INDEPENDENT SCIENCE & PARTNERSHIP COUNCIL
OF THE CGIAR
STANDING PANEL ON IMPACT ASSESSMENT (SPIA)

Research design in assessing the social and environmental impacts of agricultural research

Göran Djurfeldt, Torbjörn Fagerström, Axel Fredholm

FIINAL DRAFT, dated 19 jun 09

Contents

Introduction

Two things strike the eye in the literature and debate on the impact of agricultural research. Firstly, it is not unsurprisingly dominated by economists using cost-benefit analysis. Secondly and more surprisingly, much space is devoted to the discussion of counterfactuals, i.e. by an argument of the type: "If this research had not been made, yield development (or whatever indicator) would have been less than it currently is." Counterfactuals are shunned by historians but crop up in the literature on impact assessment for a simple reason, viz. that assessment is made *ex post* and without careful benchmarking having been done *ex ante*. This paper argues that historians are right, that counterfactual arguments in principle are dubious, at least in social impact assessment and that alternative research designs, more useful for causal attribution should be aimed for. [1]

In the first section follows a review of previous CGIAR work in impact assessment and a general methodological discussion. This is followed by a review of the methodologies, first for social impact assessment followed by an examination of environmental indicators. In the concluding section we propose a selection of case studies in which it would be possible to apply the methods and indicators suggested.

The framework of CGIAR related impact assessment

From the inception the mandate of the CGIAR focused on crop genetic improvement (CGI), but it has been gradually widened to include a range of subjects and activities which extends the scope of research and impact assessment to natural resource management, policy-oriented research, and social and environmental impacts. Methodologically, the broadening of responsibility poses a number of challenges. While issues in crop genetic improvement may be evaluated with techniques that offer a high level of credibility (e.g. econometrics), social and policy-oriented research often require methods that are unreliable in terms of generalization and different from the standards of traditional science (e.g. qualitative approaches).

Whereas the CGIAR centers generally have measured poverty by using income and consumption data, current impact assessments also include the concept of sustainable livelihoods. Such assessments seek a broader and more holistic understanding of poverty, taking into account a variety of factors that influence or impinge on people's capacity to tackle vulnerability and poverty, e.g. access to land and water, social relationships, physical safety and personal strategies to achieve context-specific and individual livelihood goals. Given the character of these indicators a combination of qualitative and quantitative methods is the most appropriate. This makes it possible to identify causal relationships, general trends, adoption rates etc. while also considering the multi-faceted nature of poverty and the indirect consequences of agricultural research. Problems arise, as we will see, in trying to scale-up to higher levels of geographical aggregation.

The potential of mixed methods is demonstrated in a recent volume edited by Adato and Meinzen-Dick {, 2007 #1895}. Based on impact assessments of various research projects within the CGIAR

[1] The report is compiled by Professor Göran Djurfeldt of Lund University, Sweden, Dr. Axel Fredholm of the same University and Professor Torbjörn Fagerström of the Swedish University of Agricultural Sciences (SLU). We have received valuable inputs from the following scholars: Malin Beckman, Kevin Bishop, Per-Anders Hansson, Jan Erik Hällgren, Richard Johnsson, Erik Karltun, Henrik Kyllin, Ulf Magnusson, Gert Nyqvist and Håkan Olsson from SLU, and Agnes Andersson, Johanna Bergman-Lodin, Thomas Hickler and Magnus Jirström, Lund University.

system, it demonstrates that qualitative and quantitative methods are mutually reinforcing and that qualitative analysis may reveal factors that are difficult to discover only on the basis of statistics. It also presents studies that assess the impact of agricultural research on poverty at the national level in India and China through econometric approaches {Fan, 2007 #1902; Fan, 2007 #1901}. In line with the general framework of *ex-post* impact assessment in the CGIAR, counterfactuals are established to estimate what would have happened without research and new technology. Depending on the availability of benchmarks and previous assessments two types of counterfactuals are used. While the first draws on studies of adopters and non-adopters and examines systematic differences among them that may influence impact and performance, the second use previous surveys or panel data and establishes a longitudinal approach to the "with/without" dilemma.

The portfolio of CGIAR related research also includes assessments of natural resource management research (NRMR). The case studies in the volume edited by Waibel and Silberman {, 2007 #1904} fall into this category. While most cases analyze the impact of various technologies associated with natural resource management, e.g. zero tillage, improved tree fallows and soil conservation technology, some may also be labeled policy-oriented since they examine the impact of research on policy-making. Here, too, the methodological framework is extensive and ranges from bibliometric and webmetric analysis to ecological modeling and cost-benefit analysis. However, as the cases generally had inadequate access to baseline data when the research began, the capacity to establish appropriate "with/without" scenarios is limited. Despite the methodological richness, therefore, the impact is often assessed on the conditions of 'best' available data or indirect indicators.

A variety of methods is also applied in policy-oriented impact assessment. The influence on policy-making due to the work of IRRI and CIFOR, for instance, is assessed using a mix of qualitative and quantitative methods (Templeton and Jamora 2007; Raitzer 2007). By establishing counterfactuals and by combining key informant interviews and discourse analysis with techniques for cost-benefit analysis, these studies explore the link between CGIAR related research and policy-making in the Philippines and Indonesia. Despite the obvious problem of establishing causality between research and policy-making they arrive to some interesting conclusions. While research at IRRI contributed to new policies in pesticide use in the Philippines and savings in private health costs, CIFOR's research reduced the presumptive costs of deforestation in Indonesia due to new policies in the pulp and paper sector.

At a more general level, then, three characteristics are evident in current impact assessment. Firstly, due to the sustainable livelihoods approach and the significance of qualitative indicators in the assessment of social impacts statistical methods are often combined with 'softer' methods drawing on anthropology and qualitative sociology. This is apparent in the volume edited by Adato and Meinzen-Dick (2007) but also in the assessment of policy-oriented research. Secondly, many studies are battling against odds due to poor benchmarks and absence of baseline data. There are, however, exceptions to this problem. A good example is the assessment of high-yield maize in resettlement areas in Zimbabwe that has a good benchmark due to evaluations in the initial phase of the project (Bourdillon, Hebinck et al. 2007). This makes it possible to establish a longitudinal assessment that is more convincing in comparison with studies that lack such reference points. Thirdly, due to deficient benchmarks many studies are using counterfactuals to deal with the "with/without" dilemma. Drawing on the general methodological toolbox of the CGIAR, this approach is a guarantor for rigorous impact assessment in cases where baseline data or continuous evaluations of research projects are missing.

Below we examine these characteristics more in detail and offer recommendations on how to move forward.

Combining qualitative and quantitative methods

Although qualitative analysis provides unique and context-specific information, it falls short when it comes to generalizations. Unless data is available in other sources such methods alone cannot estimate, for example, the percentage of adopters of a specific technology. Due to the weak qualitative research infrastructure in developing countries, it may also be difficult to find local researchers with proper skills. At the same time, qualitative methods have a great advantage in being able to provide depth descriptions and to analyze in detail the causal mechanism leading from intervention to effect. Statistical analysis, on the other hand, cannot be used for in depth descriptions, but among its advantages is the important but neglected rhetorical power of numbers. In summary, therefore, the choice between qualitative and survey-based statistical approaches is distinguished by a trade-off and we generally recommend a combination of both. In the assessment of sustainable livelihoods the combination of qualitative and quantitative methods is comprehensive and carefully planned. As the volume edited by Adato and Meinzen-Dick (2007) indicates, this makes it possible to estimate the indirect consequences of agricultural research and to establish a detailed and contextual assessment while simultaneously identifying quantitative patterns of impact. Consequently, we would recommend CGIAR and its associates to continue using combined approaches in the study of research impact on livelihoods.

Baseline data

As noticed earlier, a weakness in many studies is the absence of baseline data, i.e. information on the state of affairs *ex ante*. This is evident in the cases dealing with natural resource management, which often rely on "best" available data and indirect indicators. One explanation is that impact assessments were not planned for when the projects were launched. To mitigate this dilemma, an obvious recommendation is a more careful planning for evaluation processes in the initial phase of a project. The Environmental Impact Assessment System for Responsible Rural Production in Uruguay is one example where evaluation techniques are present in the initial phase (Rodrigues and Viñas 2007). The outcome of this project is assessed through the Rural Environmental Impact Assessment System for Uruguay (EIAR-Uruguay) and creates a benchmark by analyzing adoption rates among 500 farmers over several years. A further example is the aforementioned study on high-yield maize in resettlement areas in Zimbabwe where surveys started almost immediately and were carried out continuously (Bourdillon, Hebinck et al. 2007). Such practices may increase the availability of empirically observed data and improve credibility.

Counterfactuals

In the absence of baseline data impact assessment is tricky methodologically, not only due to poor benchmarks but because it is necessary to handle *counterfactuals*, i.e. the classical question: What would have happened had Caesar not crossed the Rubicon River? Despite much brooding, philosophers have not come up with more than one answer: We don't know! Here we are not dealing with ancient history, yet given that we will never know about the counterfactuals, the problem is if we can get to know much more about the results of more recent interventions than Caesar's.

As is clear from an entry into the *Oxford Companion to Philosophy,* counterfactual arguments are usually regarded as legitimate if they refer to a Hempelian *covering law.*[2] To use the *Companion's* example, given the covering law that sugar dissolves in water, the following counterfactual is valid: If I drop this lump of sugar in the water, it will dissolve. A *ceteris paribus* clause only has to be added. This type of argument is often used by scientists and in thought experiments and is perfectly acceptable. However, things are more complicated in the social sciences, since there are few if any covering laws. Mainstream social science has long since left Hempelian theory of science behind it. To take another example, there is no covering law from which anything at all can be concluded from the following statement: If this new crop variety is released among farmers it will be adopted. If it had been possible to make any firm predictions, the present exercise in impact assessment would have been unnecessary. This is because social systems are open, and in principle always unpredictable.

An economist might protest to this argument, because in economics there are law-like statements of the Hempelian type, for example that when the demand for a commodity increases, and if the supply does not increase sufficiently to match the increase in demand, then prices will go up (adding the *ceteris paribus* that there are no monopolies etc.). The problem is that real markets, especially agricultural markets in developing countries, seldom match the *ceteris paribus*. Therefore it is wise to be sceptical about counterfactual arguments in the social sciences. We need also to realise that such arguments often function as a normative cover-up: if they are accepted they tend to legitimate certain facts, like in the famous counterfactual: If the bombs hadn't been dropped over Hiroshima and Nagasaki, the war would have continued longer with greater costs in human suffering.

We therefore contend that the famous argument about the Green Revolution that, if it had not occurred the forested area in Asia would have been much smaller, staple food production less and food problems more severe, however much we would like to believe in it, is in principle dubious. It functions as a legitimation and, predictably, the critics are not convinced by it. More solid *ex post* impact assessments of the Green Revolution are needed than this one.

Note, however, that the above example transcends the boundary between disciplines and contains statements both about agrarian ecosystems and about society. As such it is interesting, because using covering laws one can legitimately make some inferences about what would have happened to ecosystems if yields had not increased as they did. At the same time in the example inferences are made about phenomena which are inherently social, like food problems, where inferences cannot be backed by covering laws. In other words, what would have been the social and political response to a yield crisis in Asian riziculture? We will never know! To make things even more complicated, the socio-political response would have had impacts on the agrarian ecosystems, potentially overturning the inferences drawn from a counterfactual based on covering laws.

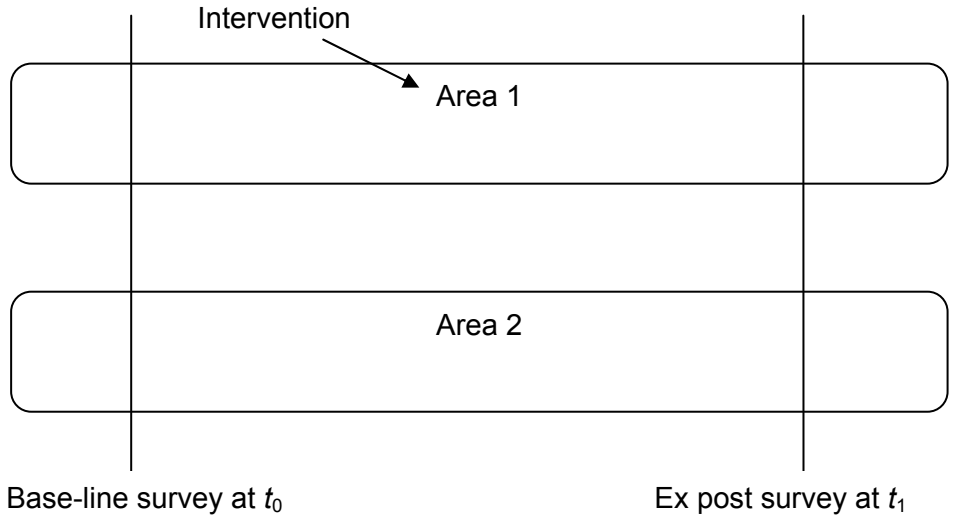
One way to get closer to an answer to the question of how to design an *ex post* impact assessment without resorting to dubious counterfactual arguments is to depart from an ideal methodology for impact assessment and then ask what happens as we remove parts of that methodology. So, for the sake of the argument, say that there is a new technology and a product of crop breeding which has been introduced in an area, for example a watershed, and add that we want to know the impact

[2] See Honderich (ed.) (1995, p. 169). The term *covering law* was coined by Carl Hempel and was influential in social science before Kuhn introduced the *paradigm* concept, varieties of which are now dominant in the social sciences (Kuhn 1962; Hempel 1965).

economically (cost-benefit analysis), socially (poverty, equality etc.) and environmentally (yields, soil quality etc). Ideally we could study that by means of a quasi-experimental design:

1. Select *two* areas, one where the intervention has been made and one control area where it has not. Selection should be based on the most similar principle, i.e. ideally the areas should have similar agro-ecology, be located in the same country etc.
2. Within the two areas make a baseline survey, giving the indicators at t_0 for the dimensions of impact we are interested in.
3. After the intervention, remake the surveys at t_1, thus getting the indicators needed for a strict comparison of before-after and for a proper causal analysis of the effects of the intervention.

The above is the only design which allows a proper separation of causes of change attributable to the



intervention and other factors which impact on the process. In this design, then, the problem of counterfactuals does not arise.

With an *ex post* design, by definition there is no baseline survey, which makes for a deviance from the ideal design. It is still possible to stick to the *most-similar cases* design and select two areas according to the principles outlined above. To the extent that the areas really are similar, effects which are present in the intervention area and absent from the control area would be attributable to the intervention. Moreover, since in a social application it is impossible to choose areas so that most variance on the indicators is *between* areas and less *within* them, the within area variance can, in a comparative statistical analysis be used to further check on the possible influence of other factors on the outcome.

Note that this design can also handle *spill-in effects*, which a pure *ex post* design without a control area cannot. In fact, this is a vexing problem in all types of project evaluation: However skilled and experienced the evaluators, spills-in may go undetected and confound what appears as a clear effect of an intervention.

Thus, a comparative case study and selection of cases by means of the *most similar* principle, makes it possible to avoid dealing with the intractable counterfactuals that arise if we choose to study only the area of intervention. Moreover, adding a pseudo-panel methodology can add further to the power of the design.

Pseudo-panels build on retrospective questions. If we are willing to regard such data as containing, with enough reliability, information about the state at t_0 or about the change between t_0 and t_1, we would get some data also for the base-line or changes that have occurred in the interim. These may be possible to complement with data from other sources, like official statistics, which would further strengthen the pool of evidence.

Retrospective questions in surveys have a bad reputation, which is not entirely deserved. They may be quite useful, given that one adapts the message to the medium, so to say. It is not much use to ask a farmer how much maize he got per hectare five years ago. However, you will get a more reliable answer if you ask whether he got more or less than last year. Precision will be higher if you tie the question to years with a symbolic significance. That is, it is better to ask what yields the farmer got in a normal year in the beginning of his farmer career compared to a normal year nowadays. Note that is no use asking about absolute amounts. The answer you get will be in the form of an ordinal variable: less, the same or more. If one wishes one can even reduce to a binary variable and get to know if yields increased or not. We are better at using that type of coarse by reliable data nowadays than we used to be, e.g. by binary logistic regression.

Our conclusion, then, is that *ex post* impact assessment need not get embroiled in arguments about counterfactuals, a peg which critics will always hang their arguments on. With these general methodological comments and recommendations, we will move over to a discussion of social and environmental indicators.

Social indicators

The following methodological discussion especially emphasizes the scale of activities, at project, programme or system level. This triad is an administrative distinction within the CG system, which upon reflection can refer to different levels of geographical aggregation. So for example, a project can have impacts on system level. Since there is no one-one relation between the administrative triad and geographical aggregation, the following discussion is specified according to the geographical dimension rather than the administrative one. The main message is that data collection problems differ considerably according to the geographical scale. Collecting data on a low level of aggregation is quite different methodologically from doing so at a higher level. In the table below we distinguish between low, medium and high levels of aggregation and list the different types of indicators accordingly. Note, however, that the table simplifies matters to a great deal, imposing a two-dimensional structure on a reality which is in fact multidimensional. The nuances will be clear from the text, while the table should be regarded as a mere summary account. We will proceed to exemplify this by discussing poverty indicators.

Table 1. Data sources for social environmental indicators[3]

			Level of geographical aggregation			
Indicator			**Low**	**Medium**	**High**	**Comments**
Poverty						
	WB definition		(X[4])	X	X	Generalisable, comparable, longitudinal
	Other operationalisations		X	-	-	Approximately comparable, can use pseudo-panel
	Alternative definitions (livelihoods)		X	-	-	Not generalisable, can use pseudo-panel
Other social indicators						
	Survey-based methodology		X	X	-	Approximately comparable, can use pseudo-panel
	Based on secondary data			X	X	Approximately comparable, can use pseudo-pane
Environmental indicators						
	Based on remote sensing:					
		Vegetative cover incl. forest	X	X	X	Approximate indicators can be developed, benchmarks
		Soil erosion	X	X	X	Approximate indicators can be developed, benchmarks
		Biodiversity I	X	X	X	Approximate indicators can be developed, benchmarks
	Based on field data					
		Biodiversity II	X	-	-	Not generalisable, no benchmarks
		Changes in water quality	X	X	X	Biological indicators increasingly used
	Model-based estimates					
		Changes in chemical pollution	X	X	X	Data only available for national levels and higher
		Leakage of nutrients	-	X	X	Data only available for national levels and higher
		Changes in greenhouse gasses	-	X	X	Data mainly available for national levels and higher
		Animal health	X	X	X	Approximate indicators can be developed, benchmarks

[3] X indicates that data are available or possible to collect.
[4] Secondary data.

Poverty indicators and income effects

The World Bank measurements of poverty have become the standard reference, both scientifically and politically, as evidenced by their being the golden standard for the Millenium Development Goals. It would have been strange if they had not been contested.[5] In recent years qualitative methods of poverty assessment have been launched as an alternative or complement.[6] In several of the articles in Adato and Meinzen-Dick (2007) referred to above, qualitative poverty assessments are used with bravado. Recently, moreover, combinations of qualitative and quantitative methods in poverty assessments have received increasing attention.[7]

A great advantage with the World Bank method is that it measures poverty levels, according to a widely accepted definition of absolute poverty. The dollar a day criterion is meant to reflect the costs to a consumer of attaining a minimum consumption of calories, in addition to a small budget for non-food expenses. Another advantage is that it can be used to study the *depth of poverty*, i.e. how far below the poverty line the poor are, on the average. In impact assessments it further enables one to compare the impact of interventions on the poor and the non-poor (given that relevant information on the interventions is available). Moreover, one can study distribution of poverty, i.e. the degree of inequality in consumer expenditure among the poor. Both these measures have implications for the aggregate increases in consumption and their distribution which are necessary for lifting a certain percentage of the poor out of poverty.

In large parts of the developing world, however, absolute poverty is increasingly being outdated thanks to economic growth. Under such circumstances it is more meaningful to measure relative than absolute poverty. Unlike the latter, which by definition is binary (poor, non-poor), relative poverty is a scale (more or less poor, more or less wealthy). Although some definitions are more current than others, this implies that levels of poverty get quite arbitrary (two dollars a day, less than half the median income etc.).

The World Bank method has been criticised for its unreliability. The method builds on lengthy interviews with consumers on their expenditure on various food items and other goods and services. Independent researchers who have tested the method are critical because questions are very difficult for respondents to give reliable answers to. Moreover, respondents, especially from non-poor households, quickly get impatient with the long interview and tend to respond less carefully after some time. In the Indian debate, concern has been expressed about the non-inclusion of consumption of non-marketed goods, both from common property resources and from official institutions. The latter would affect the validity of assessments, especially over the longer run when the relative importance of commoditised and non-commoditised consumption is likely to change.

In defence of the World Bank approach, it may very well be that the unreliability pointed to in the last instance reduces to *white noise*, affecting the precision of poverty estimates, but not necessarily the general tendencies in the results.

[5] See the early attack by Bhalla (2002). When the latest estimates were recently presented by Ravallion in the Indian journal *Economic and Political Weekly*, an interesting debate followed (see Himanshu 2008; Ravallion 2008a; Reddy 2008). See also the reply by Ravallion (2008b).

[6] See Krishna's work for an early example (2004).

[7] See e.g. Kanbur and Schaffer (2007).

From the perspective of the present paper though, the main point to be made about the World Bank method is that it is not always feasible to implement in impact assessments of the type we are presently discussing. World Bank estimates can often be used, but then more like other bits of official statistics. Information from household consumer expenditure surveys are available for most developing countries and usually in several rounds, making it possible to estimate the trends in poverty, as well as poverty rates, not only in the survey years, but by interpolation also for other years. However, one has to be content with data from the lowest level of aggregation, usually district or sub-district data. Thus, data would not normally be available for villages (at best for sample villages) or for individual households. Instead data would typically be tied to the district or sub-district level. This need not necessarily be a drawback and, if so, it may be compensated by the possibility to get trend data. To conclude, World Bank poverty data can play the same role in an impact assessment as official statistics in the more current sense of the term: It is not available at the lowest level of aggregation, but has the advantage of being generalisable, comparable, longitudinal and reliable enough. Moreover, it can provide a benchmark in poverty assessments.

If it is deemed necessary to get data for poverty on household or individual levels, this data has to be collected for the purpose. Since it is time-consuming, it is usually not feasible to collect household expenditure data. One would have to resort to income rather than expenditure data, which generally is considered to be less reliable, especially when it comes to assessments of inequality. This again is because wealthier households often have a higher propensity to conceal or distort their incomes than poorer ones, which usually do not have much to conceal. Another alternative is to go for simplied consumption data, e.g. by collecting data on the consumption of selected items in a basket of commodities. Since this is not a standard methodology, this is obviously at the cost of loss in comparability.

In longitudinal studies, the bias in income data may jeopardise conclusions about long-term trends in inequality. Another drawback with such data is that it is difficult to disaggregate to the individual level why intra-household inequality is difficult to get at.

Poverty assessment based on household income data can be tied to the World Bank definition of absolute poverty. Not only levels of poverty but also depth and distribution of income among the poor can be studied. This means that results of such assessment become at least roughly comparable with WB assessments. The drawback is obviously that it is difficult to get at trends, since data at least in the first instance would be cross-sectional. It may be added though that this can be partly compensated for by pseudo-panel data. While it is hardly meaningful to establish historical income levels for individual households, information about trends can be quite reliable: “When you compare what you can buy with the income you have today with what you could buy with the income you had before you adopted the practice we have been discussing, are you better or worse off today?” What is lost in precision may even be gained in reliability, provided there is no systematic distortion of the past, for example in the form of romanticism about a golden age, *ex ante* as it were.

Qualitative poverty assessments are very much associated with the livelihoods approach, inspired by Robert Chambers and his group at the Institute of Development Studies, University of Sussex. It needs to be noted that the philosophy of science associated with this school is some way off from mainstream science. While the latter fundamentally aims at universal or at least provisionally generalisable knowledge, the former has more particularistic aims, at least for its knowledge production. Chambers and his group has a universal philosophy stressing the capacity of the poor and

marginalised and the importance of their empowerment as a prerequisite for their emancipation from poverty, not necessarily on an individual base, but with a very strong tinge of communitarian values to tied to it. Sustainable livelihood studies aim to understand the particular and individual characteristics of the poor and the communities they are members of. The studies already referred to in the volume edited by Adato and Meinzen-Dick bear witness to the rich descriptions that are possible to attain with these methods, but their overall dilemma of being particularistic, with findings difficult to generalise is also amply demonstrated.

It would seem an obvious advantage to combine qualitative and quantitative poverty assessments. One need not go as far as those who are discussing the Q-squared (see footnote 7 above) and combine the results of different investigations in one database. The more common and accepted strategy would generally do so that one can combine, on low levels of geographical aggregation, for example a survey of household incomes with qualitative assessments of poverty and livelihoods. The advantage is that while the results of the former would be generalisable and comparable, the latter can provide important insights into the causal mechanisms lying under the different outcomes.

Can one handle downstream effects of interventions in the same way? While for example an innovation in cropping may have positive effects among the farmers who are its primary beneficiaries, the same may or may not be true further down the causal impact pathway. Since societies are immensely complex systems, interconnected globally and constantly changing, it is potentially possible to follow the downstream impacts of an innovation through many links and to places faraway from where the intervention was first made. Faced with this complexity, the first task of the researchers is to prioritize and choose to concentrate on the most important ones.

As demonstrated by Fan et al. (2007a; 2007b), mapping complex pathways require longitudinal datasets available in countries like India and China but not in many other developing countries. Even with a more moderate ambition in this regard, however, it may be necessary to go beyond the village-level in tracing causal impacts. With reference more specifically to poverty outcomes, it may be judged important to trace effects not only locally, on farmers, labourers and the local non-farm sector, but also to follow the links outside the village onto a wider region or onto the national level and into urban areas and cities. Obviously, a micro-level, field based study combining qualitative and quantitative data collection of the type discussed above will be difficult to expand in the way necessary to follow such a wide-ranging impact pathway. If at all feasible in a single study, it would call for a combination of approaches, first of all spanning the micro and medium levels of geographical aggregations (columns 1 and 2 in Table 1). Secondly, such a study would have to combine field based data collection with secondary data from official statistics and elsewhere. If several rounds of consumer expenditure surveys are available, they could be a main source of data, combined with other sources of data and in the hands of skilled econometricians who would be able to answer for example the question of consumers benefitting from advances in crop technologies.

Changes in labour conditions

Under the heading of labour conditions we will be discussing changes in drudgery and in occupational health and safety. Searching in peer reviewed articles one finds almost nothing on drudgery in agricultural work in developing countries. The catch is somewhat better in a Google search, throwing "grey" publications from various donors and NGOs. Perhaps this is an indication of how far from the harsh realities experienced by farmers and workers that most researchers and most research still are. One is reminded of A.V. Chayanov's pioneering attempt to adapt the marginalism of Walras to the

realities of Russian farmers in the early 20th century. Instead of modelling a utility calculus by a rational economic agent, he modelled a rational peasant trying to optimize between the utility of subsistence and the drudgery of an extra input of labour (Chayanov 1966).

Given that a good qualitative or quantitative description of changes in farming systems is available, the optimal way of understanding its implications in terms of drudgery would be, preferably by participatory methods or by other means of collecting data from farmers and labourers, to get an understanding of how changes in crop technologies affect them.

Occupational health and safety may be more difficult to handle since it requires medical and other expertise which social scientists do not possess. Cooperation between social scientists, doctors and engineers may be one way to handle this. Methods would probably have to be field based, combined with secondary data on for example the use of chemicals in the farming sector.

What about benchmarks in situations like these? They are probably not there, so it would be difficult to avoid the traps in drawing longitudinal conclusions based on cross-sectional, *ex post* data. One way out is pseudo-panel data (SMPP) drawing on people's own recollections about conditions *ex ante*. Even if there are biases in recollection, like romanticism of the past, skilled field workers should be able to handle this. Moreover, to quote Thomas' theorem: *If a situation is defined as real, it is real in its consequences*. In some situations, lived experience is more important than other layers of reality, which may be the primary concern of scientists, but not necessarily of lay persons.

In this connection a few words must be said about anxiety and suicides, occasioned of course by the debate on the recent suicide epidemic among Indian cotton growers. Initially and with much beating on the drums, the epidemic was attributed to changes in crop technology, more specifically the introduction of Monsanto's Bt cotton in India. This in turn has bred what threatens to be a long-lived mythology where the emerging scientific consensus has difficulties to get heard (see for example Vaidyanathan 2006 and several papers by Herring).[8]

This controversy aside, and especially as farm incomes grow, new aspects of farmers lives are bound to come to the fore, among them socio-psychological dimensions like the anxieties induced by commercial production in volatile markets and by foreclosures of non-viable farm enterprises. In the West such problems, to the extent that they do not remain buried within farmer families, are often handled by counselling by specialist staff employed by farmers' organisations, churches or the public health system. Even then, to our knowledge there is little in terms of regular monitoring of socio-psychological health among farmers, or any other occupational group for that matter. Efforts are less on preventive measures and more on acute crisis management. Thus there is little to draw on in terms of organisational models. If the SpIA considers inclusion of these elements in impact assessment, it would be necessary to draw on academic and clinical expertise in counselling and occupational psychiatry.

Food and nutrition

When poverty rates are high, farmers make up a big proportion of the hungry and a primary goal of research based innovations should be to alleviate their hunger. Like with poverty, however, farmers may be the primary beneficiaries, but they are not alone. Increased production of food crops should have a direct and positive effect on farmers' capacity for self-provisioning, but also on their marketed

[8] http://falcon.arts.cornell.edu/Govt/faculty/Herring.html.

output. The ideal combination is when output rises, preferably with the same or lower input of labour and with lower prices (all per unit of output) – with prices not falling to much that they dissuade farmers from adopting the innovation. This was the winning combination during the Green Revolution in Asia: Higher yields, increased labour productivity and lower prices (again per unit of output). Together with greater total labour absorption benefitting agricultural labourers, it made Green Revolution into, in many aspects, a win-win game.

This has not proved easy to replicate, for example in a sub-Saharan setting, although some products of crop breeding have shown promising results. Developing such crop technologies must remain a goal of CG research and therefore will call for impact assessment, if not *ex ante*, at least *ex post*. The methodological problems encountered resemble very much that in poverty assessment. Like in poverty assessments, the impact pathways are long and difficult to cover. One might add other aspects of food and food consumption, like food security, safety, quality and the effects of all this on nutritional status. Thus going from the quantity of food consumed to other aspects of food, introduces a number of methodological challenges.

Assessment of changes in nutritional status for example calls for studies of diets in various target groups. Although this is established methodology among nutritionists, we are talking about labour- and skill-intensive research, costly to run and, therefore, difficult to scale-up[9] in geographical coverage, with effects on representativity and generalisability. The perennial question of benchmarks recurs: These are unlikely to come by, and would once again force researchers to draw conclusions on processes based on cross-sectional data and problematic counter-factual arguments. Pseudo-panels may be a way out, as could be comparisons between adopting and non-adopting areas, or statistical analysis of differences between adopters and non-adopters.

As the structural transformation of the economy proceeds, food production as well as consumption is likely to diversify and value chains be elongated between primary producers and final consumers. New problems emerge, for example connected to food safety and quality. Many developing countries have rudimentary institutions controlling hygiene, pollution, adulteration etc., but in most cases their presence would be little felt on the ground and among the consumers. Exceptions here would be China,[10] India and a few others. Thus and as always there is little in terms benchmarks, or time series of crucial data to be used for impact assessments. And as always the methodological fixes are neither hard nor fast.

Welfare

Under the heading of welfare we bring together the discussion of longevity, housing, health and hygiene and children's welfare. Life expectancy is generally considered to be one of the most robust indicators of levels of living in general, at low levels of economic development primarily of quantity of nutrients consumed and basic hygiene and sanitation. At higher levels of economic development, other determinants come in, like quality of food (discussed above), education, especially of women, and social service delivery, especially in the health sector.

Another advantage with life expectancy is that fairly reliable statistics are available in most circumstances. Maybe less reliable but still quite useful statistics are also available on food

[9] India is an exception here, with a long-standing record in nutrition studies.

[10] The melanin scandal in China shows, not that the country is food safe, but that its apparatus for monitoring food safety is more developed than in an average developing country.

consumption in various aspects from FAO and on hygiene and sanitation from the World Bank and others. This also goes for some aspects of children's welfare,[11] but less statistics are produced about housing. The latter is a favourite among survey researchers, because it is easy to observe in the field and is a good indicator the economic status of a household. With longitudinal data it becomes an excellent indicator of the economic progress of a household or a village. In an African setting, the easily estimable frequency of corrugated iron roofing and the average age of the roofs are as good indicators of well-being and business climate as any other.

Welfare indicators available in official statistics are valuable in impact assessments made at higher levels of geographical aggregation and not aiming primarily to collect field level data. Housing, on the other hand, is convenient in survey based research. In the latter, however, it may be difficult to tie for example longevity indicators to individual households and the same would hold for a number of health indicators. Hygiene and sanitation standards, on the other hand are easy to observe, but as always the perennial question of benchmarks arises. Housing and sanitation would be easy to handle with pseudo-panel data, but not hygiene for example.

Gender impacts

Although the UNDP among others produces statistics on gender in development, when it comes to the many-faceted issue of gender relations, these indicators are still fairly crude. The obvious exception is gender-based statistics on education. If in an impact assessment one needs to go more deeply into gender relations for example in farming, there is no way to escape field based methods and qualitative methods. Gender relations are difficult to handle with survey methodology, unless one works with male-female teams of interviewers and a strategy to interview women independently and away from male supervision and control. Such a strategy can provide a number of indicators on e.g. division of labour or responsibility, but would call for an ethnographic type of field work to get insight into the causal mechanisms producing and reproducing gender relations in a local context. Again benchmarks would be problematic, although some of the more easily retrievable indicators may be possible to handle in a pseudo-panel approach.

Changes in empowerment of the poor

Despite the concern for non-economic indicators of human well-being in contemporary development thinking workable measurements of empowerment are still embryonic. Due to the "soft" nature of this indicator qualitative methods are probably the most appropriate approach.

Recently, however, the Poverty Reduction Group (PRMPR) at the World Bank has created an analytical framework for measuring empowerment (see Alsop and Heinsohn 2005). Empowerment is condensed into measurable components through the analysis of (1) people's capacity to realize choices and ambitions, i.e. personal agency, and (2) the surrounding opportunity structure, i.e. the institutional setting where people make choices. While indicators of personal agency are measured with "asset endowments", e.g. psychological, informational, financial, material or social assets that people use to make purposive choices, indicators of opportunity structure are measured by the existence and practice of formal or informal institutions, e.g. laws, regulations and norms, which may obstruct or support the realization of personal choices. Together these indicators measure the degree of empowerment by a

[11] The were asked to discuss time spent on nurturing children, which is really far down the impact pathway for most CG research. Unavailable in secondary data, this would require field based research to be adequately covered.

combined assessment of (1) the existence of choice, (2) the use of choice and (3) the achievement (effectiveness) of choice. Analysis and data collection may take place in three domains, i.e. state, market and society,[12] which subsequently can be divided into several sub-domains or levels depending on the location of people's choices. According to the PRMPR, this framework requires both qualitative and quantitative data and can be used in projects and interventions but also for measuring variations in empowerment at national level.

A further example is a study on vegetable and fishpond technologies in Bangladesh (Hallman, Lewis et al. 2007). Using the frequency of women's visits outside the home, their attendance of meetings and knowledge of local politics as a yardstick, it shows that quantitative methods have a potential in the assessment of empowerment.

Vulnerability

Much of the research done on vulnerability to natural disasters has been done in the sustainable livelihoods tradition. On the other hand, vulnerability to economic shocks have been quite extensively investigated by economists (see e.g. Glewwe and Hall 1998). The first mentioned type of studies have produced a number of descriptions which may prove an inspiration for an impact assessment if one is intending to go into these areas. However, in the last few years researchers have given much attention to development of indicators, which would simplify the choice of methods (for an overview and introduction, see Füssel and Klein 2006).

There are reasons to recommend a combination with survey methodology, if for no other reason to be able to contribute to the development of indicators. Since studies of vulnerability is a fairly new field in development studies, the problem of benchmarks may prove more intractable than otherwise. However, we would like to point out that teams of social scientists and engineers may be able to spot both sources of vulnerability (quality of water works, standards of housing and roads, etc.) and to estimate their development, including if they are going from poor to worse or the other way round. Causal attribution would be more complicated. Vulnerability has other dimensions which are more difficult to cover, e.g. vulnerability to economic shocks, health hazards etc. Again qualitative methods may make a contribution, but mainly at micro level.

Migration

In what connection can the effects of CG research on migration become an issue? A possible case can be foreseen, e.g. in connection with the structural transformation of the economy. It is for example an issue in South Asia that the distribution of labour power between the agricultural and non-agricultural sectors of the economy does not reflect the share of the sector in GNP (Hazell 2008). Surplus agricultural labour released from a rural sector characterised by rising agricultural labour productivity has historically been met by a growing labour demand within a nascent and expanding urban sector and constitutes an essential aspect of balanced growth (Tiffen 2003; Asian Development Bank Study Team 2005). The composition of labour demand in the urban areas and whether migrant streams fulfill the traditional selection biases in terms of age, gender and education is also important, or as has been

[12] This is a misleadingly simple triad; as Frödin {, forthcoming 2009 #2005}has shown it may be quite misleading in a field situation because it assumes an institutional setup which cannot be taken for granted, for example in Africa.

seen more recently in the context of Sub-Saharan Africa for instance, are composed also of large sections of dependents who may have more difficulties in securing employment in urban areas.

Historically or currently the question may therefore arise of impacts of agricultural research on migration propensities, streams and consequences for the migrants and for leaving and arrival areas. How would one approach a study of that issue? In a statistical superpower like India much can be done through the longitudinal databases available for example through the National Sample Survey (NSS). This would not be replicable in many other countries, except maybe China, if the authorities permit. Population statistics would always give some hints, but when it comes to the drivers and effects of migration, one would be at a loss. Linking data from such sources, especially if they are possible to disaggregate to district or sub-district level, with other data, for example consumer expenditure surveys, could potentially advance investigation somewhat. In this way one could get some estimates of the poverty effects of migration. It is possible that one can also link with farm economic data and thus at least roughly be able to discern the effects of changes in the farm sector.

If one needs to get further than this, one would be constrained to use the kind of methodologies that we are recommending, i.e. field based data collection (in this case maybe both in sending and receiving areas) and a combination of qualitative and quantitative methods and a pseudo-panel approach.

Environmental indicators

To assess the environmental impacts of new technologies on agricultural systems appropriate indicators, suited to specific regions and issues, are required. Key issues in any indicator or assessment system is finding ways to link drivers to *pressures, state, impacts, and responses*. This is called the *DPSIR system* and is widely used in Europe (Reynolds et al., 2007). The following is an attempt to adopt the thinking that lies behind the DPSIR system, the idea being, of course, that assessment also in the developing countries should eventually evolve into a level of refinement where DPSIR is fully implemented.

The possibilities of using remote sensing for assessing the impact of CG research are specifically highlighted. As we will see, new technologies in this field offer advantages, not only cost-wise, but also in terms of precision (Jensen, 2000). A surprising number of indicators are potentially available and although these indicators do not address all the needs of the CG system, they offer enough possibilities to make them highly interesting and relevant. Similarly, air-borne radar and LiDAR systems are increasingly being used for environmental monitoring. These techniques are also useful for water-related variables (McKean et al., 2008). Aerial photography continues to be immensely useful and digital air photo cameras are now taking over the market. Such imagery can be produced without the need for advanced photo labs, and they can be more easily loaded into digital photogrammetric work stations. Air photos are, however, less suitable for mapping of large areas or frequent monitoring of changes.

From January this year, digital Landsat images with pixel sizes between 15*15 and 80*80 metres and going back to 1972 are available from the US archives at USGS free of cost (http://landsat.usgs.gov/). Moreover, satellite imagery with much higher resolution (presently down to 0.5 x 0. 5 metres) is commercially available at terms which the team judges as affordable, at least in studies at low levels of geographical aggregation (www.geoeye.com, www.digitalglobe.com). The disadvantage here would

be that imagery with high level of resolution is not available in long time series[13] and that each scene also has a limited geographical extent. Another and growing advantage is that the development of new applications and methodologies is expected to accelerate with the decreased costs of satellite data.

Changes in deforestation

Changes in vegetative cover in general are obviously suited for assessment by satellite data. First of all the level of resolution is high, which makes the method useful even at quite low levels of geographical aggregation. Secondly, historical data are available, making it possible to trace changes in vegetative cover. Geographers and others have developed GIS-based modelling which is useful also for indicating the causal factors behind such changes. With these models, moreover, one can combine geographical data with economic and social ones.

As already pointed out, types of vegetation, including forest types are possible to study (Scherf et. al., 2006). Soil erosion is likewise amenable to study in fine detail with the high levels of resolution now available (Lu et. al., 2004). In general, most kinds of differences in land use (whether be it differences between adjacent sites, or differences in the same site between different points in time) can today be discovered, documented and interpreted with an extremely high level of accuracy. This holds for, e.g., forested vs. deforested areas, grazed vs. non-grazed areas, and eroded vs. non-eroded areas. Biodiversity is a potential further field of application, where remote sensing techniques can provide information about the extent, and change of extent, of biotopes.[14]

Changes in soil fertility

Soil fertility is a function of many soil properties, of which several are more or less interrelated. In most cases the term soil fertility describes the present capacity of the soil to sustain plant productivity. Soil fertility is a combination of the inherent soil quality (e.g. mineral composition, soil texture) and qualities achieved through soil formation and/or soil management (e.g. structure, soil organic matter content and phosphorus availability). There is neither a unique definition, nor a single indicator of soil fertility. However, the soil fertility concept is frequently used both by researchers (Sanchez, 1994; Vanlauwe, 2004) and by farmers (Elias, 2002).

Soil fertility can be estimated either through crop performance (yield) or through indirect indicators (e.g. content of Soil Organic Matter (SOM), indicator plants, water holding capacity). Whichever estimator is used, it is clear that bench-mark data will normally be required.

Poor or declining soil fertility can be due to a multitude of different factors. Therefore, before appropriate management methods can be developed there is a need to identify why the soil fertility is poor or declining. This is particularly relevant since even though the concept of soil fertility is firmly rooted in natural science, the drivers behind declining soil fertility are often socio-economic. Reasons for no adoption of management methods may be anything from unsuitability of the method itself, to poor governance or lack of resources in target countries.

[13] A German company is offering imagery at this level of resolution via its website: www.rapideye.com.

[14] See for example Remote Sensing of Environment Vol. 114, Issue 5, 2008: Special issue about Earth Observations for Terrestrial Biodiversity and Ecosystems.

Changes in biodiversity

Mapping of biodiversity requires time-consuming field work by skilled biologists, only feasible on a low level of geographical aggregation. Given that the relevant expertise is available, and that the geographical scale is realistic, inventory techniques can be readily adopted from the wealth of survey programs that are being run in many countries in Western Europe and the US. In general, birds and vascular plants, including aquatic macrophytes, are most easily assessed, whereas insects (and other arthropods) are examples of taxa that are considerably more difficult to deal with. Linkages between terrain and hotspots of terrestrial biodiversity have also been established (Zinko et. al., 2004).

Benchmark data is in principle needed, although this depends also on degree of precision required.[15] However, we want to point out two cases where assessments can be made *ex post*, at least in principle. One is the case where the original biotype is well-known and if changes are fairly drastic; then it should be possible to make assessments *ex post*, at least for well-studied countries. The other case is where the fauna and flora of a given region is sufficiently well known to allow for an ensemble of so called key-stone species to be identified. These are species which can be used as indicators, or early warning signals, of an ongoing decline in biodiversity (or some dimension of it).

Changes in chemical pollution

There are considerable difficulties with documenting and quantifying the effects of chemical pollution, first because benchmarks would seldom be available in a developing country context, and second because the number of conceivable chemical substances that could be relevant for assessment is formidable. However, for the purpose of this review it could be useful to distinguish between two classes of chemicals, those that pose a potential environmental problem because they are persistent and hence are accumulated in food chains, and those that are not accumulated but still potentially harmful. Typical representatives of the former class would be some insecticides, notably DDT, while the latter class encompasses nutrients among others.

Persistent chemicals

In many developing countries there is a major problem with obsolete stocks of pesticides, i.e. pesticides that cannot be used as intended because of loss of efficacy, poor formulations etc. In some cases such stocks have been dumped in developing countries as a means of removing a contamination problem in a donor country, and although national authorities may be aware of the problems locals in many cases use these stocks in ways that are not feasible. Also, pesticides may be sold under false names because the farmers want to use a familiar brand, thus concealing the fact that a new compound (which may have higher efficacy and be more environmentally friendly) is used.

Despite these caveats it is also clear that the knowledge of the dynamics of persistent chemicals in food chains and of their toxic effects on wildlife and man is substantial, since the days when DDT, methyl mercury and other persistent compounds were extensively used in agriculture in Europe and the US (e.g. Bernes 1998; Stockholm Convention against Persistent Chemicals www.pops.int). This means that if the presence of a persistent substance in the food chains in a developing country can be established, one can resort to literature data to estimate what effects that established level of the

[15] For low precision data, remote sensing is a definite possibility although some methods development is needed.

substance could conceivably exert on the ecosystems and on human consumers. For an example of an assessment of pesticide effects in a specific area, cf. Palm (2007).

Also, it should be possible from models that can be found in the literature to estimate with fair accuracy the expected dynamics of the substance subsequent to a changed exposure regime (e.g. that the emission of the substance into the environment is momentarily discontinued). This, then, would be a case of using counterfactual arguments, and would in principle suffer from the weaknesses pointed out above, but given the detailed knowledge in the literature regarding these substances the conclusions would still be reasonably valid.

Another kind of logic would be to simply take advantage of the fact that the biological effects (e.g. degree of toxicity, degree of specificity, half-life of degradation) of most pesticides are fairly well known and the information is readily available in the literature. From the point of view of environmental impact it should typically be advantageous to shift from a toxic to a less toxic substance, from a less specific to a more specific, or from one with a long half-life to one with a shorter half-life. In other words, there is, in principle, no need for any measurements in the field because the qualitative effects on the environment of a given technology shift (from one substance to another) can in principle be assessed purely on theoretical grounds.

In this context it can be pointed out that, to the extent that import and sales statistics for chemical substances are available and the geographical area where they are being used can be inferred, such statistics can be used as an accurate estimate of the level of emission of chemicals into the environment. In many countries historical data may be available so that changes in emission over time can also be estimated. Thus one would come part of the road with secondary data. A complication is, however, that sales statistics in many developing countries are not reliable since smuggling of pesticides is taking place.

Nutrients

With regard to the second class of chemicals, we would normally not expect benchmarks on nutrient leakage to be available. In their absence, one would need models for nutrient circulation in the ecosystem to be able to predict or estimate the effects of a change, e.g., in nutrient management. Thus, one is likely to be constrained to make qualitative assessments. In these, it is important to distinguish between the dynamics and chemical properties of different nutrients. For example, phosphorus is bound by most types of soils, which implies that leakage of phosphorus does not prevail in some types of soils or weather conditions, or occurs with a time delay that can be substantial. Nitrogen, on the other hand, is usually not bound nearly as hard as is phosphorus, but is readily washed out by precipitation or irrigation. For all nutrients, but particularly so for nitrogen, the content of organic matter in the soil exerts a strong influence on the dynamics and leakage.

At the moment of writing, it is unclear to what extent satellite imagery can be used for assessment of nutrient contents and leakage. However, it should be noted that in many crops the nutrient content (notably nitrogen, but also some micronutrients) of the above-ground parts is reflected by the colour (e.g. the more nitrogen the deeper the green colour). Hence, there is at least a theoretical possibility that satellite imagery can be used to estimate nutrient contents of a crop. This needs to be investigated further.

In any rate we envisage that CGIAR would not typically build in-house capacity for interpreting satellite imagery, but rather hire the expertise needed.

An interesting option for Africa is to coordinate CGIAR-related assessments with a recently financed network for deposition measurements organized by the Pan-African START Secretariat in Nairobi. One part of the work within the network will be to elucidate how much of nutrients in the African environment that comes in the form of atmospheric deposition (originating from biomass burning) compared to other sources, including agriculture.

Changes in water quality

Clearly, benchmark data are needed in order to assess changes in water quality, unless we can assume that a pristine nature prevailed before the technology shift occurred. And even then, the natural variability in pristine conditions needs to be assessed (Bishop, 2008; Erlandsson et. al., 2008) At any rate, assessment of water quality is an area where there are well-established procedures to draw on in the industrialized world.

Typically, aquatic assessments would analyze a number of inorganic parameters, such as acidity and the concentration of nitrate and phosphate ions, and a number of biological indicators (species assemblages), or specific pathogens such as *E. coli* if wastes from human settlements should be traced. They would then assess the hydrological regime and build a dynamic model of the fluxes of various substances (Håkanson and Boulion, 2002). Such models can predict, with a reasonable accuracy, the expected change in water quality subsequent to a technology shift that affects one or more fluxes from the catchment basin to the focal water body. Conversely, changes in fluxes can be inferred with reasonable accuracy from an observed change in water quality in the focal water body.

The basis for using models is firm knowledge of the systems. To get the leverage of modern methods, capacity to apply models, and the input data, are needed.

It should be noted that the section on water quality is focused on chemical indicators because we believe that these indicators are most feasible in most contexts that are relevant to the CGIAR. In the industrialized world, however, such a focus would appear to be somewhat out dated. The European Water Framework Directive (European Commission, 2000) stresses the use of biological endpoints (e.g. fish, benthic invertebrates, macrophyte, phytoplankton and benthic diatom assemblages) to assess the quality of surface waters (e.g. Allan and Castillo, 2008; Giller and Malmqvist, 1998). The directive also recognizes that aquatic ecosystems are not isolated entities, but interconnected systems (as does the EU Habitat Directive). Hence, determination of pressures and restoration effort is now focused on "fixing" the problem, which usually means altering and restoring catchments, not site-specific (habitats) integrity. In due time, the CG will undoubtedly want to move into this area, i.e. adopting the EU-approach.

Changes in greenhouse gasses

Given that benchmark data can be assumed, net effect of agrarian changes on the emission and binding of greenhouse gases can in principle be estimated fairly easily. Generally one needs to know the consumption of fossil carbon - both directly in the form of fuel for tractors etc. and indirectly in the form of fuel burnt for manufacturing fertilizers and other commodities. This consumption of fossil carbon inevitably leads to a release of carbon dioxide from the system. The biological production in the system is of course the balancing factor which consumes carbon dioxide during photosynthesis. The same carbon dioxide is again released to the atmosphere when the produce is decomposed. This is where various time delays are operating in the system. The above-ground parts of a crop are generally decomposed within a few months subsequent to harvesting (or momentarily in case burning is

practiced, as in sugar canes), whereas the decomposition of the root system is usually a much slower process. A tree in a forest may introduce a time delay of many years between assimilation and emission of carbon dioxide – a delay that can be further prolonged if the tree is used for construction purposes. Such time delays are highly valued in balances of greenhouse gases, because they allow for a net absorption of carbon dioxide today, which will only be "paid back" many years later. This is why foresters all over the world are much involved in calculating the optimal forestry strategy with respect to binding and release of carbon dioxide.

In general the relevant knowledge as to which are the key processes in a carbon dioxide balance on the farm or village scale is available. Its application to the case of a technology shift in a developing country should therefore be fairly straightforward. Some key references in the field are CEN (2006a, 2006b), Rebitzer and Ekvall (2004), Pennington et.al. (2004) and Schlamadinger et.al. (1997).

Changes in animal welfare

Besides counting heads, reproductive and productive performance of livestock is a good integrated read-out of animal health and welfare. Robust traits in domesticated mammals are birth-intervals and litter sizes (where applicable), and number of eggs for chicken.

Thus, the frequency of lambing, kidding, calving etc. and litter size could be combined into the single variable "number of offspring produced per time unit." This offers an easily available and accurate integrator of most factors that affect animal health and welfare. It should be noted that this variable is relevant and accurate even in situations where artificial insemination or other means of planned and scheduled reproduction is practiced. Moreover, it is readily available and easy to monitor and also usually accurately tracked by the farmers because of its economic importance to them.

Serological surveys of major infectious diseases are another, slightly more demanding, technique, which has the advantage of specifically measuring the disease component of animal health and welfare. Relevant laboratory techniques are available which integrate over a range of the most important infectious diseases.

Clearly benchmarks are required for both techniques. Serological samples are easily stored and in some areas it may very well be that historical data can be retrieved from such stored samples which would enable benchmarks to be achieved.

Conclusions on environmental indicators

As the review of earlier studies indicated, the state of the art when it comes to assessments of environmental impact of agricultural research in developing countries is not very advanced. On the other hand it should be clear from the very brief accounts above that the state of the art of environmental impact assessment in the industrialized world is a completely different story.

Ever since the general awareness regarding hazardous environmental effects of pollutants started to increase in the 1960s, much research throughout Western Europe and North America has been devoted to monitoring, understanding, assessing and solving these problems. Starting with the problems caused by extensive use of DDT and methyl mercury in the agricultural sector, via problems of eutrophication of water bodies caused by communal wastes or leakage of fertilizers, to changes in biodiversity caused by various practices of land use, an enormous body of scientific knowledge has today been

accumulated. Indeed, subjects like "environmental monitoring and impact assessment" have been so firmly established in academia that today they have their own curriculae, including research as well as undergraduate and postgraduate education. For example, at the Swedish University of Agricultural Sciences two or three fully fledged departments, each fully equipped with skilled staff and relevant laboratory facilities, are solely devoted to environmental impact studies. The same would be true for many other major universities in Europe and North America specializing on various aspects of biological production systems.

Hence, there is no doubt whatsoever that the CG system will be able to find the relevant expertise required in order to deepen and broaden the state of the art of impact assessment in and for the developing countries. While a great deal of methodological development is probably called for, new and refined technologies, including advances in modeling of environmental effects, also open prospects for much more effective and efficient assessment than has traditionally been the case in the industrialized world. As far as we can see, only some minor investments on the part of the CG in the development of such methods would be required in order for the CG to embark on assessment projects, preferably in cooperation with university researchers from donor countries.

Conclusions and list of potential case studies

Before coming to potential case studies, it may be worthwhile to recapitulate our major conclusions. We have strongly advocated the use of control areas or quasi-experimental methods with benchmarking (possibly with a pseudo-panel approach) as the preferable design of impact assessments. This is the only way to avoid problematic counterfactual arguments which may appear plausible to those who have already accepted the conclusion, but which will not convince those who are skeptic to the CG and its research. Since in the typical developing country context, the dearth of secondary data is likely to pose a major constraint, this would imply a survey-based approach and a combination of qualitative and quantitative methods. There are exceptions to this overall conclusion, for example, where data is less scarce, like in India or China, or where one can draw on previous research. Given that the World Bank methodology of poverty assessment is so widely accepted and used, it may also be feasible to collect data on the same (and lowest) level of disaggregation as this data (normally district or sub-district). For environmental indicators, as we have seen, feasible levels of disaggregation varies, so that when such effects are a priority, suitable research designs would come to vary with the topic of interest. However, it is a considerable advantage that the application of several of the environmental indicators can draw heavily on contemporary experience of similar environmental issues in the industrialized world. For policy impact research, finally, the scale is usually national or higher; such research moreover, as the review above has shown, varies a great deal in methodology and is difficult to manualise.

[case studies remvoved]

Save date: 2009-06-19 12.01 by GD

References

Adato, M. and R. Meinzen-Dick, Eds. (2007). *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. Baltimore, John Hopkins University Press.

Ajayi, O. C., F. Place, et al. (2007). World Agroforestry Centre. Impacts of Improved Tree Fallow Technology in Zambia. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 147-168.

Allan, J.D. and Castillo, M.M. 2008. Stream Ecology: Structure and function of running waters. *Springer*, 436 p.

Alsop, R. and N. Heinsohn (2005). Measuring empowerment in practice: Structuring analysis and framing indicators. *The World Bank Policy Research Working Papers 3510,* February 2005. Washington, D.C., The World Bank.

Asian Development Bank Study Team (2005). Regional Study on Rural, Urban and Sub-regional linkages in the Greater Mekong Sub-Region (Viet Nam, Lao PDR, Cambodia), final report, ADB study team based in Hanoi/Vientiane/Pnomh Penh/Bangkok, Asian Development Bank.

Bellon, M. R., M. Adato, et al. (2007). Improved Maize Germplasm, Creolization and Poverty: The Case of Tuxpeño-Derived Material in Mexico. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press**:** 238-284

Bernes, C. ed. 1998. Persistent organic pollutants. Swedish Environmental Protection Agency, Stockholm 152 pp.

Bhalla, S. S. (2002). *Imagine there's no country: Poverty, inequality and growth in the era of globalization*. Penguin Books, New Delhi.

Bishop, K., Beven. K., Destouni, G., Abrahamsson, K., Andersson, L., Johnsson, R., Rodhe, J., Hjerdt, N. Nature as the "natural" goal for water management: A Conversation. *Ambio* (in press).

Bourdillon, M., P. Hebinck, et al. (2007). "Assessing the Impact of High-Yield Varieties of Maize in Resettlement Areas of Zimbabwe. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press**:** 198-237

CEN (2006a). *Environmental management - Life Cycle Assessment - Principles and framework* (EN ISO 14040:2006), SIS & European Committee for Standardization (CEN). Brussels

CEN (2006b). *Environmental management - Life Cycle Assessment - Requirements and guidelines* (EN ISO 14044:2006), SIS & European Committee for Standardization (CEN). Brussels.

Chayanov, A. V. (1966). *A.V. Chayanov and the Theory of Peasant Economy*. Homewood, Ill., Richard D. Irwin.

Dalton, T. J., N. K. Lilja, et al. (2007). CIAT. Impact of Participatory Natural Resource Management Research in Cassava-based Cropping Systems in Vietnam and Thailand. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 91-117.

Dey, M. M., P. Kambewa, et al. (2007). Worldfish Centre. Impact of the Development and Dissemination of Integrated Aquaculture – Agriculture Technologies in Malawi. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 118-146

Diagne, A. and M. Demont (2007). "Taking a new look at empirical models of adoption: average treatment effect estimation of adoption rates and their determinants." *Agricultural Economics* **37**(2-3): 201-210.

Djurfeldt, G., R. Larsson, et al. (submitted). "African farm trajectories and the sub-continental food crisis." *Food Economics*.

Elias, E. 2002. Farmers' Perceptions of Soil Fertility Change and Management. SOS-Sahel and Institute for Sustainable Development, Addis Ababa, Ethiopia.

Erlandsson, M., Folster, J., Laudon, H., Weyhenmeyer, G.A., and Bishop, K. 2008. Natural variability in lake pH on seasonal, interannual and decadal time scales: Implications for assessment of human impact. *Environmental Science & Technology* 42(15): 5594-5599.

Fan, S. (2007). Agricultural Research and Urban Poverty in China and India. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press**:** 308-330.

Fan, S., C. Chan-Kang, et al. (2007). National and International Agricultural Research and Rural Poverty: The Case of Rice Research in India and China. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press**:** 285-307.

Füssel, H.-M. and R. Klein (2006). "Climate change vulnerability assessments: An evolution of conceptual thinking." *Climate Change* **75**: 301-329.

Giller, P.S. and Malmqvist, B. 1998. *The biology of streams and rivers*. Oxford University Press, Oxford.

Giordano, M. A., M. Samad, et al. (2007). IWMI. Assessing the Outcome of IWMI's Research and Interventions on Irrigation Management Transfer. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 196-216.

Glewwe, P. and G. Hall (1998). "Are some groups more vulnerable to macroeconomic shock than others? Hypothesis tests based on panel data from Peru." *Journal of Development Economics* **56**: 181-206.

Haggblade, S. and B. Zulu (2003). *The Recent Cassava Surge in Zambia and Malawi*. Successes in African Agriculture, Pretoria, Inwent, IFPRI, NEPAD, CTA.

Håkanson, L. and Boulion, V.V. 2002. *The lake foodweb: modelling predation and abiotic/biotic interactions*. Backhuys Publishers.

Hallman, K., D. Lewis, et al. (2007). Assessing the Impact of Vegetable and Fishpond Technologies on Poverty in Rural Bangladesh. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press**:** 103-148.

Hazell, P. (2008). *An Assessment of the Impact of agricultural research in South Asia since the Green Revolution*. Rome, Science Council, Consultative Group on International Agricultural Research.

Hempel, C. (1965). *Aspects of Scientific Explanation*. New York, Free Press.

Himanshu (2008). "What Are These New Poverty Estimates and What Do They Imply?" *Economic and Political Weekly* **43**(43).

Honderich, T., Ed. (1995). *The Oxford companion to philosophy*. Oxford, New York, Oxford University Press.

Hossain, M., D. Lewis, et al. (2007). Rice Research, Technological Progress, and Poverty: The Bangladesh Case. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press.**:** ", pp. 56-102 in

Jensen, J.R. 2000. *Remote sensing of the environment : an earth resource perspective*. Prentice Hall.

Kanbur, R. and P. Shaffer (2007). "Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice." *World Development* **35**(2): 183-196.

Kijima, Y., K. Otsuka, et al. (2008). "Assessing the impact of NERICA on income and poverty in central and western Uganda." *Agricultural Economics* **38**(3): 327-337.

Krishna, A. (2004). "Escaping Poverty and Becoming Poor: Who Gains, Who Loses, and Why?" *World Development* **32**(1): 121-136.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, Chigaco University Press.

Laxmi, V., O. Erenstein, et al. (2007). CIMMYT. Assessing the Impact of Natural Resource Management Research: the Case of Zero-Tillage in India's Rice-Wheat Systems. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 68-90.

Lu, D., Li, G., Valladares, G.S. and Batistella, M. 2004. Mapping soil erosion risk in Rondônia, Brazilian Amazonia: using RUSLE, remote sensing and GI. *Land Degradation & Development*. 115: 499-512.

McKean, J.A., Isaak, D.J., Wright, C.W. 2008. Geomorphic controls on salmon nesting patterns described by a new, narrow-beam terrestrial–aquatic lidar. *Frontiers in Ecology and the Environment* 6(3): 125-130

Nweke, F. I., D. S. C. Spencer, et al. (2002). *The Cassava Transformation - Africa's best-kept secret*. East Lansing, Mich., Michigan State University Press.

Palm, B. 2007. *Pesticide use in rice cultivation in Tarapoto, Peru*. Diploma thesis, Dept of Environmental Assessment, SLU, Uppsala.

Pennington, D.W., Potting, J., Finnveden, G., Lindeijer, E., Jolliet, O., Rydberg, T. and Rebitzer, G. 2004. Life cycle assessment Part 2: Current impact assessment practice. *Environment International* 30(5): 721-739.

Place, F., M. Adato, et al. (2007). Impacts of Agroforestry-Based Soil Fertility Replenishment Practices on the Poor in Western Kenya. *Agricultural Research, Livelihoods, and Poverty. Studies of Economic and Social Impacts in Six Countries*. M. Adato and R. Meinzen-Dick. Baltimore, The Johns Hopkins University Press**:** 149-197

Raitzer, D. A. (2007). *Assessing the Impact of CIFOR's Influence on Policy and Practice in the Indonesian Pulp and Paper Sector*, Draft. Rome, Standing Panel on Impact Assessment, Science Council, Consultative Group on International Agricultural Research.

Ravallion, M. (2008a). "A Global Perspective on Poverty in India." *Economic and Political Weekly* **43**(43).

Ravallion, M. (2008b). "'How Many Poor in the World?': A Reply." *Economic and Political Weekly* **43**(45).

Rebitzer, G. and Ekvall, T. 2004. Life cycle assessment Part1: Framework, goal and scope definition, inventory analysis, and applications. *Environment International* 30(5): 701-720.

Reddy, S. G. (2008). "Differently Distorted: The World Bank's 'Updated' Poverty Estimates." *Economic and Political Weekly* **43**(43).

Reynolds, K.M., Kohl, M., Thomson, A.J., Shannon, M. and Ray, D. (eds) 2007. *Sustainable Forestry: From Monitoring and Modelling to Knowledge Management and Policy Science*. ISBN: 9781845931742

Rodrigues, S. G. and A. M. Viñas (2007). "An Environmental Impact Assessment System for Responsible Rural Production in Uruguay." *Journal of Technology Management & Innovation* **2**(1): 42-54.

Sanchez, P.A. 1994. Tropical soil fertility research: Towards the second paradigm. *15th World Congress of Soil Science*, Acapulco, Mexico, pp. 65-88.

Scherf, M., Atzberger, C. and Hill, J. 2006. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment* 95(2): 177-194.

Schlamadinger, B., Apps, M., Bohlin, F., Gustavsson, L., Jungmeier, G., Marland, G., Pingoud, K. & Savolainen, I. (1997). Towards a standard methodology for greenhouse gas balances of bioenergy systems in comparison with fossil energy systems. *Biomass and Bioenergy* 13(6): 359-375.

Shideed, K., V. Alary, et al. (2007). ICARDA. Ex post Impact Assessment of Natural Resource Management Technologies in Crop-Livestock System in Dry Areas of Morocco and Tunisia. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 169-195

Spilsbury, M. J. (2007). CIFOR. The Sustainability of Forest Management: Assessing the Impact of CIFOR Criteria and Indicators Research. *International Research on Natural Resource Management. Advances in Impact Assessment*. H. Waibel and D. Zilberman. Oxfordshire, CABI**:** 217-245

Templeton, D. and N. Jamora (2007). Economic Assessmemt of IRRI's Policy-oriented Research into the Private Health Costs of Pesticide Use in Rice Farming in The Philippines. Rome, Standing Panel on Impact Assessment, Science Council, Consultative Group on International Agricultural Research.

Tiffen, M. (2003). "Transitions in Sub-Saharan Africa: agriculture, urbanization and income growth." *World Development* **31**(8): 1343-1366.

Vaidyanathan, A. (2006). "Farmers' Suicides and the Agrarian Crisis." *Economic and Political Weekly*.

Vanlauwe, B. 2004. Integrated soil fertility management research at TSBF: The framework, the principles, and their application. In *Managing nutrient cycles to sustain soil fertility in Sub-Saharan Africa* (Ed. A.Bationo). 25-42. Academy Science Publishers (ASP), Nairobi.

Waibel, H. and D. Zilberman, Eds. (2007). *International Research on Natural Resource Management. Advances in Impact Assessment*. Oxfordshire, CABI.

Zinko, U., Seibert, J., Dynesius, M., and Nilsson, C. 2005. Plant species numbers predicted by a topography based groundwater-flow index, *Ecosystems* 8: 430–441.