CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH

SCIENCE COUNCIL

Donor Demands and Uses for Evidence of Research Impact – the Case of the Consultative Group on International Agricultural Research (CGIAR)

SCIENCE COUNCIL SECRETARIAT

OCTOBER 2005

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH

SCIENCE COUNCIL

Donor Demands and Uses for Evidence of Research Impact – the Case of the Consultative Group on International Agricultural Research (CGIAR)

OCTOBER 2005

FOREWORD BY THE CGIAR SCIENCE COUNCIL, STANDING PANEL ON IMPACT ASSESSMENT

In May 2000, the Standing Panel on Impact Assessment (SPIA) of the Science Council held a workshop on the future of impact assessment in the CGIAR.[1] All 16 International Agricultural Research Centres (IARCs) were represented, along with a number of key donor representatives and outside experts. There was much discussion of donor needs for impact-related information in the context of what is technically feasible and cost effective. Several of the donor participants suggested that it might be useful to ascertain, in a more systematic manner than what had been done previously, what donors actually require in the way of impact information from CGIAR System. Thus the idea for this assessment of donor needs and wants in the area of impact assessment was born. These are referred to in the title to this report as 'donor demands'.

It was a few years later that SPIA asked David Raitzer to initiate a systematic email survey of donors, trying to ascertain their impact assessment needs and uses. The results of that survey revealed a number of interesting results, but also showed some inconsistencies in results for different questions and raised a few new issues as well (see section 3.1.2). Partly as a response to those inconsistencies and partly to obtain some more in depth understanding of the donors on impact assessment needs, SPIA initiated a second-round of information gathering from Members, this time through telephone interviews. Klaus Winkel, former Danish representative to the CGIAR, was asked by SPIA to conduct the telephone interviews. A final step in the process of ascertaining insights on donor needs was a luncheon session at the 2004 Annual General Meeting of the CGIAR held in Mexico, kindly sponsored by the UK Department for International Development and chaired by Paul Harding, Executive Secretary of the European Initiative on Agricultural Research for Development. The purpose of the session was three-fold: (i) to inform interested donors about the main findings of the survey and interviews; (ii) to clarify some key remaining uncertainties, in particular, some discrepancies arising between the email survey and telephone interview results, and (iii) to investigate elements not fully explored in the email survey and interview process.

The present document synthesizes and interprets all the results of the survey, interviews and AGM04 luncheon meeting discussion. SPIA thanks the authors for a thorough and concise analysis of all the data. We would welcome further comments from donors/investors, centres and other interested stakeholders in the CGIAR on the subject of donor needs discussed in the paper. This study represents a beginning rather than an end to dialogue with the donor community. SPIA intends to follow up on this exercise with regular efforts to ensure that the SPIA agenda is relevant to donor demands.

One limitation is worth noting up front. All Members were asked to participate in all three information gathering activities. However, responses from Members from the South were modest, and thus they are underrepresented in the results. At the same time, donors

[1] TAC Secretariat. 2001. The future of impact assessment in the CGIAR: Needs, constraints and options. Proceedings of a workshop organized by the Standing Panel on Impact Assessment of the Technical Advisory Committee, 3-5 May 2000, FAO, Rome. Rome: FAO of the United Nations.

representing more than eighty percent of the funding of the CGIAR responded in the interviews and almost 70 percent in the email survey. Thus, the donors who were included make many allocation decisions that are very important for the System.

The authors find a strong demand among donor representatives for evidence of the impacts from the IARCs. Yet, the specific ways in which such results are applied in decision making remain highly variable and, in some cases, not very clear, due to the indirect nature of the influence of evaluation results on decisions. Some of the key highlights resulting from this study include the following:

- In addition to evidence derived from ex post impact assessment (epIA) studies, many other factors and events influence Member decisions, including: political priorities of the Member country, perceptions of scientific quality and desire for funding continuity.
- EpIAs are, nevertheless an important source of information for Members, particularly in terms of maintaining a higher level of confidence in the CGIAR system and defending overall budget decisions, as distinct from helping them make the decisions.
- Members want short, concise write-ups of any kind of information they use, including from epIAs.
- With respect to epIAs, Members stressed the need for: (i) greater clarity and transparency in the studies, (ii) broader coverage in research domains (moving beyond crop germplasm improvement benefits), (iii) greater focus on mission level impacts on poverty and distribution of benefits from research. Some Members felt that epIAs should also stress more the science being assessed.
- EPMRs are an important conduit for epIA results. This means that epIA results need to be embedded as effectively as possible in EPMRs.
- There is a range of views about the value and inherent limitations of economic approaches being used in epIA. While two thirds of the members interviewed thought that the economic metric added to the utility of the impact studies read, others were more sceptical about the accuracy of past economic studies.
- Other technical criticisms of past epIAs were mentioned by some Members -- these factors will all be considered as SPIA and the centres move on to improve epIA practice and outputs in the System.

These results have implications in terms of the future of epIA in the CGIAR System. The authors have identified a number of ways in which donor needs may be better satisfied. SPIA intends to follow up on these results in several ways:

- First, in keeping with the strong, widespread feeling of a majority of donors, SPIA, in the future, will be producing short impact briefs for all of its studies. We intend to go back and also do so for past studies.
- Second, we will be expanding coverage of impact assessments to areas of CGIAR research that have not yet been adequately assessed. In fact, assessments are already ongoing related to NRM research and policy research.
- Third, SPIA will be exploring ways of increasing the transparency of impact assessments and increasing the understanding of them and, hopefully, increasing the credibility of such assessments.

- Among other things, SPIA intends to produce occasional *Impact Notes* that explain ways in which impacts are tracked and assessed, raise issues related to best practices including current limitations of impact assessment, and note any on-going progress towards overcoming such limitations.

SPIA hopes that, over time, the results of this study and the continuing quest to understand better impact assessment needs of members will lead to other changes in the ways in which impact assessments are carried out by SPIA and the centres and the ways in which their results are interpreted and presented.

Hans Gregersen, Chair
October, 2005

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH

SCIENCE COUNCIL

Donor Demands and Uses for Evidence of Research Impact – the Case of the Consultative Group on International Agricultural Research (CGIAR)

David A. Raitzer and Klaus Winkel

Prepared on Behalf of the
CGIAR Standing Panel on Impact Assessment

ACKNOWLEDGEMENTS

This paper reflects the important contributions of many more individuals than are listed as authors. Hans Gregersen, Chair of the CGIAR Standing Panel on Impact Assessment, provided guidance and feedback on essential issues of the methodological approach. Timothy Kelley, SPIA Secretary, made critical contributions to study conduct, through the provision of insights on methods and analysis, as well as survey and interview facilitation. SPIA's generous support for the analysis must also be acknowledged. Mike Spilsbury offered valuable comments on an early version of the email survey instrument. Jim Ryan provided useful remarks on a draft of this paper. The UK Department for International Development (DFID) and Jonathan Wadsworth deserve thanks for sponsoring a luncheon session at the 2004 CGIAR Annual General Meeting, at which survey/interview results were validated. EIARD's (Paul Harding) Chairmanship of this session is also appreciated. The respondents to the email survey and telephone interviews, as well as participants at the lunch session are thanked for generously donating their time and insights to this study. Without the help of these individuals, there would be little material for the present investigation to analyse.

CONTENTS

FIGURES

TABLES

SUMMARY

In an era of increasing demands for accountability, performance measurement, and results-based management, CGIAR donors repeatedly articulate demands for documented evidence that research investments are helping the poor. In response, the institutions of the CGIAR System have produced many *ex post* impact assessments (epIAs), so as to attribute and quantify benefits from research activities. But, is this information actually used in donor decisions?

A number of observers have noticed that there is little apparent relationship between impact assessment findings and the subsequent allocation patterns of donors. In fact, it has been repeatedly observed that those areas of research with the highest levels of assessed benefits often suffer from declining funding, while unproven areas of research and non-research investment receive rising funding shares. This dichotomy poses a quandary for the role of *ex post* impact assessment, which should indirectly affect funding levels by demonstrating the benefits from prior investment.

This study explores how epIA results inform donor decisions, as well as how epIA can more effectively help to influence donor behaviour. The study is based on two methods. First, an email survey was sent to representatives of all CGIAR Member (donor) agencies. This survey queried preferences for different types of impact metrics and assessment methods, and explored the relative influence of epIA studies among other information sources. Second, a series of follow-up interviews was used to further investigate how donors determine absolute and relative allocation levels for research centres and topics.

The email survey's 24 responses received from 22 Member agencies indicate that epIA results are influential and demanded. In fact, epIAs are listed (along with EPMRs) as the most important information source for funding decisions. Similarly, the metrics that are most poverty-related and squarely within the category of impact assessment (rather than process evaluation) receive the highest ratings. In addition, "magnitude and distribution of benefits" is selected as the most important determinant of use. Readership of *ex post* impact assessments was also reported as being higher on average than for other forms of research evaluation.

However, the 26 interviews of Member representatives reveal somewhat different patterns. Funding decisions are described primarily as complex, with a high degree of influence by political factors and higher bureaucratic levels. A variety of considerations are embedded in each decision, with factors, such as alignment with organizational aid strategies or continuity playing more important roles than perceptions of impact potential. Furthermore, perceived impact potential is only partially informed by perceptions of past impact. Moreover, epIA is not reported to be the most important source of information regarding perceived past impacts.

Yet, the two sets of findings are not necessarily in conflict, when it is considered that insights from impact assessment studies primarily influence donor decisions indirectly. Many prior studies of "evaluation use" have found that "conceptual use" or indirect influence is much more prevalent than "instrumental" application of findings, so as to shift decisions directly.

In the present analysis, similar patterns seem evident. For example, the primary information sources cited for perceptions of past impact, such as External Programme and Management Reviews, embed epIA findings. Similarly, the areas of research that are most frequently perceived to have generated past impact are those areas for which impact assessment has documented benefits. Thus, it seems that substantial indirect influence is likely.

Furthermore, there are also a number of indications that there is some direct influence from impact assessment. Nearly half of the interviewees and survey respondents indicate direct use of epIA findings to substantiate or inform allocation decisions. A similar proportion could also identify specific epIA studies.

The interviews further reveal that CGIAR representatives have limited time to consider impact materials, and that decision makers at higher levels in donor bodies have even less time to read study results. As a result, there was widespread demand for brief summaries of impact assessment results, so as to permit faster consideration of findings, and so as to help capture the attention of higher officials who are flooded with many forms of competing information.

In addition, although most respondents expressed strong demand for impact-related information, a significant minority also expressed scepticism or concern about the accuracy of impact assessment findings that they encountered. Certain analyses were perceived to provide too few details about the assessment methods and assessed research to permit confidence. There were also widespread demands for coverage of a broader array of research activities with approaches that focus more on poverty-related metrics and the distribution of assessed benefits.

Clearly, *ex post* impact assessment has an important role to play as an accountability tool. Although epIA is not a direct driver of specific funding decisions, the confidence in CGIAR capacity that epIA findings impart appears to be of substantial importance for continued support to the system. However, although some specific areas for improvement have been identified, this study should serve as a beginning, rather than an end to exploration of donor demands for evidence of impact. Ongoing dialogue is needed to ensure that epIA studies meet demands as well as possible, and that key constraints to influence are effectively addressed.

1 INTRODUCTION

In an era of increased accountability and performance measurement, *ex post* impact assessment (epIA) is often promoted as a means of offering evidence to funding bodies that international agricultural research investments are effectively contributing to the achievement of development goals. This is the so-called accountability function of impact assessment. Implicit in this function is the assumption that continued investments will be affected by perceptions of the performance of prior investments. In addition to this indirect role in funding determination, it is also claimed that epIA can directly offer insights for allocation decisions (Impact Assessment and Evaluation Group, 1999).

Impact assessments to serve these purposes are often requested by donors. In response, impact assessors have produced a great deal of evidence that international agricultural research is an exceptionally efficient and effective form of investment. Yet, Alston et al. (2000) observe that:

> *In recent years, a paradox has become apparent. On the one hand, we have an ever expanding volume of what appears to be generally consistent evidence that rates of return to public agricultural R&D are high—high enough to justify past support and an even greater investment of public funds. On the other hand, we have seen in recent years in most countries, rich and poor alike, a marked slowdown, if not an actual decline, in public funding for agricultural R&D. Support for international research is dwindling despite seemingly strong evidence that it pays off handsomely.*

Thus, despite impressive indications of effectiveness, donor fatigue has plagued the international agricultural research system (Anderson, 1998). Furthermore, ironically, those research areas with greatest demonstrated impact have experienced funding declines, while certain areas with limited documented impact to date receive growing shares of research budgets (Lele, 2003). This may indicate that the evidence produced has not been entirely persuasive.

It is now commonly accepted that evaluations must be used in some way to be of value (Shulha and Cousins, 1997). In fact, an entire body of peer-reviewed literature has been dedicated to the issue of "evaluation use" by the broader evaluation community (outside of agricultural research). A plethora of process-models have been developed to describe pathways from evaluation planning and implementation to improved programmatic performance. Typologies of use have been developed, and models of interaction among internal and external environmental use-determining variables have been expounded (Johnson, 1998). However, little of this analysis has been performed empirically (Shulha and Cousins, 1997), and none has dealt with impact assessment of research.

In fact, impact assessment of agricultural research has been pursued with remarkably little reference to specific demands from any audience group, other than a generic call for more impact assessment. Informal dialogue has been used to identify topics for assessment, but no systematic assessment has been previously made of preferences and uses for impact assessment results. Volumes have been written on impact assessment methods with only

passing mention of audience interests (e.g. Alston et al., 1996). Although the need for a demand-driven approach has been articulated (Özgediz, 1995), little systematic action has been taken towards satisfying this need.

1.1 Impact assessment in the CGIAR

As a network of 15 publicly-funded autonomous International Agricultural Research Centres (IARCs), the Consultative Group on International Agricultural Research (CGIAR) is uniquely organized to develop enhanced technologies and policies to help solve agricultural production constraints in the developing world. This "System" has the following mission:

> *To achieve sustainable food security and reduce poverty in developing countries through scientific research and research-related activities in the fields of agriculture, forestry, fisheries, policy, and environment.*

Operating as a publicly-funded entity that produces international public goods, the CGIAR is not driven by market forces, as the primary support for the System comes from developed country governments that 'purchase' the System's services on behalf of the developing countries' poor. Thus, in the absence of market information, empirical analysis is needed to determine if funds invested in the CGIAR's agricultural research activities are effective in making progress towards the development goals outlined in the mission statement.

EpIA is intended to fulfil this function by providing analysis that is "intended to determine more broadly whether the program had the desired effects on individuals, households and institutions and whether those effects are attributable to the program intervention" (Baker, 2000). These "desired effects" are in terms of the CGIAR goals. It should be noted that epIA is only one component of a comprehensive evaluation package[2].

Accountability and resource mobilization have long been among the primary objectives of epIA. To convince a sceptical US Congress of the US Department of Agriculture's research benefits, one of the first examples of formal economic impact assessment of research was commissioned by "Tama Jim" Wilson nearly a century ago (Alston et al., 2000). Much more recently, in 1984, when the first impact assessment that covered the entire CGIAR System was conducted, these accountability functions were clear. At the International Centers Week of 1982 an impact study was first proposed by the Swedish delegation, so as to provide "...an important means of influencing agencies in the donor countries that supply funds for the CG System" (CGIAR Secretariat, 1982).

This intended influence on donors continued to be the dominant explicit justification for CGIAR impact assessment activities during the following decades. In 1994, the Public Awareness and Resource Mobilization Committee (PARC) of the CGIAR set the seed for what would form a permanent Systemwide impact assessment entity (PARC, 1994). The PARC was established as a response to the "funding crisis" of the early to mid 1990s, during which funding fell just after three new Centres had been established. Accountability and resource allocation remain among the stated primary objectives of epIA to date (SPIA, 2004).

---

[2] This includes *ex ante* assessment at the individual IARC and System level, inter-centre (stripe) and System-level reviews, external programme and management reviews, internally commissioned reviews and medium term plan assessments by the Science Council (see Science Council (2005) for more information about CGIAR review process).

The accountability objective specifies an intended pathway of influence for epIA. Accountability in this context is primarily to donor agencies and the constituencies that lend popular support to their investments. As a consequence of fulfilling accountability demands, funding agencies are assumed to utilise information on past impacts in allocation decisions regarding future research possibilities.

This implies that donor agencies comprise a primary audience for epIA reports. For impact assessment to become a demand-driven or "utlization-focussed" evaluation activity, the demands of this audience group must be understood. To appreciate the context for the stated preferences of this body, the manner in which impact assessment findings are applied should also be explored.

1.2 The key concepts of evaluation use and influence

Considerable attention has been given to 'evaluation utilisation', i.e., use of evaluation or assessment results, by the broader evaluation community. Much effort has been devoted to the establishment of typologies of use, and this has resulted in four generally accepted categories of utilisation: instrumental; symbolic; conceptual and process (Leviton and Hughes, 1981; Shulha and Cousins, 1997).

Instrumental Use
Instrumental use is the type of use that initially preoccupied the evaluation community. This term refers to direct and traceable application of evaluation findings, so as to alter programme strategy and/or implementation.

Symbolic use
Symbolic use refers to selective application of evaluation findings, so as to support a preordained policy preference (Leviton and Hughes, 1981). Hence, such use actually makes little or no difference to programmatic implementation, and only serves to make prior policy preferences appear as the result of rational deliberative processes.

Conceptual use
"Conceptual use" or "enlightenment" (Rich, 1977; c.f. Leviton and Huges, 1981) refers to a more complex conception of how information creeps into the actions of decision-makers. Under this concept, many different information sources compete for the attention of key audiences, who have limited time for consideration of the huge volumes of information that they receive. As a result there is limited potential for a single evaluation to be directly applied so as to make substantial and fundamental changes to programmatic direction. However, an evaluation can incrementally contribute to the body of knowledge surrounding a programme or policy. This improved general understanding facilitated by evaluation findings gradually informs the context under which programmatic decisions are taken. Such application is often diffuse and difficult to trace, since several or many steps occur in the transmission of evaluative insights to future decisions.

Process use
Process use is a more recent addition to the evaluation use lexicon, which refers to the effects of involvement in the process of evaluation conduct (Shula and Cousins, 1997). For such use to occur, decision-makers must be involved in the evaluation, so as to absorb any insights that

emerge as a result of the methods employed. Such insights are transmitted directly through participation, rather than through embodiment in a final report.

Use versus influence

Some of the evaluation literature suggests that the term "influence" may be more appropriate than "use," as this places more emphasis on the outcomes of evaluation utilisation, rather than on intermediate processes. This term also is intended to convey that "influence is exercised in more subtle ways that the word "utilization" – with its overtone of tools and implements – can capture" (Weiss, 1980). Usage of the term "influence" is related to concept of conceptual use, as both stem from recognition of the indirect pathways through which evaluation results percolate into the actions of decision-makers. As a result, in order to include important potential applications of evaluation findings, analysis must include indirect pathways of influence, in addition to more direct uses of evaluation results.

1.3 Objectives

To help bring a "demand-led" orientation to impact assessment in the CGIAR, the present study has queried the demands of a key accountability audience for epIA—the Member (donor) community of the System.

The key objectives of the present study are to determine the following:

What kinds of information regarding past research impact are most relevant to the demands of donor audiences in the context of development-oriented international agricultural research?

1. **How is information on past research impact applied in allocation decisions by donors?**
2. **What specific types of information regarding past impact achieved is demanded by donor audiences for accountability purposes?**
3. **How can methods for epIA studies be oriented to effectively produce the types of information demanded by donors?**

1.4 Organization of the study

The rest of the paper is organized as follows. The following methods section describes the two data gathering techniques employed in this study - 1) an email survey of donor preferences for epIA methods and approaches; and 2) a series of telephone interviews that explore how epIA results influence donor decisions. Next, the results section provides summary statistics of trends in responses. Finally, these results are discussed in light of evaluation use concepts, and specific recommendations for impact assessment in the CGIAR are identified.

2 METHODS

Demand for impact-related information has been queried through a survey, so as to ascertain preferences for different epIA approaches, methods, and metrics. In addition, this study has used a series of telephone interviews of CGIAR Member representatives to explore the context in which epIA results are applied and the pathways by which impact findings contribute to decisions within Member agencies.

2.1 Email survey

A survey questionnaire (Appendix I) was distributed under the auspices of the CGIAR Standing Panel on Impact Assessment (SPIA) to CGIAR Members in an effort to better understand their views about the major uses of and demand for epIA in the CGIAR. Non-Member donors were not included in the email survey.

This survey was intended to explore the following issues:

- Expectations for the role of epIA
- Satisfaction with epIA studies to date
- Use of epIA in allocation decisions
- Factors that facilitate epIA use
- Demanded metrics and methods
- Readership of epIA as compared with other forms of evaluation
- Views on the demonstrated impact from natural resource management research[3]

The approach for formulating the survey format and wording was consultative deliberations between the first author and the Chair and Secretary of SPIA.

The survey was sent to all official representatives of the 63 CGIAR Members (donor agencies) in May 2003. In addition, this was supplemented by a list of CGIAR Member staff who have participated in impact assessment meetings and workshops convened by the IAEG/SPIA. The later recipients were added because the "official" representatives to the CGIAR are often senior staff who have many responsibilities aside from dealing with the System. These representatives are often informed by less senior officials, who actually use and interpret impact assessment results more directly. The survey was prepared as a Microsoft Word form, and was emailed to 160 representatives of the 63 CGIAR Members. Two follow-up reminders were subsequently sent to elicit additional responses.

Responses were entered into Microsoft Excel worksheets, from which statistical analyses were performed using *Visual Stats* and *Statsgraphics Plus* statistical software. Details are provided in Appendix III. Means and medians were computed for all quantifiable questions. Subsequently, the Wilcoxon Signed Rank Test for Paired Differences was used to compare different categories, using the respondent as the treatment block. This test was applied iteratively with

[3] It has been noted that natural resource management research and its antecedents in the CGIAR has not been subject to many epIAs, and that little impact has been shown through the few studies conducted (Anderson, 1985; Gardner, 2003; Raitzer, 2003). Therefore, queries on this topic help to reveal whether donors perceive impact, even if none has been illustrated through epIA methods.

$\alpha = 0.05$, so as to determine which categories differed statistically from one another for those questions with quantifiable responses.

2.2 Telephone interviews

Semi-structured telephone interviews of representatives of CGIAR Member agencies were utilised to further explore ambiguities and uncertainties that remained after the email survey results were analyzed (Interview question are included in Appendix II). The focus of the interviews was to probe in greater detail how investors make allocation decisions, and how impact assessment results feed into these decision processes.

The format and structure of the interviews were based on preliminary analysis of the email survey results to identify topics that remained unresolved. From this analysis, a preliminary list of survey topics and questions was developed through consultations between the authors and SPIA members. Subsequently, the survey format was tested during two trial interviews, after which adjustments were made. Interviews were preceded by a notification email, with the list of interview questions attached. Respondents were asked to read through the questions before the interviews were conducted. Klaus Winkel, a former Danish representative to the CGIAR, conducted the interviews. He conducted the interviews alone, and recorded notes from responses during the interview process.

The email notification message was sent to representatives of all 63 Member agencies of the CGIAR and follow-up reminder messages were sent to all non-respondents, with a second reminder to donors that contributed the most funding to the CGIAR System. Interviews of 25 of these Members were ultimately arranged during late September and October of 2004. A written response was provided in lieu of an interview by an additional agency, due to communication difficulties, and one agency also supplemented verbal responses with a written summary of its answers to the interview questions. Respondents for 24 of the agencies were interviewed individually, while a group of three interviewees was queried for one other agency. The duration of the interviews ranged from 20 minutes to one hour.

The following areas were explored through the survey questions:

- Investment decisions and the role of impact findings
- Donor perceptions of impact and how they are informed
- Determinants of epIA readership
- Examples of use of epIA results
- Specific demands for impact-related information

The interviews began by discussing how resource allocation decisions are made for international agricultural research, as well as how specific organizations are selected for funding. The discussion then turned to the roles of different information sources, including epIA, in informing these decisions. Next, interviewees were asked about perceptions of CGIAR research that had resulted in the greatest levels of impact in the past, as well as those areas that have the greatest potential for future impact. The sources for these judgements were then probed. Subsequently, a couple of questions attempted to analyse assessment attributes that contributed to study influence. Respondents were asked to identify epIA traits that would be likely to allow the study to have high levels of influence in the Member agency, and queries were posed about the marginal value of additional epIA efforts. Interviewees were

asked to recall one of more specific epIA studies. Questions were then posed to explore how the results from the study had influenced various decisions, and recommendations were solicited regarding how these specific studies might be improved. The final section of the interviews attempted to identify specific demands for epIA topics and methods. First, demands for economic metrics of research impact, such as internal rates of return or benefit-cost measures, was queried. Subsequently, a final question asked whether, given increasing attribution difficulties as analysis progresses from uptake of a research output to impacts on poverty, the CGIAR's performance should primarily be assessed in terms of adoption, productivity or poverty measures.

To identify the prevalence of different answers to questions posed, "content analysis" was performed upon written summaries of telephone interview responses. According to Stemler (2001) "content analysis" refers to a "systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding." To perform content analysis, responses are translated into codes or consistent categories, which can be counted and compared in frequency. To do so requires identification of information to be gathered as a first step, followed by preliminary identification of the categories to which information will be classified. In the "emergent coding" approach that was employed in this analysis, the categories are revised after review of the data to be analysed, so that responses can be appropriately categorised. Keywords are developed for each category, and the prevalence of the content identified to be within each identified category is identified through queries of these keywords. However, when counts are produced, they can only be related to frequencies (as the data are categorical), which are difficult to meaningfully appraise with statistical techniques. Therefore, qualitative inference of the significance of differences is necessitated for the results derived via content analysis. For those interval and ordinal data that are possible to compare with statistical analysis, nonparametric techniques were applied. In the case of these data, the Wilocoxon Signed-Rank test was applied in a similar manner to analysis of the email survey results (see Appendix III).

A supplementary source of information to inform the present study was a discussion with donors and other stakeholders during a luncheon session at the 2004 CGIAR Annual General Meeting. This session was designed so as to present results from the email and telephone surveys and to obtain feedback on survey and interview results from Member representatives. Representatives of 15 Member agencies participated in this session on 28 October 2004, which was chaired by the European Initiative for Agricultural Research for Development. The session did not query any additional issues. Rather, it was focused on obtaining feedback on the perceived validity of the initial analysis of the interview and survey results.

2.3 Limitations

A number of forms of bias may potentially influence the results observed. "Social desirability bias" may cause respondents to state socially acceptable and desirable responses, rather than those that indicate his or her actual beliefs (Nancarrow and Brace, 2000). In this case, respondents may have overstated the relative importance of epIA studies and metrics, so as to appeal to SPIA's interests. Analysis has been kept as neutral possible, but the authors of the present analysis may still be individually biased as a result of personal background and perspective. Thus, it should be disclosed that the primary author has had a number of consultancy assignments with the Standing Panel on Impact Assessment, while the second author has been the head of evaluation at DANIDA.

2.3.1 Limitations to email survey

The respondent sample may not be representative, if response bias has affected whether survey recipients replied to the survey. It has been repeatedly demonstrated that responses are much more likely to be sent to surveys that address issues of importance to the informant (Sheehan, 2001). Thus, it is likely that respondents may be more interested in epIA studies than are non-respondents, and such is likely to inflate responses regarding the utility of the kinds of information that impact assessment produces. Consequently, the views expressed cannot be assumed as representative of the Member population at large. However, it is unlikely that any single evaluation tool will appeal to the entire spectrum of CGIAR Members, so it may be appropriate to hone epIA approaches to those audiences that are most interested.

The response rate for the survey may appear low, but is actually rather similar to that encountered for other email surveys of a similar nature. Sheehan (2001) reports an average response rate of 24% for formal email surveys conducted during 2000, compared with the 20% rate per donor representative in the current study. The present response rate by agency compares favourably at 35%. Furthermore, as only 17% of the respondents are from developing country donors (compared with 40% of the Membership), developing country views are not well represented. On the other hand, those donors who did respond provide 69% of the total CGIAR allocation, which means that a significant share of funding decisions may be reflected.

Another concern may be that the wording of the survey may have confused respondents, and might have affected the patterns of responses received. For example, the usage of some vague terminology may have introduced inconsistency in how the terms were interpreted, which reduces the reliability of results. The range of options presented may in and of itself introduced bias, due to the "anchoring effect" of the middle choice, while the order of questions may also have introduced bias through "context effects," (Schuman and Ludwig, 1982). Careful survey construction attempted to minimise these potential biases.

2.3.2 Limitations to telephone interviews

The interviews may share the potential problems of representativeness that affect the email surveys, as only those Members that are interested in epIA of CGIAR research may have chosen to participate. Indeed, only 41% of the CGIAR Membership took part in the interview process. However, the participants do represent the vast majority of Member funding to the CGIAR (92%). Consequently, while the opinions elicited may not be representative of the entire CGIAR Member community (particularly developing country donors, which only comprise 12% of those interviewed), they do represent the most significant financial contributors to the System.

The interview techniques applied utilised a semi-structured interview approach. This method's flexibility to adapt questions so as to elicit meaningful responses may cause questions to be posed inconsistently, and may thereby limit the comparability of responses. Furthermore, there is the potential for the interviewer to ask leading questions that may impart bias in responses. However, the interviewer employed is external to the CGIAR System and has not been directly involved in any epIA conducted on its behalf, which may reduce the potential for systematic bias in his conduct of the interviews.

3 RESULTS

3.1 Email survey

3.1.1 Respondent characteristics

A total of 24 responses were received from the following 22 Members: ACIAR, ADB, Austria, Belgium, DANIDA, DFID, EIARD, EU, GTZ (2 responses), IADB, IFAD, KARI, Mexico, Morocco, Netherlands, Philippines, Rockefeller, SDC, SIDA, Syngenta, USAID (2 responses), and the World Bank. Thus, the respondents include 10 developed countries, 6 international organizations, 4 less developed countries, and 2 foundations. Thus, the Member respondent rate was about 35%, but these 22 Members represent the bulk of CGIAR allocations, as collectively they provided 69% of CGIAR funding for 2002 ($247 out of $357 million total funding from Members). Thus, despite the small number of responses, a large proportion of total allocations was represented. It should also be noted that the recipients covered a wide spectrum of disciplines, from ecology to soil science, sociology and economics. Although economics was the most common discipline of respondents (53%), agricultural scientists were also well represented (43%). Such a mixed composition should help to alleviate the influence of disciplinary biases on the patterns of responses received.

3.1.2 Expectations for the role of *ex post* impact assessments

Member audiences have a view of epIA that is both interdisciplinary and summative. Although particular Member representatives have emphasised in other fora the derivation of "lessons" as a primary impact assessment function (e.g. Matlon, 2003), this was not reflected as the prime perceived purpose among the survey sample. Rather, as indicated by Figure 1, nearly two-thirds (61%) felt that the primary role of impact assessment is "to demonstrate that research output is making significant contributions to desired development goals." The next most commonly selected purpose, "to assign credible benefit values to the impacts of research and compare these to costs of investments" (19%), is reflective of the dominance of economic approaches to-date. Interestingly, only one of the five respondents who selected this purpose had a background in economics, so this selection appears not to be attributable to disciplinary bias. The "contributions to development goals" purpose was significantly higher rated than the other three purposes, while the "to assign credible benefit levels" was rated as significantly more important than "to increase understanding of adoption", according to the Wilcoxon Signed-Rank Test (α= 0.05).

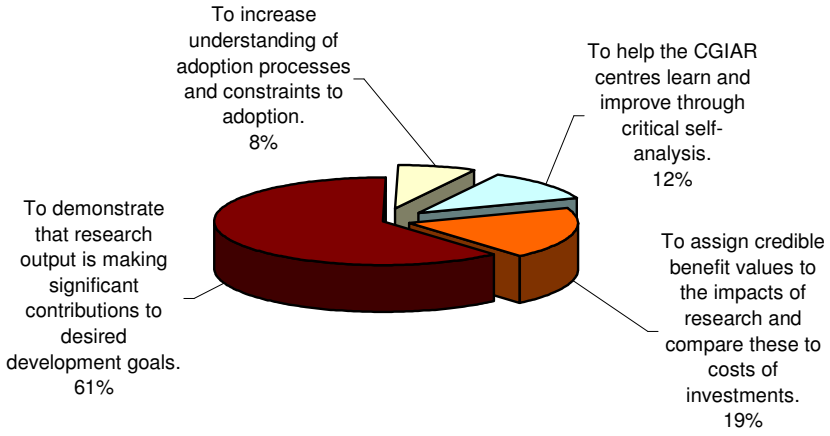
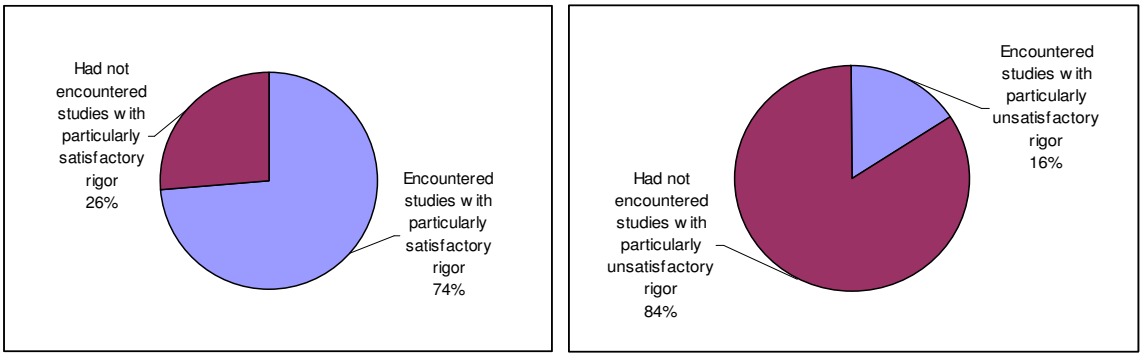


Figure 1. Member responses regarding the primary purpose of *ex post* impact assessment of agricultural research.

3.1.3 Satisfaction with epIA studies to-date

The respondents indicated moderate levels of satisfaction with epIA practices to-date in the CGIAR. Using a 10 (highest) to 1 (lowest) scale, credibility/rigour was tied with "relevance to institutional needs" with a median rating of 7.0 (means of 6.7 and 6.0, respectively). The lowest rated attribute was "comprehensiveness of programmatic coverage" with a median score of 6.0 (mean 5.6). Although these ratings appear rather similar, 71% of recipients rated "credibility/rigour" more highly than the other two attributes. According to the Wilcoxson Signed-Rank Test, credibility was significantly higher rated than comprehensiveness.

The 19 responses to queries regarding the quality of CGIAR epIAs previously encountered also suggest a fair degree of satisfaction with studies produced to date. While 74% of respondents had encountered studies of particularly satisfactory rigour, 84% had not encountered studies with particularly unsatisfactory rigour. However, this also means that 16% had encountered studies with rigour that was particularly unsatisfactory while more than a quarter (26%) had not encountered studies with rigour that was viewed to be particularly satisfactory (Figures 2 & 3).



Figures 2 & 3. Member responses regarding the rigor of impact assessment studies conducted by the CGIAR.

3.1.4 Use of epIA in allocation decisions

3.1.4.1 Reported influence of epIA compared with other forms of evaluation

EpIAs were rated as influencing resource allocations more than all information sources specifically listed, other than External Programme and Management Reviews (EPMRs) (Figure 4). For those ten respondents who rated it, the optional write-in "other" category was selected as most important. However, of these responses, five did not specify what this referred to. Of the remaining five, two referred to research proposals, two listed general factors affecting funding decisions, and one mentioned adoption constraint analyses.

EpIA, although the highest rated specific source of information, did not differ statistically in their rating from EPMRs, according to the Wilcoxon test. EPMRs also did not differ significantly from project reports produced by individual IARCs, while the latter was not significantly higher rated than annual reports from individual Centres. The four lowest rated categories did not differ significantly from each other, but were significantly lower rated than the top three categories.

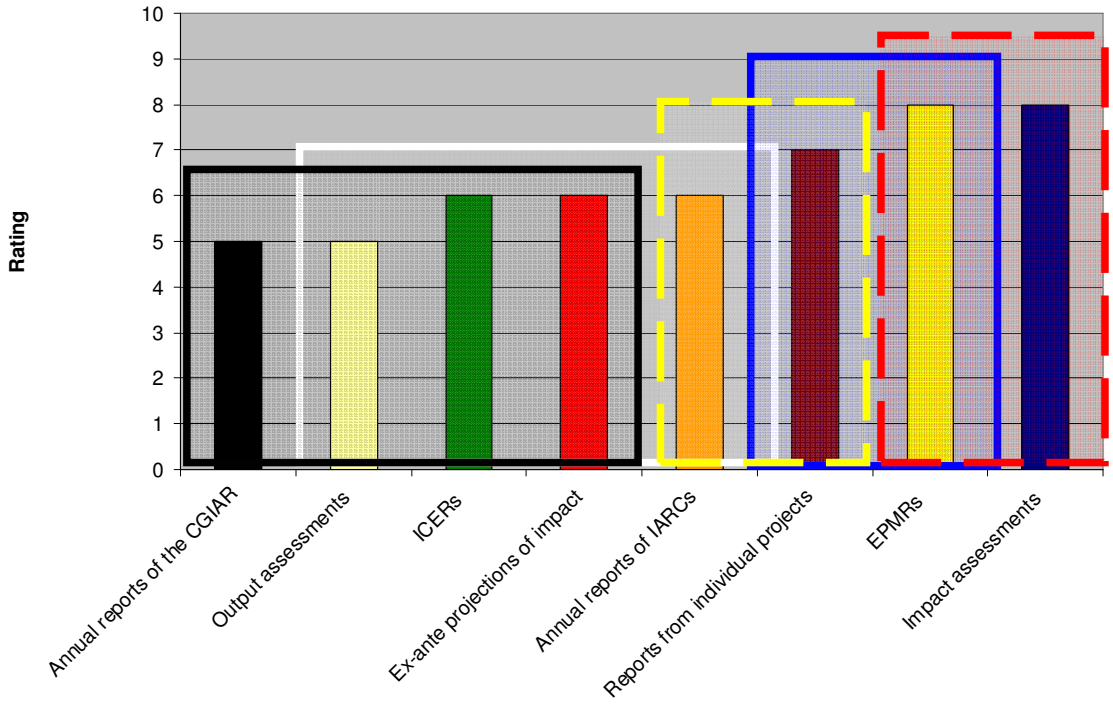


Figure 4. Median Member ratings of the importance of different information sources for influencing institutional allocative decisions across CG centres and across project/programmes within Centres. Groups of values that do not differ significantly, according to the Wilcoxon Sign Rank Test ($\alpha = 0.05$), are denoted by enclosure in boxes.

Of the 14 respondents who indicated whether CGIAR epIAs had been used for allocation decisions, 50% affirmed that specific studies had been used in this manner, while 50% indicated that no CGIAR studies had been directly utilised. In addition, one respondent who had not directly used such studies affirmed indirect influence. However, of those who declared such use, only two could provide details of the specific studies utilised.

3.1.4.2 Factors that facilitate epIA use

Respondents indicated greater utility for studies that validate large-scale effects at the mission level, rather than assessments of recent research or research within the topics that the Member is currently funding (Figure 5). "Recentness of the assessed research output" was the least important of the attributes listed for facilitating use. "Rigour of the assessment," "relevance to current priorities," and "ease of understanding findings" were tied above this with a median of 8.0. "Bearing of indicators on development goals" (9.0) was most important, followed by "magnitude/distribution of benefits" (8.5).

The highest ranked indicator – "bearing ... on development goals" was ranked statistically higher than all but "magnitude and distribution of assessed benefits." All other categories were statistically similar, except for "recentness of the research output that was assessed," which was significantly lower than all the other factors according to the Wilcoxon test.

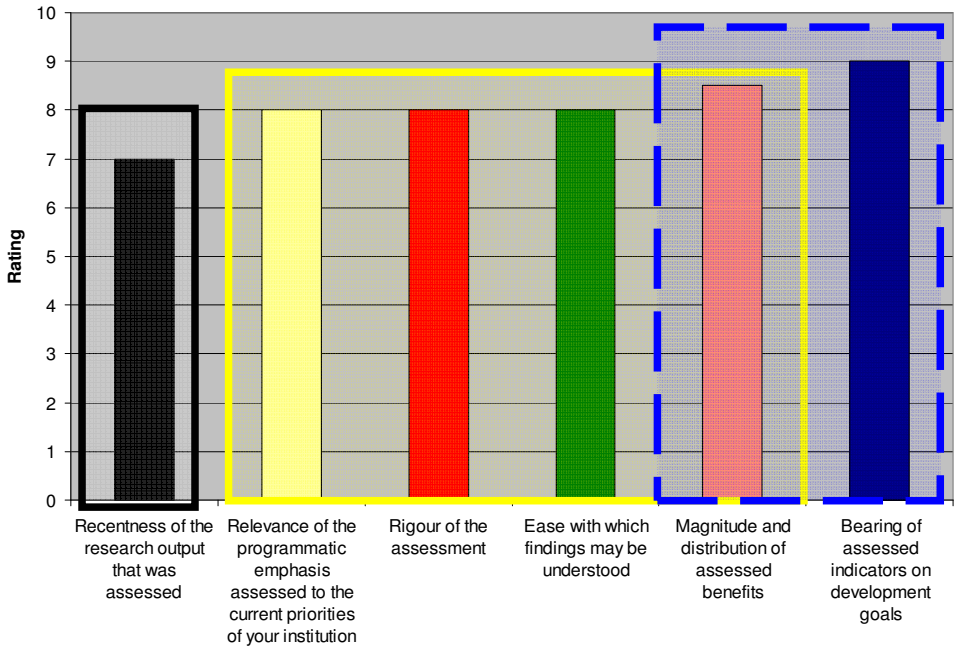


Figure 5. Median Member ratings of the importance of different factors for determining whether impact assessment results are useful. Groups of values that do not differ significantly, according to the Wilcoxon Sign Rank Test ($\alpha = 0.05$), are denoted by enclosure in boxes.

3.1.5 Readership of epIAs compared with other forms of evaluation

Readership rates did not differ in terms of median values among epIAs, assessments of research output adoption, and science quality assessments. However, in terms of mean values, readership of epIAs was the highest (4.2 in the last two years), followed by science quality assessments (4.1) and adoption studies (3.6) (n=20).

3.1.6 Demanded metrics and methods

3.1.6.1 Preferred impact metrics

Respondents indicated a general preference for indicators more closely related to mission level goals, than indicators related to the production of research outputs (Figure 6). "Reductions in risk/vulnerability," "improvements in livelihood strategies," and "changes in empowerment/capacity" were the most highly rated indicators, whereas "quantity of outputs," "measures of aggregate training," and "uptake by other institutions" were the given the lowest ratings. "Area of adoption/implementation" was rated equally to "gross economic surplus changes and "distribution of returns", while "changes in environmental factors" rated the same as "environmental values."

It should be noted that the predominant economic metric employed in *ex post* impact assessment, "gross economic surplus" did not differ significantly from any other factors than "livelihood strategies improved," "risk/vulnerability reduced," or "quantity of research outputs," according to the Wilcoxon test. The most favourably ranked two categories - "reductions in risk/vulnerability" and "improvements in livelihood strategies" were statistically dissimilar from the other less preferred indicators. No disciplinary trends were evident in the patterns of metric preferences expressed.

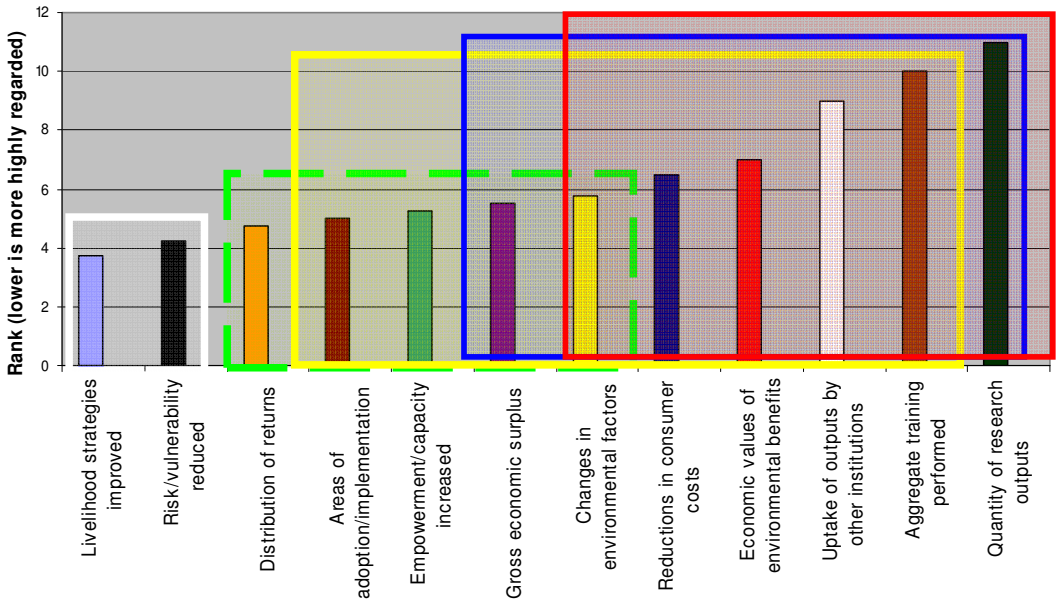


Figure 6. Median Member ratings of the importance of different impact indicators. Groups of values that do not differ significantly, according to the Wilcoxon Sign Rank Test ($\alpha = 0.05$), are denoted by enclosure in boxes.

3.1.6.2 External authorship

Fifty-eight percent of respondents indicated that assessment conduct by a specialist that is external to an organization comprises a "highly important" contribution to credibility. Only 4% indicated that external conduct is "not important" for credibility and rigour, while the remaining 38% indicated that this is "somewhat important."

3.1.6.3 Attribution

There was little enthusiasm for partitioning credit among collaborating institutions within impact assessments, as 60% of respondents felt that it was most appropriate that "all relevant institutional contributions/investments should be considered in concert, as any attempt to attribute individual complementary actions is inherently arbitrary." Twenty percent felt that it was more appropriate to partition credit by "qualitative assessment of relative contributions through key informant interviews," 13% favoured partitioning by "proportion of key outputs produced by specific institutions and programmes" and 7% preferred assigning credit according to "proportion of investment/staff time supplied by specific institutions and programmes" (Figure 7).

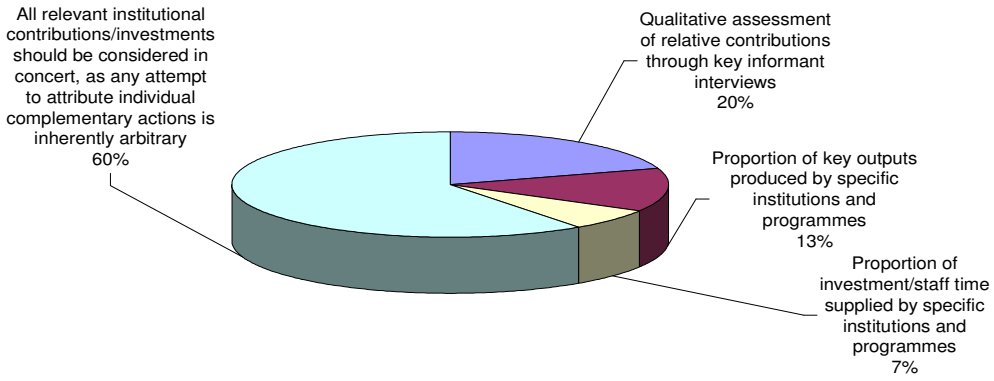


Figure 7. Member views regarding the most appropriate approach to the attribution of particular institutions involved in collaborative research.

3.1.7 Views on the demonstrated impact from natural resource management research

With regard to natural resource management (NRM) research, an area of research that has been the subject of relatively few impact assessment studies in the CGIAR, there was a general sense that some impact had been shown. More than half (52%) of respondents indicated that this research category had "moderate" or "extensive demonstrated impact", whereas 48% claimed that impact was "limited." At the same time, respondents indicated high expectations for impact assessment of this research area, as 63% felt that "NRM research, like any other form of strategic and applied research, should have mission-related benefits that can be measured and documented." The remaining 37% indicated that "the impact of past NRM research should ideally be assessed, but where attribution is difficult, demonstration of uptake is sufficient." None of the respondents indicated that it was adequate to only demonstrate science quality.

3.1.8 Unanswered or misunderstood questions

Question 8 attempted to get respondents to allocate a fixed set of resources among different evaluation methods/approaches with different costs, including epIA. However, the question was frequently misinterpreted, and very few respondents allocated the total level of resources specified. Consequently, these results cannot be considered comparable.

Only seven respondents completed the optional question 13, which attempted to query the value of different types of impact metrics, and few patterns are evident in this small number

of responses. The values assigned to different measures of impact have roughly equal averages, but very large variance in conjunction with the small number of responses restricts meaningful analysis.

3.2 Telephone interviews

3.2.1 Respondent characteristics

Representatives from 26 CGIAR Member institutions participated in the interview process, including: ACIAR, Belgium, CIDA, Columbia, DANDA, DFID, EC, FAO, France, Germany, IDB, IDRC, IFAD, India, Ireland, Japan, Netherlands, Norway, Philippines, Rockefeller, SDC, SIDA, Syngenta, UNEP, USAID, and the World Bank. Thus, among the represented institutions were 3 less developed countries, 2 foundations, 6 international agencies, and 15 developed countries. Although this includes representatives from only 41% of the 63 CGIAR Members, the bulk of the CGIAR budget is represented, as collectively these agencies provided 92% of 2003 CGIAR Member funding. Patterns of responses were similar among multilateral institutions, foundations, bilateral agencies and developing countries for every topic queried and no substantial patterns are apparent by Member size.

3.2.2 Member funding decision processes

3.2.2.1 Processes by which total CGIAR funding level is determined

The decision processes by which total funding levels were specified in each of the Member agencies were difficult to disentangle and precisely discern. All Members essentially had total allocation levels either determined at a high level in the bureaucracy concerned, or in a few cases, funding was determined through an external competitive process. For the latter cases, the mechanisms that determine allocations to the competitive funds were not investigated, as this was beyond the scope of the present study. In the former cases, overall funding envelopes were determined through some level of interaction between higher political/bureaucratic levels and the official CGIAR representative queried. In 45% of cases, this interaction was essentially described as a proposal initiated at the level of the CGIAR representative, which was then submitted to higher political/bureaucratic levels for modification and approval (n=20; Figure 11). In 45% of cases, it was indicated that higher decision-makers determined overall funding levels within which relative allocation among specific centres and activities takes place. Decision processes regarding overall funding often appeared to include little input from the CGIAR representative.

The information considered in decisions regarding overall funding levels was in most cases either not clearly described or was difficult to describe. Members indicated that this level of deliberation was essentially out of their influence, and that the deliberative processes were not necessarily systematic. As a result, the Members often had limited ability to describe how those at higher decision levels in the bureaucracy had actually derived their decisions regarding overall funding to the CGIAR System.

3.2.2.2 Decision processes for allocations among Centres and programmes

Allocations among Centres and for specific research programmes were described as decided at lower bureaucratic levels than decisions regarding overall funding. Typically, the official Member representative detailed a process whereby she or he would receive input from others with technical expertise through a series of consultations, and would thereafter finalise proposed allocations to Centres and projects, within the overall budget framework established

by higher management. Although more senior officials often needed to approve these specific allocations, this decision was typically described as within the domain of the representative to the CGIAR.

3.2.2.3 Information considered during the determination of allocation among Centres and programmes

Respondents were able to provide a more detailed description of the information they consider when making funding decisions across Centres. The most commonly claimed basis for funding decisions was 'convergence with Member institutions' priorities', as this was cited by 88% of respondents (n=21; Figure 8). The most commonly cited institutional priority was emphasis on Africa. 'Scientific quality' was the next most frequently considered factor (52%), which was often considered when assessing specific grant proposals from the Centres. Frequently, the assessment of proposals involved external peer review. Continuity and past performance were also considered relevant by 29% of the Members. Continuity was often manifested in the form of a desire to maintain funding for Centres or research areas that had traditionally received support from a specific Member in the past. Past performance includes responses referring to reputation and feedback from experts/partners, as well as documentation and analysis of the performance of individual research activities. Past impact was reported as a a relevant information source by about a quarter of respondents reporting.

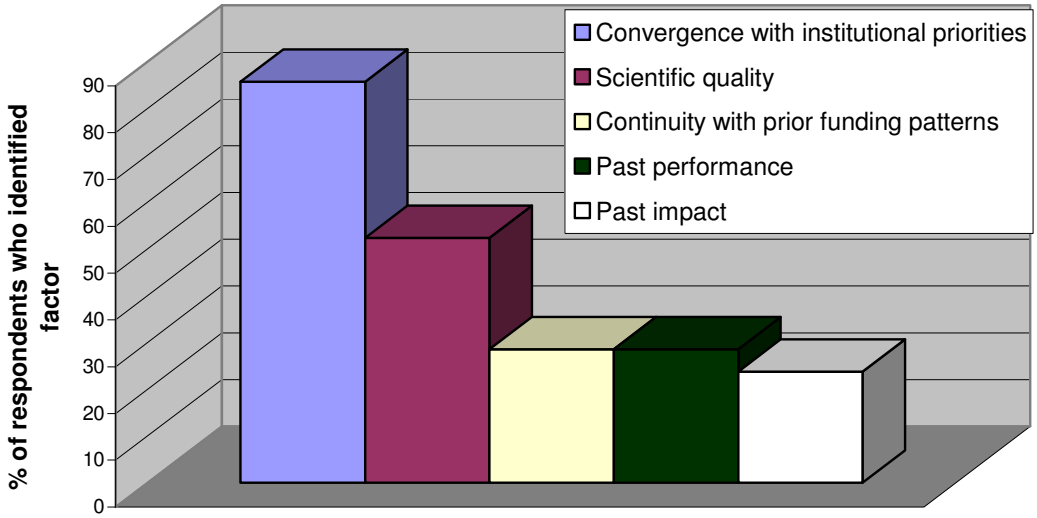


Figure 8. The five most frequently reported kinds of information considered by Members when making decisions to allocate resources among CGIAR research activities (n=21).

3.2.2.4 Flexibility to respond to new information

While it could be expected that Members would not have much flexibility to shift funding from one programme to another, due to institutional priorities and desire for programmatic continuity, this was not reported. Rather, 74% of the Members reported at least moderate levels of flexibility to shift allocations in response to new information (n=19; Figure 9).

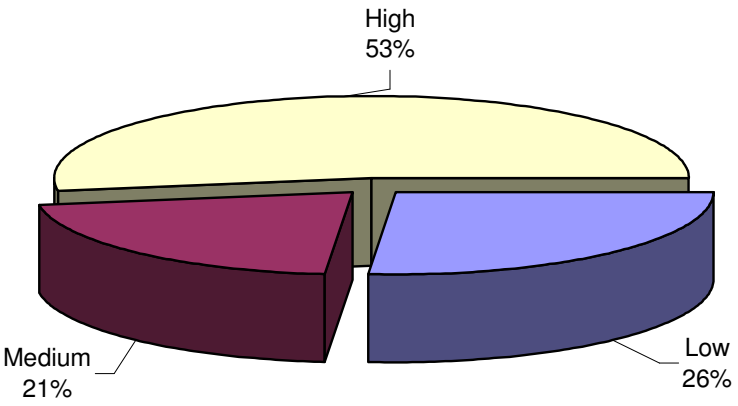


Figure 9. Reported flexibility to shift allocations among research topics, in response to new information (n=19).

3.2.3 Use of epIA in funding decisions

3.2.3.1 Overall direct use

Within responses, 50% of the Members made specific reference to applications of epIA to substantiate budget decisions (n=26; Figure 14). These responses often made use of the terms "defending" and "justifying" funding decisions to higher decision-making bodies. The half who made reference to defence of budgets represent 58% of CGIAR funding.

Interestingly, only 19% of respondents indicated that epIA may directly affect allocation decisions regarding the CGIAR (n=26). Those who reported such influence provide 24% of CGIAR funding.

3.2.3.2 General readership

When queried about an actual example of past use of an epIA study, slightly less than half the Members could actually name specific epIAs that they had read (n=26; Figure 10). An additional 15% cited evaluation studies, while 12% promotion summaries rather than impact assessments. It should be noted that the latter two categories embed *ex post* impact assessment results in their analyses. The remaining 27% could identify no specific study.

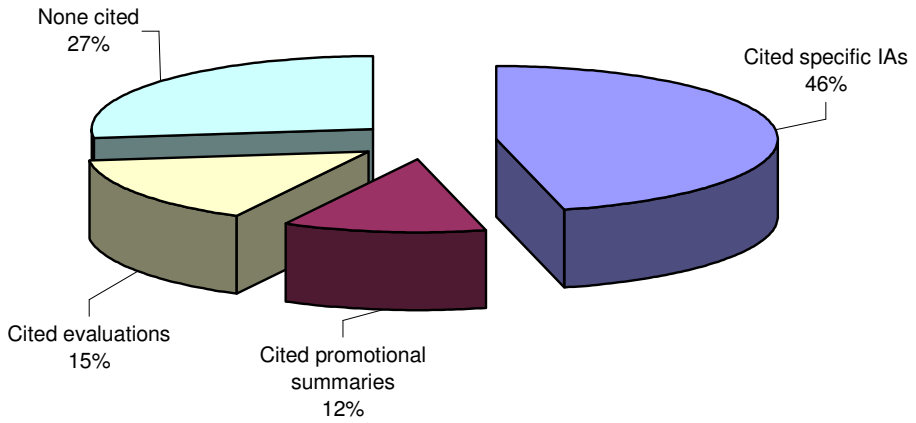


Figure 10. Types of studies mentioned by Members when use of specific epIA studies was queried (n=26).

3.2.3.3 Indirect forms of use

A majority (58%) of respondents indicated that epIA results are more useful in the context of other forms of information (n=26). These other kinds of information often included more details about the research assessed or different types of evaluation. Of those who expressed the utility of combining epIA results with other information, 53% specified that this should be analysis of scientific processes. The analysis of science demanded was most frequently in terms of partnerships (half of the 53%). Other information claimed to be useful for contextualizing epIA findings included information on failures (12%), alternative investment possibilities outside of the CGIAR (4%), and policy contexts for uptake (4%). The remainder of those interviewed mentioned the usefulness of considering epIA results in the context of other forms of information, but did not provide additional details.

Only 12% of respondents reported applications of epIAs outside of funding decisions (n=26). The uses mentioned included the formulation of programme strategies, and guidance to domestic institutions.

3.2.4 Perceptions of impact potential and the role played by epIA

3.2.4.1 Perceptions of impact to date

When queried about research areas believed to have resulted in the greatest impact to date, there was some loose convergence between perceived past impacts, and those plausibly demonstrated by epIA studies (Figure 11; Raitzer, 2003). Although germplasm improvement only received 17% of 2003 CGIAR allocations, it was ranked as an important source of impact by 88% of respondents -- more than was any other area of research (n=24). NRM research was next most frequently cited, with mention by half of the respondents, even though it has been repeatedly noted that there is a dearth of documented evidence of impact for this research category (Kelley and Gregersen, 2003; Lele et al., 2003; Raitzer, 2003). Policy and social science research follows, and was mentioned by 38% as a major source of impact, even though, here again, there is limited documentation of impact for this research area (Lele et al., 2003). Germplasm conservation is cited as the fourth most important source of impact, as it was mentioned by almost a third of those interviewed.

A recent meta-analysis of all large-scale economic *ex post* studies of CGIAR research impact found that germplasm improvement generated 84% of large scale plausible documented impacts, followed by pest management with 15% and 1% for NRM research (Raitzer, 2003). When these results are considered in concert with other observations of limited documented impact for policy and NRM research, it is far from clear what specific information has mainly informed those perceptions.

3.2.4.2 Perceptions of future impact potential

Perceptions of future sources of research impacts deviated slightly from perceptions of past impacts. Germplasm enhancement declines from mention by 88% of respondents as a sources of past impact to 64% who recognise it as a source of future impact, while NRM research rises from 50% to 68% (n=22). Similarly, policy research also rises from 38% to 54%. Other categories have little change.

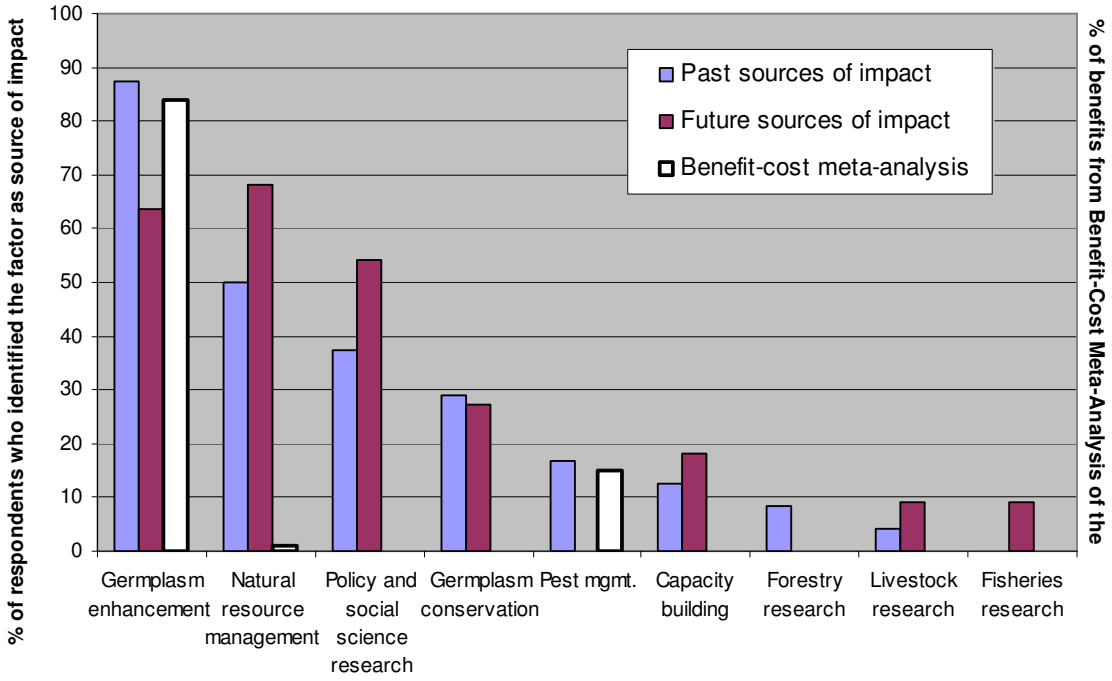


Figure 11. Reported perceptions of past (n=24) and future (n=22) sources of CGIAR research impact, as compared with those demonstrated to have large magnitude economic impact in Raitzer, 2003.

3.2.4.3 Information sources for impact perceptions

When queried about the information sources used to inform perceptions about past impact, epIA was mentioned with relatively low frequency (Figure 12). EPMRs were reported with the highest frequency (68%), followed by expert opinion (55%) (n=22). Interaction and direct feedback from Centres came next, and were mentioned by 36%. Member-commissioned evaluations and epIAs followed those and were mentioned by 23% of the respondents. However, it should also be noted that these categories are not completely distinct. EPMRs embed epIA findings, and expert opinions may be based on many information sources including epIA studies.

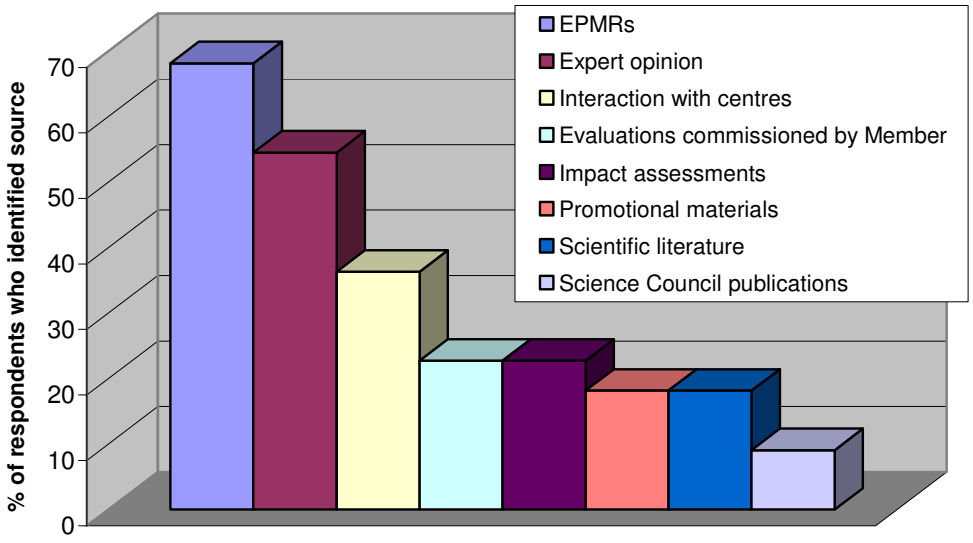


Figure 12. Sources of information cited by Members for perceptions of past and future impact potential (n=22).

3.2.5 Demands for impact information and study characteristics desired

3.2.5.1 General information needs

Members had a diversity of demands that they wished to see realized in future epIA studies (Figure 13). Most commonly, Members expressed a desire for summaries or briefs. These were needed for allowing information to reach higher decision-makers or the public more effectively, and to allow faster consideration of impact findings at the level of CGIAR Member representatives. These demands ranged from a desire for public-relations oriented material to more academic types of publications, and were spontaneously expressed by 54% of respondents (n=25). Following this, the most substantial content-related demand was for analysis of mission-level impacts on poverty (stated by 46% of Members). Improved methods and more details about the science being assessed were mentioned as demands by 31% each, while 12% wanted more information at a System level.

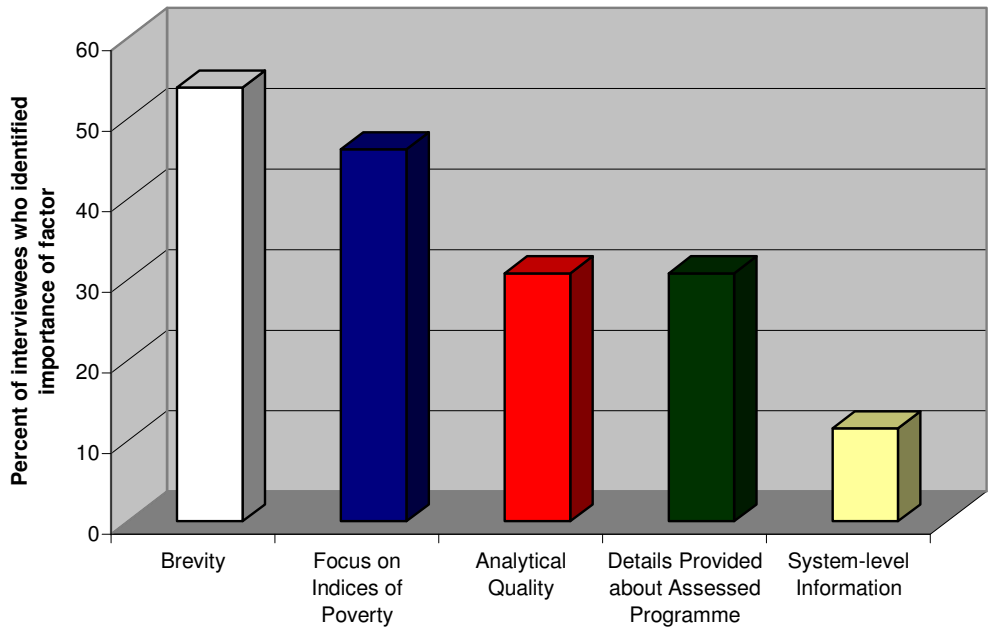


Figure 13. Demands expressed by Members for attributes of future *ex post* impact studies (n=25).

3.2.5.2 Methodological preferences

Members acknowledged the utility of both studies that focus on large-scale estimates of adoption and productivity effects, as well as smaller scale analysis of detailed effects at the household level: 64% reported that both types of studies are useful, with 40% claiming that both have equivalent utility (n=22; Figure 14). Of those who expressed a preference, 69% expressed preference for analysis of effects at the household level, and 31% claimed that large-scale estimates have greater utility.

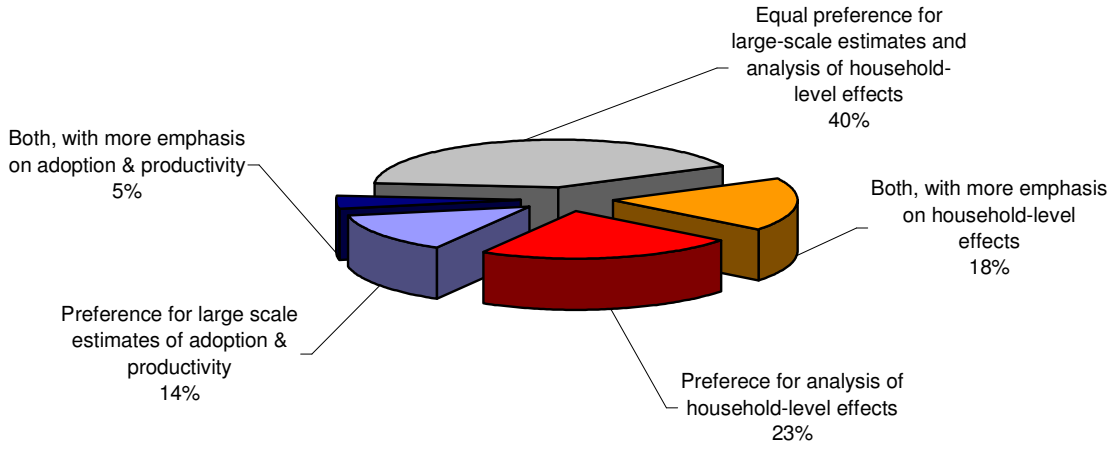


Figure 14. Preferences stated by CGIAR Members regarding emphasis on large-scale studies or small scale analyses of household level effects (n=22).

Economic metrics were reported to be useful to a majority of the Members interviewed. Two thirds of respondents indicated that such metrics add to the utility of studies (n=24).

3.2.5.3 Concerns

Although most respondents indicated no concerns regarding credibility or quality of epIA studies produced to date by the CGIAR, a substantial minority (27%, which represented 25% of the CGIAR budget), did express some reservations, in various ways (n=26; Figure 15). Many of these statements concerned possible bias, such as the impression of an "old boy's network," or skepticism towards "assumption" laden economic analyses that were impossible to validate. Others also expressed concern about the fact that economic metrics do not generally inform about the distribution or social implications of research benefits.

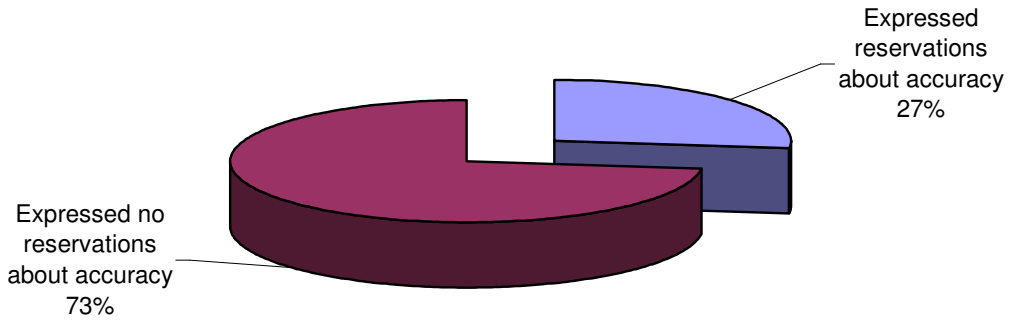


Figure 15. Percentages of Member representatives who expressed reservations about epIA accuracy or reliability (n=26).

3.2.6 Lunch session results

During the lunch session at the 2004 CGIAR Annual General Meeting, there was consensus among the 13 donor agencies represented that *ex post* impact assessments are important for confidence in the CGIAR System, although they do not play important roles in allocation decisions among Centres and programmes. Three specific donor comments indicated that epIA influence is often indirect. Two donor comments indicated the importance of epIA for continued investment in the CGIAR. A single comment pointed concerns about epIA accuracy, while another noted that epIA would be more useful for allocation decisions if coverage of research topics were greater.

4 DISCUSSION

4.1 Member decisions and the influence of epIA

The interviews illustrate that agricultural research funding decisions within Member institutions are highly complex (see summary of findings in Table 1). At higher decision-making levels, there is considerable ambiguity in answers regarding how decisions are made as to overall funding to the CGIAR System. This may be reflective of either limited knowledge of higher decision dynamics among the interviewed population, or it may indicate that these decisions are seldom systematic (or both). It has been repeatedly noted that factors other than information, such as ideologies, interests and institutions affect how information is processed in organizational decisions (Weiss, 1999), so it should not be surprising if the latter proves true.

Table 1. Summary of findings from the email survey and telephone interviews of CGIAR Member representatives.

Issue	Email survey (24 responses)	Interviews (26 repondents)	
Expectations for the role of ex-post impact assessments	80% affirmed the primacy of accountability objectives		88.5% of respondents only referred to investment decisions or accountability as epIA uses
Satisfaction with ex-post impact assessment studies to date	26% of respondents never encountered a study with very satisfactory rigor. Moderate ratings were given for comprehensiveness, relevance, rigour of epIAs produced to date.		27% expressed concerns about reliability or bias in studies produced to date.
Use of ex-post impact assessments in allocative decisions	EpIAs and EPMRs rated as most important, 50% claimed use of epIA's in allocation decisions	**Allocation processes**	Total funding level to CGIAR System usually set from above the CGIAR representative. Consultations used to determine allocation among centres, projects, with strong influence by the CGIAR representative.
		Degree of flexibility in allocation decisions	Flexible possibilities for allocation among centres, projects (74% rated freedom as moderate or high), but total funding level typically inflexible.
		Ability of epIA to affect allocations	19% (5 donors representing 24% of CGIAR budget) claimed that epIA could or do directly affect allocation patterns in their agencies
		Relation of perceived past impacts to perceived future impact potential	Percieved past impact loosely follows epIA results, while perceived future impact oosely follows perceived past impact
		Use of epIA in judgements regarding impact potential	Impact assessment is only the 5th most important source for opinions on impact potential and past impacts (cited by less than 1/4 of interviewees)
Study attributes that facilitate ex-post impact assessment use	Bearing on development goals most important demand		The most commonly cited demands were report brevity (54% of informants), bearing on poverty alleviation (46%), quality of assessment (31%), and details of the research provided (31%)
Demanded metrics and indicators	Metrics most directly related to poverty most highly rated		68% preferred assessments of household level benefits over large scale adoption/productivity studies
Demand for economic indices	Little difference between stated preferences for adoption data and economic indicators		67% affirmed that economic indicators add value over adoption estimates
Readership of ex-post impact assessments as compared with other forms of evaluation	On average, 4 epIAs read over last two years - roughly same as adoption studies or assessments of scientific quality		46% of interviewees were able to cite specific epIAs, 15% cited evaluations, 12% cited PR materials, and 27% cited no specific study.
Attribution	60% of respondents stated that credit should not be partitioned among collaborators		19% explicitly expressed desire for information on how CGIAR is working with partners
EpIA use other than for CGIAR funding	inconclusive		Only 12% reported other uses, in terms of direction to other institutions, internal strategies, etc.

At lower decision levels, more details of the basis of allocation among Centres and programmes were provided, probably due to the greater roles played by respondents in these decisions. In addition, moderate to high levels of flexibility regarding these decisions were reported among most of those interviewed. However, despite this freedom, it is clear that the consideration of past patterns of impact in these deliberative processes is only marginal. Other factors, such as political priorities, scientific quality, and desire for funding continuity play much more prominent roles in these decisions.

Patterns of actual allocations can be expected to derive at least in part from perceptions of impact potential, if Members are truly interested in efficient allocation of resources. It can also be logically supposed that documented past impact should inform perceptions of impact potential, if indeed epIA directly influences allocation decisions (as discussed below). However, the information sources cited as a basis for perceptions of impact potential appear to focus primarily on scientific quality, as EPMRs, expert opinion, and interaction with Centres comprise the primary basis for perceived impact potential. EpIA is only the fifth most important source for such perceptions. While this appears to diminish the role of epIA, compared with other statements of allocative importance for epIA, it is also largely consistent with the information claimed to be considered during allocation decisions.

However, there are also some important caveats to these observations. While only a small percentage of Member representatives reported direct use of epIA in allocation decisions, those that did include major Member organizations, such as USAID and SIDA. As longstanding pillars of the CGIAR System, these Members often set the "agenda" that other Members follow. Therefore, such influence may carry far more significance than the small proportion reporting direct use indicates.

Furthermore, the information sources listed as the basis for perceptions of impact do embed epIA results, so it is unclear what contribution to these perceptions is actually provided by *ex post* assessment. Perceived past patterns of impact do indeed at least loosely follow the results of *ex post* assessment. The link becomes more tenuous when extrapolated to perceived future impact potential, and it remains uncertain what role perceptions of past patterns of impact play in formation of perceptions of future impact potential. Unfortunately, the interview process did not attempt to differentiate between information sources for past and future impact perceptions.

When the context of agricultural research is considered, it may be appropriate for perceptions of future impact potential to be only partially informed by past patterns of impact. Research is a dynamic process, as research methods, topics, and approaches continually evolve over time. Lag times from the production of research outputs to impacts on mission-level goals are long and unpredictable, while the process of research itself is uncertain and may involve many "dry holes" before successful innovation can be fostered (Alston et al., 1998). When innovation occurs, it is often the product of host of complementary factors that allow for research to translate into impact by assisting in diffusion, and by making conditions conducive to adoption (Ekboir, 2003). As a result, it is very difficult to relate historical patterns of impact to the potential of an ever-evolving research portfolio. Information concerning current performance and relevance may indeed have important roles to play.

A couple of respondents indicated that information on past impact is difficult to utilise if it does not allow for comparative appraisal of performance across research activities. As noted

by Kelley and Gregersen (2003), Lele (2003), Raitzer (2003) and others, epIA coverage is heavily concentrated on a small subset of the CGIAR research agenda related to crop germplasm improvement. It may hence not be appropriate to expect that such partial coverage is appropriate to information needs concerning comparative allocations across activities, especially when the onslaught of other information facing decision-makers is taken into account.

However, this does not necessarily negate demands for impact-related information voiced during the interviews. Rather, there are strong indications that impact-related information is of value, although not for the purpose of directly informing funding decisions. In the email survey, epIA was reported to be the most important source of information listed for allocation decisions, although only half the Members reported this type of direct use. An equal proportion of interviewees mentioned the importance of epIA for defending budget decisions, while an even greater proportion stated that epIA is more valuable in the context of auxiliary information. Furthermore, stated readership levels do indicate that there is some consideration of epIA findings, as survey responses report average readership of two studies annually.

4.1.1 Instrumental use

In prior studies of evaluation use (Patton, 1977; c.f. Leviton and Hughes, 1981), it has been observed that intended users of evaluation reported high rates of use in decision-making, but could not specify how evaluation findings were precisely applied. In fact, evaluation use as a body of systemic inquiry arose largely out of the failure of evaluation to produce observable influence on public policies and projects. However, when indirect patterns of influence are considered, it becomes clear that evaluation does have substantial effects. Yet, they are largely indirect and involve intermediate uptake events and recombination with other forms of information. Evaluation is thus recognised to not exert influence in isolation. Rather, evaluation (like policy research) complements and competes with many alternative sources of information. In this vein, it has become recently vogue to shift the evaluation lexicon from the term "use" to the term "influence," so as to capture this more nuanced relationship between evaluation results and application in target decisions. Similarly, the importance of "contextual factors" for determining use has also been highlighted in recent literature, as use is rarely uniform across evaluations, and may only occur with the right mixture of timing and circumstance (Henry and Mark, 2003).

Accordingly, the interviews and email surveys did not identify many examples of specific donor decisions instrumentally shifted by epIA findings. This is similar to the virtual absence of observations of instrumental use in the literature.

One of the primary reasons for which epIA is unlikely to exert direct instrumental influence in decisions is also evident in the interviews. In most donor agencies there are few staff that deal with the CGIAR, as the IARCs receive a very small share of most Member agencies' development assistance budget. In the context of the many forms of information that these few staff receive, this means that there is little time that can be devoted in most agencies to the consideration of impact assessment studies.

4.1.2 Conceptual use

However, use of evaluation in a less direct "conceptual" manner has been much more frequently documented. For example, Rich (1977; c.f. Cousins and Leithwood, 1986) found that evaluation frequently was used to improve the general understanding of decision-makers. Hawkins et al. (1978) observed that key decision-makers regarding a drug-treatment programme were much more prone to consider information embedded in personal communications, rather than in paper reports. Siegel and Tuckel (1985) found that assessment of the impact of evaluation research based only on direct effects on target policies was "spurious" as more significant influence is exerted through "refocusing" and other shifts in the overall understanding of issues concerning the evaluated programme. Boyer and Langbein (1991) found that influence is not continuous, and may happen at critical junctures when evaluative information complements other information sources well. Furthermore, Patton (1977; c.f. Leviton and Hughes, 1981) found that a principal conceptual form of influence concern reductions in uncertainty regarding decisions to be taken.

The implication of the prevalence of "conceptual," rather than "instrumental" use for the results observed is that direct application of evaluation findings in specific decisions may not be readily observable, even if evaluation does play an important role. As a result, the fact that direct changes in CGIAR allocation patterns are not observable after epIA indicates the value of particular research areas is not surprising. However, this does not necessarily refute the influence of such studies. Rather, the influence may be exerted through enhanced understanding of how research can foster impact, or how the agricultural research system functions. As epIA exerts influence amid a wide array of other forms of information, it is very well possible that the without epIA counterfactual may be declining support to those research areas with demonstrated impact. Consequently, the consistently stated importance of epIA for donor decisions should not be dismissed simply because allocations do not follow demonstrated impact.

When comparing the interview and survey results, a number of contradictions regarding the role and stated importance of epIA for donor decisions become apparent. While epIA is stated to be the most important information source for allocation decisions in the survey, it is given lower ranking as a source of information for perceived impact potential in the interviews. Furthermore, the surveys indicate that only half of respondents have directly applied epIAs during funding decisions. This dichotomy may results from differences between questions that imply instrumental and conceptual patterns of use. The more specific questions that imply more direct forms of use may receive responses that indicate less application of epIA results, compared with those that allow for influence through intermediate pathways, such as EPMRs (which reference epIA findings). If such is the case, it appears that impact assessment may have influence that is primarily "conceptual."

Furthermore, optional comments by several respondents in the surveys also appear to indicate that conceptual use is common for epIAs. For example, one respondent noted that "it is vague stuff in this institution, but knowledge that such work is going on helps, probably" in allocation decisions. Another respondent who responded that epIAs had not been directly used clarified that:

This does not mean though, that impact studies have no influence. For instance, a convincing impact study can contribute to decision making because it illustrates the competence of a center and its scientists or because it clarifies the state of the art and the problems that still need research.

Furthermore, in the interviews, there was general awareness of impact demonstrated through formal assessment, as a clear majority of respondents could successfully identify the areas with the highest levels of assessed returns. While impact assessments were not cited by a majority of interviewees as the direct sources for these perceptions, the information sources most frequently cited do embed epIA results. Thus, it appears that many of those who indicated no direct application of findings during budget decisions may still have these decisions influenced by epIA through more indirect means.

4.1.3 Symbolic use

Casual observations of "symbolic use" or application of evaluation to legitimize decisions that would be undertaken anyway in the absence of the evaluation findings are common. Quantified patterns of such use are rarer, however. Knorr (1977; c.f. Leviton and Hughes, 1981) provides some quantitative evidence of the prevalence of this form of use, and this has been substantiated later by Shulock (1999). Initially considered as a form of "misuse," Shulock recasts this use as a legitimate application of findings in a manner that contributes incremental influence, and which assists in the appearance of rationality by decision-makers.

Donor demands for accountability-related information could be seen in this light as users of epIA findings who only apply them in a persuasive or legitimising fashion. In fact, the way in which "accountability" is described as a purpose for impact assessment actually gives the impression of "symbolic use," as "justification of research investment" is a primary role for epIA (Impact Assessment and Evaluation Group, 1999). Half of the interviewees also spontaneously mentioned the use of epIA for the "defence of budgets," a symbolic application. Furthermore, this form of use may also explain how high rates of use of impact assessment could be reported, while allocations appear to bear little influence. As a result, it seems that symbolic use may be common.

4.1.4 Process use

"Process use" is actually a form of "conceptual use" transmitted through involvement in evaluation by intended users. Greene (1988; c.f. Shulha and Cousins, 1997) and Patton (1978) suggest that involvement in evaluation stimulates use, while Turnbull (1999) produces quantitative evidence to substantiate this argument.

Process use is largely irrelevant to epIA for donor audiences, as donors have insufficient human resources to become involved in the scores of epIA studies produced annually by the Centres. In fact, for their accountability objectives, involvement would even be negative perceived, due to the high value that donors place on independence in assessment conduct. Accordingly, there was little indication of process use in the survey response or interviews.

On balance, the evaluation use literature, the survey results, and the interview findings indicate that direct application of evaluation findings for programmatic change is rarely observable. Rather, the primary pathways of influence are indirect and involve incremental improvement to the general understanding of programme functions. Such conceptual

influence usually involves combining evaluation findings with other forms of relevant information, often through repackaging in intermediate dissemination avenues.

4.2 Preferences for methods and metrics

Despite the apparent prevalence of indirect forms of use/influence, there are clear preferences for epIA approaches among the Member representatives. In particular, there is strong stated demand for evidence of impact at the 'mission-level' among the respondents. Consistently, in the survey responses, indicators most directly related to the CGIAR mission of poverty alleviation were highest ranked, i.e., studies that beared most directly on development goals were rated as most useful, impact assessments were considered highly important for allocation decisions, and nearly two thirds thought that epIA should primarily "demonstrate that research output is making significant contributions to desired development goals." Similarly, in the interviews, Members expressed preference for analysis of impacts at the household level, and bearing on development goals was the most consistently demanded aspect of study content. Nearly half of the interviewees spontaneously mentioned the need to include more analysis of impacts on poverty-related metrics or more distributional analysis. According to these results, it seems that additional studies on the impacts of CGIAR research are warranted, even if applied only in indirect ways by Members, so as to reduce uncertainty surrounding investments.

Since epIA has been dominated by economic methods (Pingali, 2001) one particularly counter-intuitive finding of the survey is that economic approaches appear not be strongly preferred over the physical measures applied in benefit calculations. In terms of impact metrics, areas of adoption were rated similar to gross economic surplus measures, while physical metrics of environmental benefits were also considered equal to calculations of environmental values. Given that considerable effort has been spent on calculating gross economic surplus measures from adoption estimates, this is surprising.

However, the interviews appear to explain some of this quandary. Two-thirds of the interviewees felt that economic metrics add utility to biophysical measures of productivity enhancement. Yet, most of those who did not express such sentiments, clarified that the economic benefit estimates provided to date have been too assumption-laden, and are not transparent enough to validate. A few interviewees noted that they felt that non-CGIAR attributable influences were insufficiently credited, that the approaches were biased, or that quality had been lacking. These remain as essential challenges to be addressed in future epIA studies, as a substantial minority of those queried echoed these concerns. Furthermore, other analyses (Cooksy, 1997; Raitzer, 2003) note some of these same areas for methodological improvement. In addition to these concerns about accuracy, a few of those who were sceptical about economic metrics pointed out that they provide little detail on the distributional implications of benefits generated.

There was also pervasive demand for brief summaries of epIA findings in the interviews conducted. Given the apparent preponderance of conceptual and symbolic uses of epIA results and the fact that decision-makers are flooded with many competing forms of information, these demands appear almost self-evident. In the context of copious competition from other flows of information, it is imperative that results are presented in a manner that attracts attention, and which minimises the time necessary to glean findings. However, it remains unknown whether such may sway use towards advocative applications.

Demand for the pairing of epIA results with other forms of information was also widespread. In particular, it appeared that there was demand for information that bridged epIA results with the findings of other forms of evaluation, as a large share of respondents wanted more details about the research programmes assessed, in terms of partnerships and performance. Consequently, it appears appropriate that comprehensive evaluations embed epIA findings as one component of a larger evaluation portfolio. To some extent, current EPMRs already attempt to do this.

These audiences also appear to demand epIAs of a broader range of research activities. Of the rated epIA attributes in the email survey, comprehensiveness of programmatic coverage was considered the least satisfactory, while respondents overwhelming felt that NRM research must have documented benefits. In the interview, a number of respondents indicated that limited coverage of different research topics served as a constraint to epIA use in decisions. It thus appears that a key priority should be to expand analysis of impacts at the mission-level to as wide a spectrum of research activities as possible, if Member demands are to be satisfied.

According to survey responses, it is also important for credibility that epIAs be conducted by assessors external to the assessed institutions. This stands in contrast to certain recommendations for "learning-oriented" evaluations that are conducted internally and involve programme staff, so as to maximise internal process use (Horton and Mackay, 2003). This, in combination with preferences for indicators far down the impact pathway, suggests that Member demands necessitate somewhat separate approaches for accountability and learning oriented studies, as data on such indicators cannot be provided in the timeframe of research decisions. Long lags of a decade or more imply that the research agenda will have substantially evolved by the time such impacts are evident, and that lessons related to such historical research may be of limited relevance.

Accordingly, Members did not view epIA so much as a vehicle for learning but, primarily, as a tool for accountability. Nearly two-thirds of the Members responding affirmed the primacy of the latter purpose. Consequently, it may be appropriate to focus on epIA on mission level goals for this audience, with separate efforts tailored towards internal learning.

It should be noted that this is not the first study to recognise that separate efforts are merited for accountability and learning purposes, as the evaluation literature has already recognised this dichotomy (Cracknell, 2000 [c.f. Hall et al., 2003]). At higher "strategic" or "policy" decision levels, there has been observed greater demand for information on overall programme effectiveness (Nielsen, 1975). Conversely, at lower levels of management or implementation, greater demand for evaluation of details regarding the implementation of specific programme elements has been observed (Wholey, 1970, Brickel, 1974; c.f. Leviton and Hughes, 1981). The preferences observed in the present study for information on impact achieved essentially represent similar demands among policy-level decision-makers for outcome-related information.

4.3 Implications for epIA in the CGIAR

Clearly, epIA in the CGIAR has an important role to play, according to the strong demand for impact-related information among Member audiences. However, responses allude to a number of areas for potential improvement.

According to the results obtained, it should not be expected that epIA has primary utility as a direct "instrumental" input into mechanistic Member funding decisions. Rather, epIA plays a more subtle role of building confidence and reducing uncertainty, which is often embodied in the "accountability" function of epIA findings. This implies that epIA should not be extrapolated as the sole source for specific recommendations for action, but rather should be produced as one of many inputs into deliberative processes.
For accountability purposes, it appears that studies should be more frequently focussed on mission-relevant impacts on poverty, so as to meet Member demands. To do so requires greater analysis of the distribution of benefits generated, so as to identify more precisely the proportion of benefits accruing to poor populations. Greater analysis is also needed, so as to discern how these benefits are realised within beneficiary households.

In so doing, it appears that it may be necessary to divorce some "accountability-oriented" studies from studies oriented towards direct "internal learning." The strong stated demand for analysis of end outcomes among the Member population runs counter to demands for "formative" (process) evaluation documented previously among programme implementers. Furthermore, accountability audiences have strong preference for externally-conducted evaluation, whereas internal evaluation is widely appreciated as more suitable for internal uptake of lessons.

It appears that greater investment in study quality may be necessitated, so as to satisfy the substantial minority who voiced concerns about credibility. In particular, greater transparency about the nature of assumptions underlying analysis may effectively address many concerns.

The results here confirm the widely acknowledged need to expand epIA to a broader array of research activities. As noted by a number of Members, epIA would be of much greater utility if it permitted greater comparative analysis of different research options.

Finally, it appears that dissemination practices for epIA studies could be improved. As noted by a majority of interviewees, brief summaries of key epIA findings should be developed to facilitate the advocative roles that epIAs often play. Not only would such outputs increase the likelihood that findings reach key decision-makers, but in many cases the CGIAR focal persons stated that even they themselves did not have the time to read lengthy reports. Hence, it seems that attractive summaries and briefs of only a few pages may markedly help to improve dissemination.

5 CONCLUSION

EpIAs are repeatedly produced that show outstanding efficiency from investment in a small subset of CGIAR research activities, yet funding does not increase for these research areas relative to investment in areas of research for which no or few impact results are available. Does this mean that epIAs are not used, despite strong stated demand for these analyses?

The findings of this study suggest otherwise. Although epIA is not a direct driver of specific funding decisions, the confidence in CGIAR capacity that epIA findings impart appears to be of substantial importance for continued support to the system. The CGIAR has successfully maintained stable funding growth, even as resources for international development have become scarcer, and other "fads" in international development have come and gone. It is likely that epIA has helped to contribute to this outcome. However, the precise pathways by which epIAs help engender confidence remain enigmatic, as are influences on shifts in funding among different areas of research.

Much remains to be learned about the utility and influence of epIA studies, as the contribution of such to decision-making is still somewhat unclear. Effort is needed, so as to better understand the incremental value of investment in study scale, study rigour, and distance down the impact pathway. While it may be apparent that, all else being equal, metrics more directly related to mission-level goals are preferred, it is not clear how these preferences are affected by tradeoffs involved when actually selecting methods and approaches for epIA. Additional investigation is needed to clarify these issues.

However, a number of clear messages do emerge from donor audiences. Studies that focus on poverty-relevant metrics are particularly demanded, as are brief summaries of impact assessment findings. Increased study transparency and external conduct could help to alleviate concerns about accuracy voiced by a substantial share of respondents. Impact assessment coverage of a broader array of research topics would make epIA more relevant to funding choices. More generally, many donors have emphasised that continued impact assessment is essential for continued support.

The present exercise should be regarded as a beginning rather than an end to exploration of Member preferences and uses for evidence of research impact. Ongoing dialogue is needed to ensure that studies meet demands as well as possible, and that key constraints to influence are effectively addressed.

6 REFERENCES

Alston, J. M., B. J. Craig and P. G. Pardey. 1998. Dynamics in the creation and depreciation of knowledge, and the returns to agricultural research. *EPTD Discussion Paper No. 35*. Washington, DC: International Food Policy Research Institute.

Alston, J.M., C. Chan-Kang, M.C. Marra, P.G. Pardey & T.J. Wyatt. 2000. A Meta-Analysis of Rates of Return to Agricultural R& D: Ex Pede Herculem? *Research Report 113*. Washington, DC: IFPRI.

Alston, J.M., G.W. Norton and P.G. Pardey. 1996. *Science under scarcity: Principles and practice for agricultural research evaluation and priority setting*. Ithaca, NY: Cornell University Press.

Anderson, J.R. 1985. Assessing the impact of farming systems research: Framework and problems. *Agricultural Administration* 20(4): 225-35.

Anderson, J.R. 1998. Selected Policy Issues in International Agricultural Research: On Striving for International Public Goods in an Era of Donor Fatigue. *World Development*. 26(6): 1149-1162. Baker, J. 2000.

Baker, J. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington, DC: World Bank.

Boyer, J.F. and L.I. Langbein. 1991. Factors Influencing the Use of Health Evaluation Research in Congress. *Evaluation Review.* 15(5): 507-532.

CGIAR Secretariat. 1982. *CGIAR International Centers Week, Washington, D.C., November 8-12, 1982: Main Conclusions Reached and Decisions Taken*. Washington, DC: CGIAR Secretariat.

Cooksy, L. 1997. *Methodological Review and Synthesis of Existing Ex Post Impact Assessments*. Rome: Impact Assessment and Evaluation Group, CGIAR.

Cousins, J. B. and K.A. Leithwood. 1986. Current empirical research in evaluation utilization. *Review of Educational Research*. 56(3), 331-364.

Ekboir, J. 2003. Why impact analysis should not be used for research evaluation and what the alternatives are. *Agricultural Systems* 78(2): 166-184.

Gardner, B. 2003. The CGIAR at 31: An Independent Meta-Evaluation of the Consultative Group on International Agricultural Research. Washington, DC: World Bank Operations Evaluation Department. 53 p.

Hall, A., V.R. Sulaiman, N. Clark, and B. Yoganand. 2003. From measuring impact to learning institutional lessons: an innovations systems perspective on improving the management of international agricultural research. *Agricultural Systems* 78(2):213-242.

Hawkins, J.D., R.A. Roffman, and P. Osborne. 1978. Decision maker's judgments: the influence of role, evaluative criteria and information access. *Evaluation Quarterly*. 2(3): 435:454.

Henry, G.T. and M.M. Mark. 2003. Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*. 24(3): 293–314.

Horton, D. and R. Mackay. 2003. Using evaluation to enhance institutional learning and change: recent experiences with agricultural research and development. *Agricultural Systems* 78(2): 127-142.

Impact Assessment and Evaluation Group, 1999. Impact Assessment of Agricultural Research: Context and State of the Art. *Paper prepared for the ASARECA / ECART/ CTA Workshop on Impact Assessment of Agricultural Research in Eastern and Central Africa, Uganda, November, 1999*.

Johnson, R.B. 1998. Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*. 21: 93-110.

Kelley, T. and H. Gregersen. 2003. NRM impact assessment in the CGIAR: meeting the challenge and implications for ICRISAT. In Shiferaw, B. and Freeman, H. A. (Eds). Methods for assessing the impacts of natural resource management research. A summary of the proceedings of an International Workshop, 6-7 December 2002, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, India.

Lele, U. 2003. *The CGIAR at 31: An Independent Meta-Evaluation of the Consultative Group on International Agricultural Research*. Washington, DC: World Bank Operations Evaluation Department. 68 p.

Leviton, L.C. and E.F.X. Hughes. 1981. Research on the Utilization of Evaluations: A Review and Synthesis. *Evaluation Review.* (4): 525-548.

Matlon, P.J. 2003. Foreword. *Agricultural Systems* 78(2):123-126.

Nancarrow, C. and I. Brace. 2000. Saying the "right thing": Coping with Social Desirability Bias in Marketing Research. *Bristol Business School Teaching and Research Review.* 3. [http://www.uwe.ac.uk/bbs/trr/Issue3/Is3-2_2.htm, accessed 26/5/2005].

Nielsen, V.G. 1975. Why evaluation does not improve program effectiveness. *Policy Studies Journal*. 3(4). 385-389.

Özgediz, S. 1995. Strengthening *Evaluation in the CGIAR: Needs and Options*. CGIAR Task Force on Impact Assessment. Washington, DC: CGIAR Secretariat.

Patton, M.Q. 1978. *Utilization-focused Evaluation*. Beverly Hills, CA: Sage.

Pingali, P.L. 2001. Milestones in Impact Assessment Research in the CGIAR, 1970-1999. Mexico, D.F.: CGIAR Standing Panel on Impact Assessment. 37 p.

Public Awareness and Resource Mobilization Committee. 1994. *Minutes of the Public Awareness and Resource Mobilization Committee (PARC) Meeting*. Washington, DC: CGIAR Secretariat.

Raitzer, D.A. 2003. Benefit-Cost Meta-Analysis of Investment in the International Agricultural Research Centres of the CGIAR. Rome, Italy: Science Council Secretariat, FAO. 45 p.

Schuman, H. and J. Ludwig. 1982. The Norm of Even-Handedness in Surveys as in Life. Proceedings of the Survey Research Methods Section, American Statistical Association (1982). Alexandria, VA, USA: American Statistical Organization. [http://www.amstat.org/sections/srms/Proceedings/papers/1982_002.pdf, accessed 26/5/2005].

Sheehan, K. 2001. E-mail Survey Response Rates: A Review. *Journal of Computer-Mediated Communication*. 6(2). [http://www.ascusc.org/jcmc/vol6/issue2/sheehan.html, accessed 26/5/2005].

Shulha, L. M. and J. B. Cousins. 1997. Evaluation Use: Theory, Research, and Practice since 1986. *Evaluation Practice* 18(3): 195–208.

Shulock, N. 1999. The Paradox of Policy Analysis: If It Is Not Used, Why do We Produce So Much of It? *Journal of Policy Analysis and Management*. 18 (2): 226-244.

Siegel, K. and P. Tuckel, P. 1985. The Utilization of Evaluation Research: A Case Analysis. *Evaluation Review*. 9(3): 307-328.

Standing Panel on Impact Assessment. 2004. *Report to the Annual General Meeting of the CGIAR from the Standing Panel on Impact Assessment (SPIA) of the Science Council*. Rome, Italy: CGIAR Science Council Secretariat.

Stemler, S. 2001. An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). [http://PAREonline.net/getvn.asp?v=7&n=17, accessed 22/9/2005].

Turnbull, B. 1999. The mediating effect of participation efficacy on evaluation use. *Evaluation and Program Planning*. 22(2): 131-140.

Weiss, C.H. 1980. Knowledge Creep and Decision Accretion. *Knowledge: Creation, Diffusion, Utilization.* 1(3): 481-404.

Weiss, C.H. 1999. The interface between evaluation and public policy. *Evaluation*. 5(4): 468-486.

APPENDIX I: FORM USED FOR EMAIL SURVEY OF CGIAR DONOR REPRESENTATIVES

Survey of Investor Demands for Ex post Impact Assessment (epIA) and Evaluation within the CGIAR

Please complete the included questions by answering in the grey fields, **save the file**, and return the document to D.Raitzer@cgiar.org as an **email attachment**. Thanks in advance for your input!

Please also note that this document does contain macros intended to simplify the process of entering survey responses, so for maximal ease of data entry, macros should be enabled in Microsoft Word (this setting may be modified under "Tools -> Macro -> Security"). Thanks for your cooperation!

Name:	_____
Title:	_____
Institution:	_____
Disciplinary Background:	_____

1) Please **rank** the following purposes for *ex post* impact assessment[4] in the CGIAR (1st = most important, 4th = least important):

a. select rank To assign credible benefit values to the impacts of research, and compare these to costs of investments.

b. select rank To demonstrate that research output is making significant contributions to desired development goals.

c. select rank To increase understanding of adoption processes and constraints to adoption.

d. select rank To help the CGIAR centres learn and improve through critical self-analysis.

Additional comments (optional):

[4] ***Ex post* impact assessment** may be defined as evaluation of progress towards mission-level goals facilitated by specific past activities (mission-level goals are food security and poverty-alleviation in the context of the CGIAR). Such is distinct from **ex ante impact assessment**, which projects potential benefits contingent on possible future courses of action.

2) Please **rate** the current range of CGIAR *ex post* impact assessments (for all topics), in terms of the following attributes.

	Poor			Fair				Good		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Credibility/rigour	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Comprehensiveness of programmatic coverage	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Relevance to your institutional needs	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Additional comments (optional):

3) How important are the following factors in determining whether epIAs are useful? Please **rate**.

	Not important			Somewhat important				Most important		
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Relevance of the programmatic emphasis assessed to the current priorities of your institution	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Recentness of the research output that was assessed	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Rigour of the assessment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Ease with which findings may be understood	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bearing of assessed indicators on development goals	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Magnitude and distribution of assessed benefits	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Additional comments (optional):

4) To what degree does authorship by an assessor **external** to the institution being assessed contribute to the credibility of an epIA if the methods employed are otherwise transparent and sound?

Not important			Somewhat important				Highly important		
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Additional comments (optional):

5) Please **rate** the importance of the following information sources, in terms of influencing your institution's funding decisions across CG centres and across project/programmes within Centres:

	Not important			Somewhat important				Most important		
	1	2	3	4	5	6	7	8	9	10
Annual reports of the CGIAR	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Annual reports of the individual CGIAR Centres	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reports from individual projects within CGIAR Centres	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
External programme and management reviews	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Internally-commissioned external reviews	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Ex ante projections of impact	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Output assessments (number of publications, trainees, etc.)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Impact assessments	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other (please specify)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Additional comments (optional):

6) Please **rate** the importance of the following indicators for demonstrating the impact of CGIAR research:

	Not important					Somewhat important									Most important					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Quantity of research outputs (number of publications, improved varieties, etc.)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Aggregate training performed	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Uptake of research outputs by other researchers and/or institutions	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Areas of adoption/implementation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gross economic surplus due to changes in unit costs/productivity	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Attributable reductions in consumer costs	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Distribution of returns to producers/resource managers/consumers	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Changes in environmental factors (erosion rates, changes in species diversity, etc)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Economic values of changes in environmental factors	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Risk/vulnerability reduced	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Empowerment/capacity increased	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Livelihood strategies improved	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Additional comments (optional):

7) Please **give a rough estimate** of the number of CGIAR evaluation studies that you have read over the past two years, by category, if possible:

Category	Number of Studies
Science quality/output assessments	
Adoption/uptake studies	
Ex post impact assessments	
Total	

Additional comments (optional):

8) Evaluation takes place within the context of limited resources, which means that it is not feasible to comprehensively evaluate every activity of a research Centre. In addition, certain forms of evaluation require greater resources than others. Please **allocate a total of 100 units among** the following types of evaluation, according the needs of your present institution.

	Cost per study	Number of studies
Science quality/output assessments (assessment of scientific methods applied and/or quantity and quality of research outputs)	10 units	
Adoption/uptake studies (assessment of whether output has been utilised by target audiences)	20 units	
Implementation studies (assessment of whether output has facilitated discernable changes in resource management practices)	30 units	
Ex post impact assessments (assessment of benefits realised due to changes in resource management practices attributable to output)	50 units	

Additional comments (optional):

9)

a) Has your institution found any CGIAR impact assessment to be characterised by particularly satisfactory rigour?

Yes: ☐	No: ☐

If readily accessible, please provide a citation for the study that your institution has found particularly satisfactory:

b) Has your institution found any CGIAR impact assessment to be characterised by particularly unsatisfactory rigour?

Yes: ☐	No: ☐

If readily accessible, please provide a citation for the study that your institution has found particularly unsatisfactory:

---

10) Please **select** the statement that **best** matches your view:

Natural resource management (NRM) research conducted by the CGIAR has had:

a. ☐ no *demonstrated* impact to date.

b. ☐ limited *demonstrated* impact to date.

c. ☐ moderate *demonstrated* impact to date.

d. ☐ extensive *demonstrated* impact to date.

Additional comments (optional):

---

11) Please **select** the statement that **best** matches your view:

a. ☐ There is no need to assess the impact of past NRM research, as environmental considerations alone are sufficient to justify such investment, as long as the quality of NRM science is demonstrated.

b. ☐ The impact of past NRM research should ideally be assessed, but where attribution is difficult, demonstration of uptake is sufficient.

c. ☐ NRM research, like any other form of strategic and applied research, should have mission-related benefits that can be measured and documented.

Additional comments (optional):

Supplementary questions (if time permits)

12) Activities of the IARCs are often pursued in a collaborative manner with a number of national/local partners, and impacts may be contingent on the delivery of complementary factors. Thus, the assessment of research impacts often involves attempts to attribute portions of the impacts to different institutional contributions. **Please rank the following means for crediting specific research activities by specific institutions** (1^{st}=most appropriate, 4^{th} =least appropriate)**.**

a) select rank Qualitative assessment of relative contributions through key informant interviews

b) select rank Proportion of key outputs produced by specific institutions and programmes

c) select rank Proportion of investment/staff time supplied by specific institutions and programmes

d) select rank All relevant institutional contributions/investments should be considered in concert, as any attempt to attribute individual complementary actions is inherently arbitrary.

Additional comments (optional):

13) For the following questions, please select the **maximum research investment** that would be justified by the following theoretical research impacts/outcomes, if such were convincingly demonstrated:

	Maximum Justified Investment (thousands of $)										
	<100	**100 - 200**	**200 – 300**	**300 – 400**	**400 – 500**	**500 – 600**	**600 – 700**	**700 – 800**	**800 – 900**	**900 - 1000**	**>1000**
Averted displacement of 10000 people from the agro-forest systems upon which they depend for their sustenance	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Averted deforestation of 10000 hectares for 5 years of primary tropical forest in a Southeast Asian developing country through catalysed improvements in corporate behaviour	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Catalysis of 500000 additional hectares of forest under Forest Stewardship Council Certification through improvements to national certification guidelines	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Contribution of a key paper on Forest Biological Diversity to the Convention on Biological Diversity (CBD) that formed the rationale for the creation of an Ad Hoc Technical Expert Group on Forest Biological Diversity, which, in turn, formed the basis for the CBD work program on forest biological diversity.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Mentoring for a total of 17 PhD student-years and 21 MSc student-years, 8 person-years of short training courses, and 27 trainers trained in adaptive collaborative management techniques.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Additional comments (optional):

14) Has your institution used any CGIAR impact studies when justifying budgetary decisions?

Yes: ☐	No: ☐

If you know which studies were used, please provide citations for the most frequently utilised studies or sources:

APPENDIX II: QUESTIONS FOR TELEPHONE INTERVIEWS OF CGIAR DONOR REPRESENTATIVES

Question group #1: Investment decisions and the role of impact findings

1. Please describe the number of staff within your organization who deal with the CGIAR, and the percentage(s) of their time spent on CGIAR matters.

2. What sequence of decisions is used to allocate funds to the CGIAR System, and at what level can year to year changes be significant? What are the roles of: a) politicians, b) aid management, and c) scientific community? At what level of detail (if at all) is the CGIAR explicitly mentioned in an appropriation bill approved by parliament? Against what organisations is the CGIAR competing for funds?

3. Please describe the kinds of information considered when 1) selecting overall investment in agricultural research; 2) specifying total allocation to the CGIAR; 3) distributing funds among CGIAR Centres.

4. How important is the performance of past investments, as compared with other considerations? How does your office appraise the performance of past investments? If there is little documented evidence of impact, how is the past performance appraised?

5. (If relevant) Please explain how the answers to 2-5 differ in the case of smaller restricted allocations, as compared with unrestricted funding decisions.

6. How important is it that *ex post* impact assessments are produced for your agency to continue to substantially fund the CGIAR (extremely important, very important, important, somewhat important, not important)?

Question group #2: Donor perceptions of impact and how they are informed

1. What areas of past CGIAR research have generated the greatest level of impact in terms of CGIAR goals, in your opinion? Please pick the top three of the following, and rank them (water management research, soil and crop management research, policy/social-science research, pest management research, livestock research, forest management research, fisheries research, germplasm conservation, ecological research, or crop genetic improvement).

2. What areas of CGIAR research will lead to the greatest future impact, in terms of CGIAR goals, in your opinion (please use the same categories as above, if possible)?

3. What sources of information have you used to make determination for 1 and 2 above (e.g., expert opinion, feedback from research partners, independent scientific publications, EPMRs, project reports, your own evaluations, centre or system level impact assessments, etc.)?

Question group #3: Determinants of epIA readership

1. What type of impact study (in terms of topic and method), if any, would be most successful in influencing key decision-makers to allocate more funds to the CGIAR from your agency?

2. Do you feel that enough *ex post* impact assessment is being conducted within the CGIAR for your agency's needs?

Question group #4: EpIA use in context

1. Is it possible for you to identify one or more *ex post* impact assessments that you found to be of particular value? Please name the ones that you remember. Why are these valuable?

2. (Assuming the answer is yes in the previous question) Did the study (or studies) contribute to your decisions regarding the CGIAR or other research institutions (for example, in terms of decision taken at AGM, funding priorities, funding levels, etc.)? If so, how? If not, why not?

3. Can you think of any other ways in which you've used the study (e.g. in the formulation of internal strategies, in decisions regarding other research institutions, in decisions regarding complementary programmes, in public-awareness material for your institution, in papers that you've written, etc.)? Did the study improve your understanding of any issues? Did it contribute to your confidence in the Centre's capacity?

4. Can you identify any ways in which the quality and relevance of studies that you read could have been improved, so as to better meet your needs?

Question group #5: Specific demands for impact-related information

1. Does an economic metric for investment efficiency (e.g., a rate of return or benefit-cost ratio) based on adoption and unit benefit data confer much additional value for you, as compared with adoption data alone? Why or why not?

2. Against which of the following should the CGIAR's performance primarily be assessed, considering that research contributions become much more difficult to precisely attribute as analysis progresses from a) to c)?:

    a) how widely research products are utilised, and the quality of the processes that resulted in the products

    b) productivity gains and other gross benefit levels stemming from the use and application of research outputs

    c) contributions to ultimate impacts on the welfare of the poor

APPENDIX III: DESCRIPTION OF THE WILCOXON SIGNED RANK TEST FOR PAIRED DIFFERENCES

The Wilcoxon Signed Rank Test for Paired Differences was used to compare different categories, using the respondent as the treatment block. This non-parametric test is appropriate for ordered categorical data where observations may be ranked, and is used to test the magnitude of the median difference in paired data. The null hypothesis for the test is that the distribution of differences between the pairs is symmetric around zero, which means that the two data sets are not significantly different. The alternative hypothesis is that one set of observations is represents a significantly higher median than the other.

The formula for the test statistic z is as follows:

$$z = \frac{T_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$ whereby n is the number of pairs, and T_+ is the sum of positive ranks.

T_+ is calculated by pairing the treatments for each treatment block, and subtracting one treatment from the other. The absolute values of these differences are ranked, of which the ranks of the positive values are summed to produce the T_+ value.

Under the two-tailed test applied here, the rejection region for the null hypothesis is either $z < -Z_{\alpha/2}$ or $z > Z_{\alpha/2}$.

This test was applied iteratively with $\alpha = 0.05$, so as to determine which categories differed statistically from one another for those questions with quantifiable responses.

APPENDIX IV: LIST OF ACRONYMS

ACIAR	Australian Centre for International Agricultural Research
ADB	Asian Development Bank
CGIAR	Consultative Group on International Agricultural Research
DANIDA	Danish International Development Agency
DFID	UK Department for International Development
EIARD	European Initiative for Agricultural Research for Development
epIA	*Ex Post* Impact Assessment
EPMR	External Programme and Management Review
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
GTZ	German Technical Cooperation Ministry
IADB	Inter-American Development Bank
IAEG	Impact Assessment and Evaluation Group
IARC	International Agricultural Research Centre
ICER	Internally-Commissioned External Review
IDRC	International Development Research Centre
IFAD	International Fund for Agricultural Development
KARI	Kenya Agricultural Research Institute
NRM	Natural Resources Management
PARC	Public Awareness Resource Mobilisation Committee
SDC	Swiss Development Corporation
SIDA	Swedish International Development Agency Corporation
SPIA	CGIAR Standing Panel on Impact Assessment
UNEP	United Nations Environment Programme
USAID	United States Agency for International Development

SCIENCE COUNCIL SECRETARIAT
c/o FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, Italy
e-mail: sc-secretariat@fao.org
Tel: +39 06 5705 6696
Fax: +39 06 5705 3298