IPM Impact Assessment Series

Guidance Document

Introduction to the "Double Delta" Approach

June 2008

The IPM Impact Assessment Series aims to provide guidance to impact assessment of Integrated Pest Management (IPM) projects and programmes.

This Guidance Document: "Introduction to the “Double Delta” Approach", was prepared by Marc Voelker, University of Hannover, in close consultation with Prof. Hermann Waibel, University of Hannover, and Harry van der Wulp, FAO Global IPM Facility.

The preparation of this guidance document was commissioned by the FAO Global IPM Facility.

Table of Contents

1. What is this Guidance Document about and who should read it?

This Guidance Document describes the purpose, methodology and procedure of impact assessment of Farmer Field Schools (FFS) on Integrated Pest Management (IPM).

It explains the methodology of econometric impact assessment studies, notably the "double delta" approach, which is also referred to as difference-in-difference. Special attention is given to the study design, conducting surveys and analysis and interpretation of data. The final chapter identifies pitfalls and provides lessons learned from past studies.

This Guidance Document is aimed at staff of national IPM programs and field practitioners involved in the design and interpretation of impact assessment studies.

Farmer Field Schools on Integrated Pest Management

Farmer Field Schools (FFS) are a participatory approach to educate farmers, for instance in understanding and practicing Integrated Pest Management. Instead of being given "top-down" instructions by administrators and researchers on what to do or not to do, farmers learn to make informed management decisions based on their own observations and ecosystem analysis. During a typical FFS, farmers meet in weekly sessions from planting to harvest. They conduct simple experiments to develop a better understanding of the functional relationships between pests and natural enemies, and between crop damage and yield. They observe and discuss the developments in their fields and decide upon crop management interventions. A trained facilitator guides activities.

2. Why Impact Assessment of Farmer Field Schools is needed?

The basic aim of impact assessment of farmer field school programs is to determine to what extent such programs have achieved their objectives and brought about the desired effects. This includes verification whether those effects can be actually attributed to the program.

Impact assessments serve to inform different interest groups about the results of the programme concerned. Interest groups can be identified at three different levels:

- **Farm-level**: Participating and non-participating farmers.
- **Programme level**: Project or program staff at the institutional level, ranging from field trainers to national programme coordinators.
- **Policy level**: Policy makers and donor agencies.

Farmer field school programs need to provide benefits to the participating farmers. Investment in the programme should be economically justifiable from the society's point of view and compare positively to other similar investments.

While farmers tend to base their decisions on personal experience, scientific impact assessment might be required by policy makers and donor agencies to quantify the benefits arising from investment in FFS programmes and to help them determine where to best allocate their scarce financial and human resources. This is particularly relevant in cases, where decisions are to be taken regarding substantial investments for the scaling up of FFS programs.

Furthermore, impact assessment provides an important tool for review of the effectiveness of programs. Findings will help identify areas for improvement. Training quality and learning achievements can be monitored and the findings analysed and jointly assessed by program facilitators and administrators, and farmer groups. If problems are identified, training curricula and work plans can be improved for program continuation or design of subsequent farmer field school programs.

Because different interest groups, as mentioned above, have different reasons to conduct impact assessment studies, it is important to note that one cannot respond properly to the variety of needs with a single approach. Different interest groups with different objectives for impact assessment require different methodological approaches. The Double Delta approach is of particular importance if scientifically rigorous data are required for policy-level review.

3. What is special about Impact Assessment of Farmer Field Schools?

Impact assessment of FFS on IPM is complex due to the abundance of impact parameters and because evaluation can take place at different levels. Main success indicators include yield, pesticide use, input costs, profit, and the variation in yield and profit. Further indirect parameters, which are more difficult to measure, but not less relevant, are effects on agricultural biodiversity, soil quality, occupational and public health and the environment, as well as social effects and impact on policy development. The levels at which evaluation can take place are self-evaluation by farmers, self-evaluation by a project and external evaluation.

A considerable number of studies have assessed the impact of farmer field schools. These studies varied in focus, approach, methodology and statistical robustness. This diversity produced a range of outcomes from different perspectives. However, it is difficult to compare the results of these studies because studies were designed to be either statistically rigorous (but with a restricted scope, focussing on a few variables) or comprehensive (looking at a broad range of impact indicators, but without solid statistics), but never both. In order to evaluate benefits from more than one perspective and enhance the comprehensiveness of the overall evaluation, methods or results of different data sources need to be combined. See "Farmer Field Schools: a Review of 25 Impact Studies" for further information on this.

This Guidance Document explains how to set up a statistically rigorous study. The "double delta" methodology is a tool to reduce the risk of biased and ambiguous results in key parameters. This allows for a scientifically sound and qualified assessment of the actual impact of the training, which yields valid, meaningful and comparable results. Other perspectives can be gained, for instance, from in-depth case studies, which provide a broader scope of information but are less statistically rigorous.

4. Who should design and carry out an Impact Assessment study?

If an authoritive study is required, then it is advisable that an institution that is external to the FFS program should design and conduct the impact assessment study. Otherwise results can be biased since program staff may make choices and interpretations that positively influence the actual outcome of the study.

Universities and private consultants are examples of such external institutions that may be able to design and carry out impact assessment studies. A major requirement for such institutions is to have sufficient expertise and experience in agricultural economics and the statistical techniques to be applied in the study. This is of critical importance since mistakes in the design and implementation of the study are likely to lead to higher costs, time lags and biased results. Therefore the external institution to design and conduct the study should be selected carefully.

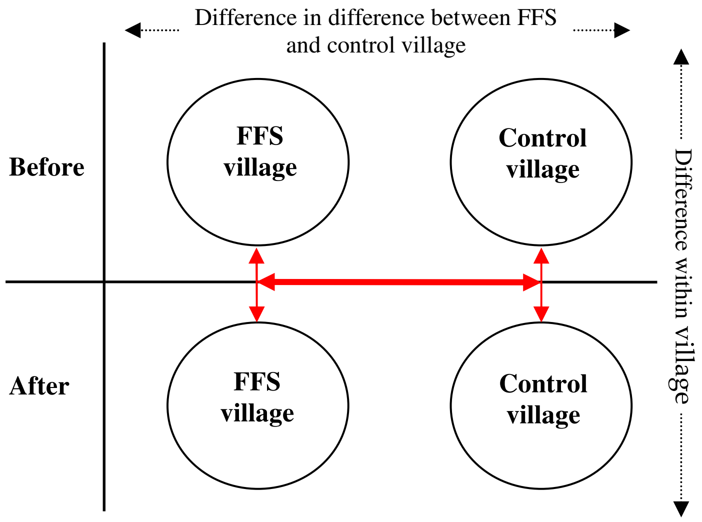
Assigning a university might have the advantage that its staff is more neutral, careful and on the cutting edge of science in its approach to impact assessment than private consultants. However this cannot be generalized. It is recommendable to only assign an institution that can provide suitable references, which show that it has the required expertise and practical experience. Different external institutions can be assigned for one impact assessment study. A university, for instance, can be responsible for the design of the study while local consultants carry out the survey.

In addition, it is important that the institution conducting the study has a good understanding of IPM and the FFS approach, or otherwise informs itself thoroughly and designs the study in close consultation with project staff.

5. "Double Delta": How does it work?

The basic idea of the "double delta" approach is to compare the situation before and after training for FFS participants and a control group of non-participants. Survey data are entered into a model to estimate the difference in performance on selected success indicators (e.g. amount of chemical pesticides applied) between the two groups. This way, any effects caused by other factors than training are "differenced out". For instance, if a drought occurs in the survey region, or if there are considerable changes in input prices, it will have the same effect on the yield of participating and non-participating farmers. It will also "difference out" any effects of policy measures that affect all farmers, such as, for instance, the banning of certain pesticides, or the removal of subsidies for pesticides.

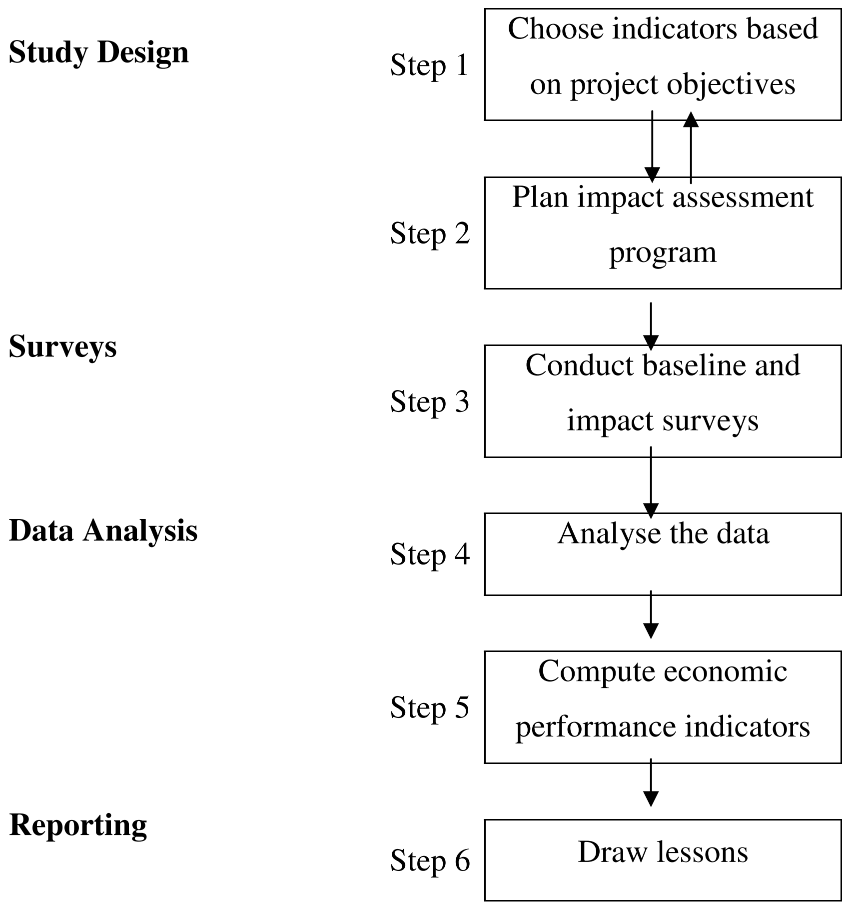
This is necessary to avoid that changes that result from external factors are attributed to the training. For instance, if pesticide prices increased between the base-line and the actual survey, one would want to know to what extent pesticide use reduction can be attributed to the project and to what extent to the price increase. Likewise, if yields after training were disappointed because of drought, then one would want to know how yields of affected FFS farmers compared to those of an affected control group.



Keep in mind that it is important to collect data from the control village not only after, but also before, the training. Although the farmers in both villages are untrained before the training, the FFS village cannot substitute the control village at that point of time. Although the villages should be basically equal, small differences in the socio-economic characteristics are likely to exist and need to be taken into consideration in the "double delta" model.

6. Steps in farmer field school impact assessment

The procedure of impact assessment of FFS on IPM can be divided into 6 steps:



Step 1: Choose indicators

As a first step, meaningful success indicators need to be identified that are suitable to measure to what extent a FFS program has achieved its objectives. The choice of success indicators is determined by the objective of the project and the purpose of the impact evaluation.

Generally, the main objectives of IPM farmer field schools are:
- to increase production in a sustainable manner,
- to reduce pesticide use and associated health and environmental risks,
- to improve farmer income,
- to empower farmers to understand the agro-ecosystem of their crop and to be able to take crop management decisions independently.

Farmers should become better managers of their fields by shifting their focus from pest control to maintaining crop health. They should identify the causes of the build up of pest populations that are resulting in yield losses. They should be enabled to make informed decisions based on an improved knowledge of the agro-ecosystem and the properties of input factors. Furthermore, farmers should learn how to build new institutions for farmers-to-farmers knowledge transfer (social involvement), solve problems independently and implement their own locally developed FFS programs.

Each interest group has its own particular reason for assessing impact. Therefore the choice of indicators also depends on the interest groups of the impact assessment study. Program staff would want to know whether their training has the desired effect and where it can be adjusted or improved. They will, thus, have a particular interest in the impact of their training on farmers' knowledge and behaviour. Donor agencies and governments want to assess whether the project contributes to broader development objectives related to rural development, sustainable intensification of agricultural production and reduced health and environmental risks related to pest management. Farmers, generally have their own set of indicators as they can directly determine the impact of the training they experience in their livelihood situation.

The effectiveness of FFS training in terms of improving participants' knowledge on IPM can be assessed by knowledge scores based on structured questionnaires. Through such questionnaires, farmers are asked how to identify the major pests in their fields, how these reproduce and how to control them. Further questions can be posed regarding farmers' understanding of different warning labels on pesticides and the precautions they take in applying and storing them. A score of one is assigned to the correct response to each of the questions and the sum of scores is the knowledge score. Thus, if a farmer, for example, answers correctly 10 questions his knowledge score will be 10. The higher the knowledge score is, the better the tested knowledge of the farmer. Knowledge questions can be posed in different ways. Multiple-choice questions with several right answers, for instance, can also be posed. Thus respondents can get a score of more than one for one question.

However, improving farmers' knowledge generally is only an important first step towards change in behaviour needed to achieve the program objectives. Table I shows how improved knowledge can change farmers' behaviour and which indicators can be used to measure related effects on economic well-being, health and the environment. Keep in mind, however, that this causality is assumed and that it can be challenged.

Improved knowledge →	**Change in behaviour** →	**Success indicators**
Crop management	Adoption of new approaches and technologies; adaptation of technologies to local conditions; use of beneficial insects	Yield; output prices affected by product quality; net profits
Knowledge on insect pests	Use of non-chemical pest control and more efficient use of pesticides in terms of quantities and timing	Use of non-chemical pest management; amount and types of pesticides used; expenditures on pesticides; number of applications and correlation with pest pressure
Knowledge on pesticide toxicity	Use of less toxic pesticides, use of personal protection	Share of highly toxic pesticides used; EIQ; soil and water quality; insect population counts (biodiversity); signs and symptoms of acute poisoning and correlation with spray events; pesticide-related health costs

Table I: Impact of improved knowledge on farmers' behaviour and indicators for measuring related effects

The profitability of farmers' efforts can be measured by calculating their net revenue. Therefore, their production costs and revenues need to be computed. Keep in mind that revenues are influenced by many factors in addition to FFS training, such as output prices. Remember that a control group is used in the "double delta" approach in order to take into account these factors. When calculating production costs the inclusion of labour costs is difficult. One must distinguish between hired and family labour. Costs for hired labour are the wages paid to the workers. However workers can also be paid in kind. Costs for family labour can be measured by taking into account the opportunity costs of labour. Opportunity costs are the benefits that a farmer could have received if time spent on FFS had been used for other activities that would have generated income. For instance, a farmer could have worked off-farm for some time during the day, thus receiving a specific wage. Since opportunity costs are difficult to measure it should be considered to exclude family labour costs from the computation of production costs.

Positive impacts on the agro-ecosystem and health are due to reduced pesticide use and selection of less hazardous pesticides. Changes in the quantity of pesticides used can be directly measured in terms of cost, dose (expressed in active ingredient or formulated product per surface unit) or number of applications. Composite risk indicators can be used to assess changes in the overall risk to health and the environment. An example of such an indicator is the Environmental Impact Quotient (EIQ), which reflects potential impact on farm-workers, consumers and fish, bees and beneficial arthropods. Empirical results have shown that the EIQ can capture more benefits of FFS on IPM than other pesticide reduction indicators because it reflects the combined effect of reduction in pesticide use and improvement in selection of pesticides. Thus, the use of less toxic, but more expensive, pesticides would come out as an improvement, while comparison on pesticide expenditure would show a deterioration. For detailed information about the use of EIQ in IPM impact assessment see the Guidance Document on the Use of EIQ.

Instead of using a composite environmental indicator that averages out several risks, such as EIQ, one can also look at indicators for a specific environmental or health impact. For instance, the ratio of natural enemies to insect pests can be measured by simple population counts. Measurement of effects on soil and water quality can be carried out by taking samples and conducting relevant tests. Poisoning incidence of livestock and fish can be observed. Health effects can be measured by surveying farmers' expenditures on medical treatment, days of work lost due to poisonings, signs and symptoms of acute poisoning and correlation with spray events, hospital records and health exams for professional sprayers.

Finally, the implementation of new institutions for farmer-to-farmer knowledge transfer, improved capabilities of farmers to solve problems independently and the development of locally FFS programs by the farmers themselves need to be measured. This can be done by monitoring the frequency and attendance rates of farmers planning meetings to solve pest problems, inter-village and intra-village forums and research committees for scientific learning, locally developed FFS and similar activities.

As mentioned before, it must be stressed that due to practical restrictions one cannot take into account all indicators in one impact assessment study because the more indicators are included in the analysis the higher the costs of the study will be. If different types of studies are conducted simultaneously and refer to the same farmer population, the results of these studies can be combined later.

When choosing indicators one also must consider the costs associated with the use of different indicators. Some indicators, such as the direct measurement of environmental effects in the environment, are more costly than others such as calculating the EIQ. Thus the budget available for the impact assessment study may also influence the choice of indicators.

Step 2: Plan impact assessment program

After having chosen meaningful indicators for the success of an FFS program as described in the previous step, it needs to be decided from which farmers the respective data should be collected. The basic procedure of FFS impact assessment is to collect information from a portion of the population of interest by taking a sample of elements from the larger group. Based on the information collected from the sample, conclusions can be drawn about the whole population of interest.

In order to draw a sample of a population it is first necessary to define the population about which conclusions should be drawn. Since the "double delta" model will be used to analyse the collected data, information must be collected from at least two groups of farmers: 1.) farmers who participate in the FFS training (direct impact group), and 2.) farmers who do not participate in the FFS training and who live in a control village with similar socio-economic conditions (control group). To guarantee that knowledge, which participants obtain in the training, and skills and practices, which they learn, will not be diffused to the control group it is required that only minimal possibility of information exchange between the village where the FFS training takes place and the control village exists. For example, the villages should not have a common market, church, mosque or temple where farmers regularly meet and where information exchange can take place. Optionally information can be collected from a third group of farmers who live in the village where the FFS training takes place, but who do not participate in the training. This group is referred to as "exposed farmers" or "indirect impact group". Thus, possible diffusion of knowledge can be assessed. A further option is to analyse the data by gender, class or other socio-economic factors in order to understand who is benefiting. Data should be collected at least at two points of time, right before the start of the FFS training and in the year after the FFS training. When the data are collected in the year after the FFS training farmers should have grown the respective crops at least for one season. Additionally, data can be collected, for instance, three years after the training in order to investigate its long-term effects, which however requires more time and financial resources.

Sample selection and sample size

Furthermore, it needs to be determined how the individual farmers to be included in the sample will be chosen and what the sample size will be. The farmers to be sampled should be chosen in such a way that they can be considered as representative for the population of interest. Choosing the participating farmers to be sampled in the FFS village should be carried out by first identifying those farmers who participate in the training and then randomly selecting a group of farmers from this subset (stratified random sampling). Since some farmers who are initially supposed to participate in the FFS may ultimately not participate, more farmers than theoretically needed should be selected. Thus later dropouts will not weaken the explanatory power of the econometric analysis. The non-participating farmers to be sampled from the FFS village should be chosen in an equivalent way while those from the control village can be chosen randomly from the village population. However it should be kept in mind that they should be farmers growing the same crops as those participating in the FFS.

The size of the sample needs to be large enough to enable researchers to draw valid conclusions. If the sample size is very small it might happen that, for instance, only a subset of farmers is surveyed, who for some reason were less successful in reducing pesticide use after FFS training, while average pesticide reduction of the whole population of farmers who participated in FFS is larger. Having a larger sample size increases the probability of getting a representative sample of the whole population of interest. A suitable sample size can theoretically be calculated by taking into account three factors:

- The margin of error the researcher is willing to accept in the study
- The level of risk the researcher is willing to accept that differences revealed by statistical analyses do not exist in reality (also referred to as significance level)
- The variance in the primary variables of interest in the study

The text below explains how to calculate a suitable sample size. Examples are given in order to make the calculation procedure easier to understand. Do not be scared by formulas however! In practice, econometric software packages are available that do the calculations for you. All you have to do is to specify the acceptable margin of error and significance level, and to enter a dataset of the particular observations of the variable of interest.

For those interested, details are provided in the grey box below:

As an example, let us assume that pesticide use, measured in kilogram per hectare, is the primary variable of interest in a study on the impact of FFS. First of all, the margin of error the researcher is willing to accept in the study needs to be determined. As a rule of thumb, 3% margin of error is acceptable for continuous data. This margin of error needs to be multiplied by the expected mean of pesticide use in the study, which can be derived from previous studies. If previous studies indicate an average pesticide use of 15 kilogram per hectare, the acceptable margin of error for the expected mean will be $0.03 \times 15 = 0.45$. This means that the pesticide use to be estimated in the sample should be within 0.45 kg per hectare of the true pesticide use of the whole population of interest.

As a next step, the significance level needs to be specified. A significance level of 0.05 is appropriate and commonly used in research studies. It indicates that differences in pesticide use revealed by statistical analysis cannot be applied to the whole population of interest in 5 out of 100 cases. Further significance levels used in research studies are 0.1 (accepting a greater risk of error) and 0.01 (accepting less risk of error). In order to calculate a suitable sample size the significance level needs to be considered by utilizing the critical value of the t-distribution for the significance level selected. Please refer to step 4 for an explanation on how the critical value of the t-distribution is used to assess whether differences in pesticide use revealed by statistical analysis can be applied to the whole population interest. Tables of critical values of the t-distribution can be found in any book on statistical analysis or on the Internet (e.g. http://en.wikipedia.org/wiki/T-table). The critical value of the t-distribution for a significance level of 0.05 is 1.96 for sample sizes above 120, which can be generally expected in research studies.

Finally, the standard deviation of the variable that you are interested in needs to be estimated. The standard deviation indicates how much variation there is away from the mean value of a set of values. Two samples with the same average pesticide use of 15 kilogram per hectare can have different standard deviations. A sample where the pesticide use of farmers varies only between 14 and 16 kilogram per hectare has a smaller standard deviation than a sample where most farmers have pesticide use levels of less than 14 or more than 16 kilogram per hectare.

Thus, a small standard deviation indicates that there is relatively little variation away from the mean value of a set of values. The standard deviation can be computed by taking the square root of the variance in the variable of interest. Conducting a pilot study or using data from previous studies of the same or a similar population can obtain an estimate of the variance. Data from previous studies, for instance, could indicate a standard deviation of 4.5.

The required sample size for the study can then be calculated by applying the following formula:

$$n_0 = \frac{t^2 \times s^2}{d^2} = \frac{1.96^2 \times 4.5^2}{0.45^2} = 384$$

where n_0 = required sample size, t = t-value for the selected significance level, s = standard deviation and d = acceptable margin of error for the expected mean.

If the required sample size exceeds 5% of the whole population of interest the sample size needs to be corrected by using the following formula:

$$n_1 = \frac{n_0}{(1 + n_0 / Population)} = \frac{384}{(1 + 384/1600)} = 310$$

where n_1.= final required sample size after correction, and population size = 1600.

Furthermore, it should be taken into account that for every independent variable there should be at least 10 observations when regression methods are used in the study. This means that if a researcher wishes to use 14 variables in a regression analysis, the sample size must be at least 140. The more independent variables are used the larger the sample size must be. However, increasing the sample size above the level indicated by the above formulas raises the risk of errors in the analysis. Therefore, it is recommendable, for instance, not to use more than 11 independent variables in a multiple regression analysis if the above formulas suggest a sample size of 111. In general, only as many variables as necessary should be included in a regression model. Please refer to step 4 for more information on regression methods.

However, in practice information on the variance is often not available. Moreover, costs, logistics, availability of staff and the willingness of organisations or communities to get involved can be restricting factors. For instance, there be only sufficient staff available to properly manage a sample of 300 farmers, although the sample size formulas suggest a sample size of 400. In such a case practical concerns should be taken into consideration when determining the size of the sample. It is better to have a small, properly managed and carefully analysed sample than a large sample, which is poorly supervised and never fully analysed for lack of time. No "magic ideal sample size" exists, which can be applied if the calculation of the sample size is not possible due to a lack of information on the variance, but generally 300 to 500 farmers are required including exposed farmers and the control group. This could be less if no group of exposed farmers is included in the model. The share of the control group on total sample size ideally should equal the respective share of farmers participating in the FFS. If the data is analysed by socio-economic factors, such as gender, the sample size needs to be larger in order to have a sufficient sample size for each socio-economic group.

Preparation of the survey includes determining the scope of information to be collected. General socio-economic data should include key parameters, which show whether the sample is representative of the population of interest and whether FFS participants, exposed farmers and control farmers are similar in their socio-economic characteristics. This is important in order to avoid that the difference-in-differences between farmer groups is due to factors other than FFS training.

In order to define the key socio-economic parameters the basic questions to be asked need to be identified. The incorporation of unnecessary parameters should be avoided since it will complicate further analysis. Table II gives examples for key socio-economic parameters and shows how these parameters provide information on factors, which are important with regard to the representativeness of the sample.

Parameter	Why is it important?
Farm size	The ownership of farm land is an indicator for wealth and thus for the ability of farmers to invest in new production technologies.
Access to markets	Farmers with good access to input markets have better opportunities to purchase high quality crop varieties and production inputs.
Education/ Experience	Well-educated and experienced farmers are likely to achieve better production results than poorly educated farmers or those who have little experience in growing their crop.
Access to similar projects	Farmers with access to similar projects (on irrigation, extension, rural credit, etc.) are likely to achieve better production results than other farmers.

Table II: Examples for key socio-economic parameters

Step 3: Conduct baseline and impact survey

Two surveys are required in order to generate a data base that contains the data needed to estimate the difference between success indicators before and after the FFS training for both FFS participants and non-participants. A baseline survey needs to be conducted to obtain data about success indicators before the FFS training. It should be carried out shortly before the start of the FFS training. Data about success indicators after the FFS training need to be acquired by conducting an impact survey during the year after the FFS training. The methods available for data collection are the same for both surveys.

Data collection can be conducted applying different techniques. Farmers can be surveyed by means of recall surveys. In this case a questionnaire, which is prepared prior to the data collection, determines the course of a standardized interview. See table III for a possible outline of a respective questionnaire through which data for a "double delta" analysis should be surveyed. It is important to focus on essential questions, which really serve a purpose.

Section	Questions on...
Demographics	Respondents name, gender, age, household size, number of full-time labourers, landholdings,…
Production system	Main crops, plot size, main problems, production inputs such as pesticides and fertilizer (amount and costs), pest control measures other than pesticide use, main pests and diseases, output, poisoning symptoms from pesticide use,…
FFS training experience (for FFS participants only)	Members of households participating in FFS, target crop and facilitator of FFS
Knowledge on FFS (for non-participants only)	Knowledge on FFS that a non-participant might have obtained from other sources, knowledge on related farming practices,...

Table III: Outline of a questionnaire on the impact of FFS in IPM

The demographics section includes questions on indicators through which the representativeness of the survey should be tested. Data for the "double delta" model should be collected in the production system section. The last two sections focus on indirect additional impacts of FFS training, which cannot be measured directly in the "double delta" model.

The interviews should be carried out by enumerators who are to be hired by the external institution that is in charge of conducting the survey. Enumerators must be capable of carrying out the interviews properly; otherwise results can be incomplete and imprecise. Personal interviews should be held with applicants for enumerator jobs in order to verify their qualifications. The socio-economic profile of enumerators can be important since their gender, for example, might influence responses. This however depends on the country and region where the survey will be conducted. It is important to find out which kind of person is a suitable enumerator in the country and region of interest and then to hire persons of that type. In general, it is recommendable to hire enumerators from the region where the survey will take place since they know best the area and how to deal with the local farmers. Prior to the actual survey, test interviews should be conducted in order to train and test the enumerators and to find out how to further improve the questionnaire.

Although recall surveys are widely used in impact assessment studies, they have some disadvantages. Information on yield and inputs can be biased because it is difficult for farmers to precisely remember details, such as the amount and product names of pesticides they applied on their plots during the season. Participatory monitoring and self-evaluation methods can be used in order to avoid such biases. Handing out monitoring sheets just before the season, would allow farmers to record all their crop production activities (e.g. pesticide use, seed varieties, irrigation, yield) directly after the fieldwork. Farmers are then visited regularly (e.g. every second week) by enumerators in order to review the completed monitoring sheets together with the farmers and to check these for consistency and completeness. Any omissions or errors can then still be corrected. The enumerators then take the forms with them for processing. A preparatory meeting before the season is required to train farmers on how to fill in the monitoring sheets. Table IV shows a sample for a monitoring sheet on pesticide use. If a recall survey is conducted, the form should be completed as soon as possible after the season to be surveyed. Doing a recall survey 3 seasons after the season of interest is definitely too late as many farmers will no longer accurately remember their pesticide usage. In such a case, it is better to start a new study with a proper baseline.

Table IV: Monitoring sheet on pesticide use

Cropname: ....

Date for each application	Trade name of pesticide	Common name (active ingredient)	Target pest or pests	Area treated (unit of area)*	Labour input (person hours)	Pesticide amount (kg or litres)	Pesticide costs (cost per unit of weight)*

*Units are to be defined depending on what farmers are familiar with.

If desirable, additional data can be obtained from experiments that can be carried out in research stations or farmers' fields. Experiments in research stations, for instance could indicate the yield effect and allow conclusions on costs and returns of pre-designed pest control measures. They aim at determining thresholds for the economically optimal use of input factors. However, data obtained from such experiments would need to be used with great care because of the difference between the conditions met in experiment stations and those in farmers' fields. Finally, ecological, bio-economic and economic models can generate information but are rarely available in practice.

Step 4: Analyse the data

The "double delta" approach can be applied using both linear and regression methods. A simple linear approach is to take the mean value of each group's success indicator before and after FFS implementation and then compare the differences in means between the groups as shown in table V.

	Mean		Change between 2004 and 2005		
	2003	2005	Absolute difference	Change in %	t-value
FFS participants	13.6	9.9	-3.7	-27	2.7*
Exposed farmers	13.4	12.1	-1.3	-10	2.1*
Control group	13.7	14.2	0.5	2	0.7
Note: * = Significance at 5% level; sample size N = 300					

Table V: Change in pesticide use (kg per ha)

The above fictive example indicates that the average pesticide use of FFS participants in the sample for 2005 (after FFS training) decreased by 27% compared to 2003 (before FFS training). Exposed farmers accounted for a decrease of 10% while the average pesticide use of the control group increased by 2%.

Since the sample comprises only a fraction of the whole population of interest, a t-test can be conducted in order to assess how strong the sample evidence is. This is of great importance because although the above results indicate that pesticide use by FFS participants in the sample after the training was much lower compared to before, this does not necessarily mean that this applies to all FFS participants in the study area. By computing the t-value we can assess whether, given a certain significance level, the calculated difference in means applies to the whole population of interest.

Box: Computation of t-values

T-values can be computed by applying the following formula:

$$t = \frac{D}{\sqrt{\frac{s^2}{N}}}$$

t = t-value, D = difference in means at the two points of time, s = standard deviation of the compared differences of the observed farmers and N = number of observed farmers.

Any standard econometrics software can calculate t-values, requiring only a dataset of the particular observations. Table VI shows how the necessary data can be tabulated in a worksheet file. Using an example of pesticide use in kg per ha, data on pesticide use before and after the FFS training are entered by each individual farmer. Dummy variables indicate whether a farmer belong to FFS participants, exposed farmers or the control group. In a dummy variable which, for example, indicates whether a farmer participated in an FFS program, a farmer is given a value of 1 if he participated and a value of 0 if he did not participate.

Farmer ID	Pesticide use before	Pesticide use after	FFS participant	Exposed farmer	Control group
18	13.6.	9.5	1	0	0
19	13.2	10.2	1	0	0
27	13.9	10.0	1	0	0
28	13.8	9.8	1	0	0
39	13.4	10.0	1	0	0
41	13.2	12.5	0	1	0
43	13.4	11.9	0	1	0
49	13.9	12.1	0	1	0
50	13.0	14.2	0	0	1
51	13.5	14.1	0	0	1
56	13.3	14.3	0	0	1
.	.	.	.	.	.

Table VI: Example of data preparation for a double delta linear model on pesticide use
(Source: modified from Walter Echols and Praneetvatakul 2008)

The calculated t-value then needs to be compared to the critical value of the t-distribution that applies to the desired significance level and the sample size. As explained in step 1, the critical value of the t-distribution for a significance level of 0.05 is 1.96 for sample sizes above 120. If the computed t-value is higher or equal to 1.96 the calculated difference in means can be applied to the whole population of interest. Variables with t-values higher or equal to 1.96 are marked with a star in table VIII. The significance level of 5% in the above sample indicates that the difference in means of FFS participants and exposed farmers applies to the whole population of interest in 95 out of 100 cases. Generally, a significance level of 5% is considered adequate although significance levels of 1% or 10% are used in some studies.

By employing a regression framework the change in success indicators from before to after FFS training is measured as a growth process, which depends on various factors. Most econometrics packages can perform a "double delta" regression analysis requiring only a dataset of the particular observations of the variables to be included in the model (see Table VII).

Farmer ID	Pesticide use before	Pesticide use after	FFS participant	Exposed farmer	Product price before	Product price after	Pest pressure
18	13.6.	9.5	1	0	5.0	5.5	0
19	13.2	10.2	1	0	5.1	5.3	1
27	13.9	10.0	1	0	5.1	5.4	0
28	13.8	9.8	1	0	5.0	5.4	0
39	13.4	10.0	1	0	5.1	5.3	0
41	13.2	12.5	0	1	5.1	5.3	1
43	13.4	11.9	0	1	5.1	5.5	1
49	13.9	12.1	0	1	5.0	5.4	0
50	13.0	14.2	0	0	5.1	5.3	0
51	13.5	14.1	0	0	5.0	5.4	0
56	13.3	14.8	0	0	5.1	5.3	1
.	.	.	.	.			.

Table VII: Example of data preparation for a double delta regression model on pesticide use
(Source: modified from Walter Echols and Praneetvatakul 2008)

Before converting the data file into regression software, e.g. SPSS, STATA, LIMDEP, further columns need to be added which contain the differences between before- and after-variables. Given the example in Table VI, the difference in pesticide use before and after the FFS training then needs to be selected as dependent variable in the regression software. The dummy variables for FFS participants, exposed farmers and pest pressure, and the difference in product price before and after the FFS training need to be selected as independent variables. Finally, the regression can be run by entering the appropriate command for the regression software that is used.

The price of the pesticides, for instance, is an important factor affecting the use of pesticides. The price of the pesticides will generally differ slightly within each of the three farmer groups. A regression model values the effect of different factors, including FFS training and, for instance, the price of the pesticides, on pesticide use by computing a coefficient for each of these factors. A positive coefficient indicates that the respective factor had an increasing effect on pesticide use and vice versa. Table VIII shows a typical output of econometrics software, which gives the outcome of a "double delta" regression analysis. The negative coefficients of "Dummy for FFS participants" and "Dummy for exposed farmers" in the example indicate that both FFS participation and exposure had a decreasing effect on the growth rate of farmers' pesticide use.

In order to assess whether the results apply to the whole population of interest, t-values are computed by dividing the coefficient by the related standard error. In the example, t-values show that, given a significance level of 5%, only FFS participation had a reducing effect on pesticide use in the whole population of interest. Sample evidence regarding the effect of FFS exposure on pesticide use is not strong enough to draw respective conclusions on the whole population of interest. Since a regression model takes into account the effect of factors other than FFS training on the success indicators it is more rigorous than linear methods. Note that "Dummy" in table VIII means that a dummy variable was used to distinguish between, for instance, farmers participating in the training and other farmers. Using a dummy variable all participating farmers were marked with a "1" in the dataset and all non-participants were marked with a "0".

	Change in pesticide use		
	Coefficient	Standard error	t-value
Intercept	0.48	0.24	2.0*
Dummy for FFS participants	-0.25	0.12	2.1*
Dummy for exposed farmers	-0.11	0.08	1.4
Product price	-0.32	0.46	0.7
Pest pressure	0.45	0.19	2.4*

Sample size N = 300
Notes: * = Significance at 5% level

Table VIII: Change in pesticide use, regression model

Step 5: Compute economic performance indicators

Taking into account the results of the "double delta" calculation and following the methodology of cost-benefit analysis (CBA), economic performance indicators could be computed to assess whether the FFS project achieves a satisfactory rate of return for the investment made. This would allow funding institutions to compare the FFS project's rate of return to those of other projects. In addition, farm-level CBA can be applied in order to assess the project's rate of return for single farmers.

Although CBA could be an interesting tool in impact assessment of FFS on IPM, caution should be exercised when using indicators such as the Internal Return Rate (IRR) for impact assessment of FFS. These are difficult to calculate because the result depends on several assumptions that create a degree of uncertainty. Furthermore a single indicator cannot adequately reveal the complex impacts of FFS training, which always involves effects in a range of economic and social areas. See Review Document: IPM Farmer Field Schools: A synthesis of 25 impact evaluations.

The basic procedure of CBA is to aggregate farm level benefits and value them at economic prices, calculate net benefits by taking into consideration the costs of the program and, finally, to compute the indicators on the economic performance as shown in table IX.

If, for example, the "double delta" analysis revealed that FFS participation had a positive effect on yield, the related economic productivity increase and thus the added benefits can be calculated by multiplying the increased output with the respective output price of the commodity. Assumptions then need to be made about the impact of FFS training on crop yields, external input use and health status due to reduced pesticide use in the years after the evaluation. Different scenarios (optimistic, in-between, pessimistic) should be used. Costs of the FFS program include general program costs, costs for the training-of-trainers course, training costs and possibly opportunity costs of time of farmers.

Box: Calculation of costs

General program costs may include:
- Salaries of project staff, foreign and national consultants
- Overhead costs required to provide the program infrastructure
- Transport of program staff and materials
- Production of training materials, including booklets, forms, etc.
- Specific studies in direct support of the programme
- Depending on the situation, cost of counterpart specialists, extension workers[1] and the organisational infrastructure for the fieldwork
- Monitoring and evaluation
- Reporting

Training-of-trainers (TOT) costs include:
- Costs of the training of specialists and field staff

FFS training costs may include:
- Salaries and transportation costs of trainers
- Transport
- Materials, including teaching materials and inputs for demonstration plots.
- Food for the participants during the training sessions, if provided.

[1] Sometimes extension workers already exist before the start of an FFS program. FFS then aim to help them to become more effective. In such a case the difference in cost of extension staff between before and after the implementation of an FFS program need to be included in the calculation of costs of the FFS program.

Year	0	1	2	3	4-10
Benefits					
Productivity increase (economic value)		250,000	500,000	900,000	1,000,000
Health benefit		20,000	50,000	90,000	100,000
Costs					
Program costs, general	800,000	800,000	800,000	800,000	50,000
TOT course	800,000				
FFS training (100 MU per farmer)		150,000	170,000	130,000	
Opportunity cost of time of farmers		100,000	90,000	80,000	
Added benefits		270,000	550,000	990,000	1,100,000
Added costs	1,600,000	1,050,000	1,050,000	1,010,000	50,000
Added net benefits	-1,600,000	-780,000	-500,000	-20,000	1,050,000
NPV (10%)	1,103,261				
IRR	16%				
BCR	1.27				

Table IX: Example for an economic evaluation of a fictive three-year IPM-FFS project (US$)

Opportunity costs of time are income that farmers forgo when attending FFS training instead of working on their fields or off-farm. In case a farmer would not do anything productive instead of attending the training, however, opportunity costs of time do not accrue. The same applies if productive activities are undertaken at a different time instead, and do not go at the expense of, for instance, weekly productivity.

The net benefits of the program can be used to compute the net present value (NPV), internal rate of return (IRR) and/or benefit cost ratio (BCR) of the project. NPV and IRR provide an indication of the economics of the project. Because of methodological inconsistencies the BCR is less often used.

Assume that the example in table IX has been calculated applying an optimistic scenario for the years following the FFS training. The IRR of 16% means that even if the money spent on the project has to be borrowed at 16% investing in the program is still economically meaningful. Applying less optimistic scenarios could result in lower IRRs. In general, an IPM FFS program in a developing country would be expected to yield a minimum rate of return of 8 to 10%. Only strong indicators for additional benefits, such as positive effects on the environment, which could not be economically valued, can justify a lower rate of return.

Once again it needs to be stressed that caution should be exercised when using such indicators for impact assessment of FFS. Since results depend on several assumptions that create a degree of uncertainty, indicators such as the IRR are difficult to calculate. Furthermore a single indicator cannot adequately reveal the complex impacts of FFS training. Indirect benefits related to health, environment, social structure, empowerment of women, general usefulness of research, etc. are difficult to measure by using a single indicator.

Step 6: Draw lessons

After analysing the data and computing economic performance indicators lessons should be drawn from the results. With the help of the calculated success indicators it can be concluded which objectives of the FFS training have been achieved and which have not been achieved. The next step is then to find out why things worked or did not work. It therefore needs to be analysed what internal and external factors, capacities, and decisions influenced or constrained outcomes and impact. Such findings can finally be used to adjust the programme and/or to improve the organisation or curriculum of the training. They could also contribute to further decision-making regarding the continuation or scaling up of a program.

7. What can go wrong?

There are several pitfalls that can jeopardise the impact assessment. Below is a list based on lessons learned from past studies. It is important to be aware of such potential pitfalls throughout the procedure of impact assessment in order to avoid incomplete or biased results.

- An impact assessment study should either focus on self-evaluation by farmers, self-evaluation by a project or external evaluation since all levels require different methodological approaches and are both time and money consuming. Problems can arise in trying to carry out an impact assessment study at all three levels at which evaluation can take place at once.

- Incorporation of unnecessary parameters in the impact assessment study should be avoided since it will further complicate the analysis. It makes no sense to have a very broad dataset without having the financial and human resources to analyse it.

- The control group could be severely diminished during the course of the study if the geographical coverage of the FFS program expands to control villages.

- Farmers, who were surveyed before the start of the FFS, may dropout of the FFS. As a result, there can be an insufficient number of trained farmers for a meaningful before and after comparison.

- Before and after comparison can not be made if a farmer changes crop. Impact assessment should therefore focus on stable key crops and not attempt to include too many different crops. This is point particularly applies to vegetable growers.

- The selection of farmers for the impact assessment study will be biased if only those farmers are selected who are the most promising with regards to the successful application of IPM practices.

- The objectives of the study may not be completely clear for the external institution (e.g. consultants) in charge of the data collection. It is important that the external institution that will conduct the data collection knows why the various data are being collected and how they will be used. Only then will it be able to give clear instructions to its enumerators.

- The quality of enumerators could be insufficient due to improper training or selection by the respective external institution. Differences in the style of interviewing between enumerators can occur and enumerators can change between pre- and post surveys. Enumerators can have a lack of motivation to do their work properly because they are not well paid. Hence collected data can be incomplete and imprecise. Information on indicators such as yield, which is required to compute gross margins, can be missing. "Zero" values can misleadingly be marked as missing values. The conversion of local measurement units to the standard units used in the entry form can cause problems. Keep in mind that serious monitoring takes time and requires expertise. In case a data set is found to be incomplete or imprecise a "clinic session" can be held with the external institution, which conducted the survey. The aim of such a session is to identify and discuss problems in the data set that compromise impact assessment analysis and jointly find solutions to improve the quality of the data set.

- Results could become biased if FFS program staff are involved in the impact assessment survey since they have an interest in results that shed a positive light on their work.

- The questionnaire used in the survey may be too long and farmers can be unwilling to spend a large amount of time for the interview. Thus, answers to questions at the end of the interview can be biased because farmers are not concentrated anymore.

- Farmers may be unwilling to provide information.

- Farmers' awareness of study objectives might introduce a bias in direct reporting (participatory approach). If farmers, for instance, know that one of the objectives of the training is to reduce pesticide use, they might understate their pesticide use in order to please project facilitators.

- Evaluation results could be biased due to the "pampering" effect. Farmers participating in FFS training receive, to some degree, exceptional attention of program facilitators, journalists or expatriate scientists. Thus positive training effects are aggravated during the time of the program. After the FFS program, in contrast, political interest is likely to stop. Furthermore old habits, like applying great amounts of pesticides (which could be promoted by pesticide salesmen), might be taken up again by farmers. The socio-economic and technical context can also change.

- Loose monitoring or questionnaire sheets can be lost. It is important to make sure that sheets are complete by either tacking them or using an electronic version on a notebook.

- The time given for data entering may be too short. Hence mistakes are more likely.

- Excel sheets for data entering might be too complex. Keep them as simple and easy-to-understand as possible.

8. Summary

The purpose of impact assessment of FFS on IPM is to inform different interest groups about whether a FFS program has brought about the desired effects and whether those effects can be attributed to the program. Assessing the impact of FFS on IPM is difficult because of the abundance of impact parameters and because evaluation can take place at different levels. Previous impact assessment studies were designed to be either statistically rigorous or comprehensive, but never both, which made it difficult to compare results.

The "double delta" approach can be used to set up a statistically rigorous study, which is scientifically sound and qualified, thus yielding valid, meaningful and comparable results. Its basic idea is to model the effect of FFS training by estimating the difference between success indicators before and after the training for both FFS participants and non-participants and then comparing the difference between the two groups.

The procedure of impact assessment of FFS on IPM can be divided into 6 steps. In step 1, meaningful success indicators are identified, which relate to the objectives of the FFS programme and are suitable to measure whether the program has achieved its objectives.

The design of the impact assessment program is the second step. The groups of farmers to be surveyed (FFS participants, exposed farmers, control group), sample size and key socio-economic data to be collected are determined.

In step 3, data is collected by means of standardized interviews, participatory monitoring and self-evaluation methods, or experiments that can be carried out in research stations or farmers' fields.

Step 4 is to analyse the collected data by applying the "double delta" approach, which can be done using both linear and regression methods. Economic performance indicators (net present value, internal rate of return, benefit cost ratio) are computed in step 5. However, caution must be taken when using economic performance indicators for impact assessment of FFS on IPM as there always are several impact areas and results depend on several assumptions.

Finally, the outcome of the analysis can be used to identify which objectives of the FFS program have been achieved and which have not been achieved. Moreover the reasons why things worked or did not work can be found. Decisions on the continuation or scaling-up of a program or on how to improve the organisation or curriculum of the training can then be made based on these findings. Throughout the procedure of FFS impact assessment several pitfalls can occur, which need to be taken into account in order to avoid incomplete or biased results.