinton, Owens, and van Ravensway, 1999; and Cuyno et al., 2001), or a choice experiment could be run. Alternatively, a benefit-transfer method could be used to value environmental benefits. Unfortunately, some studies have gravitated toward simple scoring models or weighted scales for environmental impacts of technologies; for example, the Environmental Impact Quotient (EIQ) that is commonly calculated by pest management specialists for technologies that affect pesticide use contains so many arbitrary weights it is essentially useless (Kovach et al, 1992).

**Nutrition and Health** -- Nutrition or health impacts of an agricultural research intervention can also be assessed using multiple methods. It is difficult to establish a causal link between a research intervention and specific nutrition and health outcomes due to the complexity of the impact pathway. Economic incentives also affect adoption of alternative nutritional interventions. The approach used in impact assessment could combine an RCT or IV study to establish the change in nutrient consumption, and then an approach such as calculation of Disability-adjusted life-years (DALYs) used to assess the implied health effects. The DALY approach, first described by Murray and Lopez (1996) as a means of capturing health effects in a single index, combines the estimated number of years of life lost and the estimated number of years lived with temporary or permanent disability due to a given health problem. DALYs are calculated with and without the intervention, and can be useful for evaluating micro-nutrient

\(^2\) When aggregated over all households, the process produced an actual and counterfactual distribution of well-being for bean-producing households.
interventions. DALYs can be given an economic value as well and included in an economic surplus and benefit cost analysis. Zimmerman and Qaim (2004) calculated the economic value of DALYs saved in the context of rice biofortification in the Philippines. Subsequent studies have applied the method to vitamin A, iron, and zinc for varietal improvement of several other crops and countries (e.g., Manyong et al., 2004; Stein et al., 2005; Nguema et al., 2011). The DALY formula is also a type of scoring system, but the weights have at least been established through extensive debate by nutritional and medical experts.

For interventions that affect macro-nutrient consumption, an alternative approach is to take the results of analyses of production and income changes due to the research intervention (using an RCT, other on-farm trial data, and secondary data) and combine those changes with a food demand system to project changes in food consumption (by quintile or other disaggregation depending on the degree of heterogeneity of consumption changes in the population). Those changes are then combined with information on nutrient contents of foods to get an aggregate picture of nutritional change (Pinstrup-Anderson, 1976; Mutuc, 2003). The results too could be used to calculate changes in DALYs.

**Gender Impacts** – Agricultural interventions can impact women in various ways, including their power, income, time allocation, and consumption among others. If RCT, ES, and other methods are used to measure total income benefits from an intervention, additional information would be needed on the shares of female production and consumption of the commodity in question, and on adoption of the resulting technology to sub-divide benefits by gender. Application of rigorous methods for assessing gender impacts of agricultural research interventions has been limited to date, with many gender studies relying on qualitative assessments or on arbitrarily weighted indices of gender impacts as opposed to careful measurement. Progress has been made on household interview techniques to ensure that views of women farmers are adequately represented in impact assessments (Doss, 2013). One of the reasons for the limited empirical progress has been the added complexity associated with farm-household analyses that address intra-household resource allocation, consumption, and power.

**Risk** – Agricultural interventions can benefit producers and consumers due to reductions in yield variance from drought and other factors. Kostandini et al (2011) and others have demonstrated relatively straightforward methods for valuing risk reduction due to agricultural technologies. For example, assuming their mean income does not change after the technology but the yield variance does, the monetary value of the benefits to farmers relative to their mean income is related to the difference in the coefficient of variation squared for the income distribution before the new technology and the coefficient of variation squared for the income distribution after the new technology:

\[
B/Y_0 = 0.5R (Y_0) \left( \sigma^2_{Y_0} - \sigma^2_{Y_1} \right)
\]

where B is the money value of the reduction in income variation, R is the coefficient of relative risk aversion, Y0 is the mean of the income distribution before the technology and Y1 is the mean after the new technology, and \( \sigma^2_{Y_0} \) is the coefficient of variation squared for the income distribution before the new technology and \( \sigma^2_{Y_1} \) is the coefficient of variation squared for the income distribution after the new technology. As more and more technologies are aimed at reducing risks from climate change, these types of benefits become increasingly important to measure.
Other objectives – Donor agencies increasing ask for measured impacts on objective such as “improved livelihoods” and “food security.” The difficulty posed for impact assessment is that the terms are vague (sometimes purposely so) and must be defined more explicitly before impacts can be measured. In some cases, it is possible measure an impact such as poverty reduction of reduced income risk as proxies for achieving those objectives.

Aggregation – Assessment of aggregate benefits of agricultural research involves several measurements or projections, including spatial and temporal diffusion and depreciation of the technology, market domains affected, and probabilities of research success. For ex ante analysis, projections may be made for specific technologies or projects (which later are aggregated up into potential research programs), or they are made for aggregate research programs. Decisions must be made about whose expert opinions are valid for the projections and how complementarities and substitutability of projects are addressed if the projects are aggregated into programs.

Aggregation and Probabilities of Success -- Agricultural research programs consist of combinations of projects. Ex ante evaluations often must assess the expected values of these combinations, with some projects unrelated to each other, others substitutes (parallel), others complements (serial), and others partial complements or substitutes. In ex ante evaluation of research projects, it is not known in advance whether any of the projects will succeed scientifically or commercially. Therefore probabilities of success must be attached to per-unit cost changes for individual projects or for the whole program.

To examine what would happen without the agricultural research program, it can be useful to disaggregate the program into its component projects because: (a) scientists projecting results of a program may be more familiar with some components than others and so have difficulty answering questions about the research program as a whole, (b) research and adoption lags may vary by technology, and (c) decisions may need to be made about relative resource allocations to the various projects (Alston, Norton, and Pardey, 1995). However, evaluating a program at the aggregate level is often required because: (a) the joint impact of a set of program components may be greater or less than the sum of the contributions of the components, and (b) the number of individual projects can be very large compared to the number of programs, and key resource allocation decisions for agricultural research may be made at the program level.

Because sometimes would seem useful to evaluate individual research projects for a potential program portfolio and then to combine these evaluations into overall program priorities, it can be instructive to examine the implications of the alternative relationships that individual projects may have to each other, as these relationships significantly influence their individual benefits and the expected value of the overall program. Research projects can be categorized into four groups: (1) unrelated (non-interacting) projects, (2) perfect complements (serial projects), (3) perfect substitutes (parallel projects), and (4) related but neither perfect complements nor perfect substitutes. Assume there are two projects: a and b. Let B(1,1) denote benefits when both projects succeed, B(1,0) denote benefits when project a succeeds and project b does not, B(0,1) denote benefits when project a fails and project b succeeds, B(0,0) denote benefits when neither project a nor project b succeeds.
With serial projects, all projects must succeed or there are no benefits: \( B(1,1) = B, B(1,0) = 0, B(0,1) = 0, \) and \( B(0,0) = 0. \) With parallel projects, if one or more succeeds, all benefits are captured: \( B(1,1) = B, B(1,0) = B, B(0,1) = B, \) and \( B(0,0) = 0. \) With unrelated projects, benefits are additive over projects: \( B(1,1) = A + B, B(1,0) = A, B(0,1) = B, \) and \( B(0,0) = 0. \) If we want to measure the expected value of a combination of projects, can we measure the values of individual projects, multiply by probabilities of success and add up? No unless projects are unrelated. For serial projects the benefits of the group would be overestimated, for parallel projects underestimated, and for related but partially serial or parallel, it is unclear. Consider two projects that can either succeed (1) or fail (0). Assume that the probabilities of success or failure are independent. The probability of success of project a is \( Pa \) and for project b it is \( Pb. \) Program benefits depend on the combination or failure of individual projects. The expected value (EV) of the program is:

\[
(6) \quad EV = Pa Pb B(1,1) + Pa (1-Pb) B(1,0) + (1-Pa) Pb B(0,1) + (1-Pa)(1-Pb) B(0,0).
\]

For serial projects, \( EV = Pa Pb B \) and is easy to calculate based on the entire program succeeding or failing. For parallel projects, \( EV = (1 - (1-Pa)(1-Pb)) \) or \( B((Pa + Pb) - Pa Pb), \) which is also easy to calculate based on the entire program succeeding or failing. For unrelated projects, \( EV = Pa A + Pb B, \) which is easy to calculate based on individual projects succeeding or failing. In the general case, it is necessary to evaluate all \( B(, ) \) terms and assign probabilities and not do so on just one project at a time. Therefore if evaluating a research portfolio or setting research priorities, it is necessary to (1) combine projects into programs and evaluate non-interacting programs or (2) evaluate benefits of all combinations of project success and failure, multiply by probabilities, and add them up over all combinations. The former does not allow project specific analysis, but is doable if experts can provide believable program probabilities of success. The latter allows project specific analysis but requires \( S^N \) separate assessments, where \( S \) is the number of levels of success per project and \( N \) is the number of projects. \( S^N \) can get big fast. Thus one needs a simple model, not a complex one for research portfolio analysis or priority setting. To obtain probabilities of success even for a single technology, program, or small set of projects or programs is difficult because only a few people are really knowledgeable about a specific research topic and those people, if they are researchers, may be biased in their views. Therefore checks and balances must be built into elicitation of probabilities of success and ranges applied.

**Aggregation from field to farm to markets at various levels** – Aggregating benefits assessed at the plot, field, or farm level in a larger population requires a conclusion about whether the sample results are general enough to apply across a range of human and bio-physical conditions given the presence of random variation and observational errors. Drawing conclusions depends on the sample size and homogeneity of the conditions. The higher the level of aggregation, the greater the potential for market level effects from the research. Simply adding up farm level impacts could mislead due to price effects and the need to discount benefits and costs over time. If econometric estimation of research impacts begins with secondary data obtained for political regions, assumptions about potential spillovers of benefits across political boundaries must be made as well, and assumptions about the nature of the research, adoption, and depreciation lags and about the discount rate. If primary farm level survey data are used, economic surplus (ES) analysis is often used to account for the price effects when aggregating up to the market level, and assumptions are made about lags and the discount rate. In some cases, it
may make sense to use a general equilibrium (GE) approaches (see for example, Hareau et al., 2005) to address additional impacts throughout the economy, especially in factor markets.

**Accounting for research lags, depreciation, and discounting** – Many econometric studies using secondary data have regressed agricultural production or productivity against agricultural research and extension variables for purposes of using the results to estimate rates of return to research. The specification of the lagged relationship between research investments and production reflects dynamics assumed for knowledge and technology creation, diffusion, and depreciation. Various lag structures have been assumed with an eye on parsimony in parameters (due to collinearity concerns) although often with little theoretical justification presented for lag shape or length. Recent studies by Alston et al (2010) and by others have found, for theoretical and empirical reasons, that total lags of 30 years or more with a gamma distribution may be appropriate. For evaluation of specific technologies, shorter lags are usually assumed, in part due to the nature of practices evaluated, but also due to available data and to the fact that with reasonable discount rates, benefits more than 20 years out are worth little.

**Cost of conducting research evaluation** – Perhaps the biggest constraint to rigorous research impact assessment is cost. Randomized Controlled Trials in particular are expensive due to the need to initiate them at the beginning of a project and the difficulty of applying to more than a few technologies when even a modest agricultural research program typically includes a dozen a more technologies, usually with new ones added each year. As a result, even an RCT that is used to evaluate a trivially small part of a research program can cost several hundred thousand dollars and still not provide much information (e.g., Duflo et al, 2008). The popularity of using existing field trial data and expert opinion is due in no small part to the cost issue. Donors often request rigorous impact evaluations with little sense of the minimum costs required to complete them. However, with careful integration into the biological part of agricultural research programs, it is possible to reduce impact assessment costs (although some might worry about loss of objectivity in that case). An RCT on the USAID funded IPM Innovation Lab program in Bangladesh is close to completion at a cost of about $50,000 per year for three years with a panel of 800 farmers in 100 villages surveyed each year. Survey costs are dropping as electronic tablets for enumerators (with programs such as Surveybe©) become the norm, as tablets reduce time per interview and with data entry.

4. Results

A large body of evidence has demonstrated that agricultural research has generated high economic returns with countries around the world benefiting enormously from productivity growth in agriculture due to technological change. Alston et al (2000) sampled 292 studies that reported a total of 1,852 estimates of rates of return to agricultural R&D, from which they reported an overall median internal rate of return of 44 percent. Other reviews of the literature have found similar results (Evenson, 2002; Fuglie and Heisey, 2007). Estimated rates of return to agricultural research vary widely, in part because some use questionable assumptions, but also because the level of aggregation affects the estimates, with the lower the level of aggregation of research programs being evaluated, the higher variance in rates of return. Most research projects yield a zero rate of return, and the estimated benefit from an aggregate research program is the average of the benefits from many projects with zero benefits, some
with modest gains, and a few with high (and in some cases astronomical) gains. Agricultural research evaluation studies that assess benefits of specific projects or technologies seldom list the unsuccessful projects; hence, the level of aggregation is a critical factor when comparing studies on rates of return.

A good example of an agricultural research and extension evaluation study that estimated a high level of benefits for a small research project was Myrick et al (2014), which estimated the benefits of a biocontrol program for papaya mealybug in Southern India. Farmers in Tamil Nadu, India had reported that a new pest was affecting papaya (Carica papaya L.) and several other crops. Numerous applications of insecticide were made, but crop losses were severe and the pest spread rapidly. The pest was identified by researchers as the papaya mealybug (Paracoccus marginatus). The parasitoid Acerophagus papayae was imported from Puerto Rico, multiplied, and released. Excellent control of the papaya mealybug was obtained within five months of the initial release, pesticide usage was reduced substantially, and production and income were increased. The annual economic benefits for the five most important crops affected by the biocontrol program in Southern India were conservatively estimated at well over $100 million, and the net present value of benefits over five years totaled more than $500 million. The total cost of the research and dissemination program was about $500,000. These high returns were conservatively estimated because the pest continued to migrate north in India and then into Bangladesh, but the beneficial parasitoid moved with the pest and therefore there was no economic damage in the new areas. Occasionally, research evaluations are accused of cherry picking the most successful project to evaluate. However, selective evaluation of research benefits often make sense because it would be too costly to evaluate all projects in a research portfolio and often the benefits a just a few projects more than pay for the whole program and therefore even if the other projects were assumed to have a zero rate of return, the program could be justified.

The fact that a high rate of return is realized for one research project or small research program does not imply that the same return can be obtained for a larger future project or program. One reason that economists often use net present value rather than rate of return to rank investments is that it is unfair to compare rates of return for projects involving thousand dollar investments with projects involving million dollar investments. Even if those with smaller budgets have higher marginal internal rates of return, those returns may well decline if the investment were to be scaled up. A related bias in rate of return studies for agricultural research was pointed out by Alston et al. (2011) and Rao, Hurley, and Pardey (2012) in the context of calculating rates of return to agricultural research. Internal rate of return calculations assume that returns can be reinvested over time at the calculated internal rate of return. However, it might make more sense to assume that the returns can be reinvested in the future at the rate of return on alternative social investments, and that the cost of capital may vary over time. Rao, Hurley, and Pardey (2012) recalculate the rates of return to agricultural research for a large set of studies globally assuming the reinvestment rates and cost of capital rates are about 3 percent. This modified internal rate of return calculation reduced the average rate of return to agricultural research in the United States from 33% in those studies to 12 percent—still a decent return, but much lower than the previous estimate. Perhaps in part because of the potential for bias, fewer studies over time have reported internal rates of return. Many of them estimate net present value of agricultural research
investments in which a real social rate of return of 3–5 percent is typically used to discount costs and benefits.

5. Conclusion and Observation

Agricultural research evaluations have found high rates of return on investment, almost regardless of the assessment methods applied. Many research evaluation methods complementary, but the key is careful identification of the counterfactual: What would have happened without the research. Where methods they are substitutes, such as RCTs, IV estimations, double difference methods, and the use of expert opinion, a careful description of the relevant impact pathways and a close eye on matching the problem and data to the method can help guide choice of method(s). For example, an initial ex ante analysis of the potential benefits of an IPM technology might use expert opinion and economic surplus analysis. A subsequent ex post impact analysis might involve a farm-household survey with IV or PSM analysis of the impacts of that intervention or of others. Data from on-farm trials with randomized and replicated plots could also be combined with adoption data in an economic surplus analysis to assess market level impacts. An RCT might be used to assess impacts of a previously developed technology in a new area. An RCT might also be used to assess the benefits of alternative technology diffusion approaches. Gender, nutrition, environmental, and poverty analyses can be conducted using various combinations of methods mentioned above. In summary, impact assessment methods have expanded and been refined over time, demand for non-efficiency objectives has grown, but there are many pitfalls and analysts must be ready to match appropriate combinations of methods to given situations. This paper has said little about evaluating research that produces improved institutions. While there is literature on the subject, unfortunately there has been relatively little progress on that subject since a few studies that focused primarily on valuing research information using Bayes Theorem (Pardey and Smith, 2004; Schimmelpfennig and Norton, 2003). Another concern is the limited progress toward measuring impacts of systems research and in identifying research impacts on broad objectives such as food security and livelihood improvement.

References


