Ministry of Foreign Affairs of Denmark
Danida

FROM DANIDA'S EVALUATION DEPARTMENT - DEVELOPMENT COOPERATION, SEPTEMBER 2007

ASSISTANCE TO DEVELOPMENT RESEARCH

EVALUATION STUDY 2007/1

Assistance to Development Research

September 2007

Centre for Economic and Business Research[1]
Porcelænshaven 16A
DK-2000 Frederiksberg
Denmark

[1] Valuable comments and suggestions for review have been provided by Martin Paldam, Henrik Secher Markussen, Holger Bernt Hansen, Agi Csonka, Eva Broegaard and Hanne Foss Hansen. All errors and/or omissions are, of course, the responsibility of CEBR.

Table of Contents

1. Introduction

"Knowledge for development", "Partnerships at the leading edge" and "Research for poverty reduction"[2] – the catchphrases are manifold when it comes to describing the relationship between research and development assistance. In recent years many development cooperation agencies have included "knowledge-based aid" as part of their answer to the challenges posed by the information revolution, aid fatigue and globalisation.[3] As a result, support given to research related to the promotion of economic development in poor countries is met with increased attention and demands.

Funding research to promote specific development agendas is, of course, not a new invention. The mission oriented public research programmes during the 1960s and 1970s such as the Green Revolution and the campaign to eradicate smallpox were early examples of research put to use on specific development challenges. The utilization of research for development purposes was further institutionalised in the 1980s with the establishment of strategic research centres such as the CGIAR centres and the World Bank Research Committee. The instrumental approach to research funding found its interim culmination with the introduction of knowledge-based aid during the 1990s – research was now seen as a central element of donor responses to their own perceived weaknesses.

As can be seen by the tables in Appendix C (see page 42) aid flows to development research are neither large nor growing rapidly.[4] It is, however, still a strategically important form of assistance for the following three reasons. First, the initiatives and

[2] "Knowledge for Development" was the title of the 1999/1998 World Development Report, "Partnerships at the Leading Edge" was the title of the Hernes report (Danida, 2001) and "Research for Poverty Reduction" was the title of the DFID research policy paper (DFID, 2002).
[3] See McGrath and King (2004).
[4] The source of data is the OECD CRS database. Given the dependence on donor self-reporting, different ways of categorising aid flows and different levels of detail, a potential bias and/or inaccuracy must be accepted using these data. Although it is neither possible to determine the direction of bias nor the size of error committed, these data still represent the best available indication of both the size and trend in development assistance allocated to development research.

results within this area can have significant importance for the efforts and strategic changes in other areas of development assistance. The discovery of effective antiretroviral therapies against HIV/AIDS and the subsequent drop in prices, leading to international pressure to provide this kind of treatment in developing countries is an example. Second, the assistance given to development research is significant compared to the public and private funding otherwise given to this area. Again, an example can be found within the field of health where the research into many tropical diseases would otherwise not be undertaken by, for example, the international pharmaceutical companies. Third, new knowledge and evidence has the potential to change the world. The process of improving and refining seeds initiated during the Green Revolution (see below) is an example of this kind of research. The assistance to development research may be small in magnitude but it holds great potential for change.

Key questions in this context are: Does it work and can we measure it? Can we, for example, measure and compare the impact of agricultural research with the effects of supporting a rural micro finance institution in a developing country? Questions like these are of interest to both the development cooperation agencies (and the public opinion in donor countries) that funds the research, the researchers who conduct it, and the people in the developing countries who are to benefit from it. The pressure to answer such questions is moreover accentuated by the recent trend towards performance-based management. Based on the idea that peer reviews and benchmarking can be used to create incentives to drive long-term performance improvements, performance-based management has permeated both private and public sectors. The result is increased pressure to clarify the impact of various measures (or initiatives) – including the funding allocated to development research.

Evaluating research funding is, however, not easy. Research is per definition an open-ended process that is not always amendable to evaluation frameworks. The often long time lag between when research is conducted and when research findings have a measurable impact on, say, maternal health in sub-Saharan Africa is another complicating factor. Moreover, research often occurs in international networks and alliances rather than in individual offices or laboratories. This introduces considerable issues of attribution, inter-dependencies and secondary effects, making evaluation more complex.

Further, development research is complicated by its global objectives and ambitions. In contrast to other forms of public research funding, the unit of measurement for impact evaluations of development research should be global not national. In addition, the impact of the research might never materialize because the necessary conditions (e.g. infrastructure, educational support, technical equipment, complemen-

tary inputs) are missing in the poor countries. One can argue that some of these deficiencies could have been taken into consideration and mitigated by the researchers. They can, for example, take the selling price and limited local purchasing power into consideration when developing a new seed. It would, however, not be fair to expect that a research project can solve all problems of missing infrastructure, local unrest and/or erroneous beliefs about the attributes of the new products. By implication, the absence of impact does not necessarily mean that the research programme has not met its objectives. As a consequence, evaluating assistance to development research as well as the impact of this research in developing countries is challenging.

Against this background, the objective of this report is to provide an overview of the state of affairs in the evaluations of support to development research. This is achieved through a survey of the existing evaluations. More specifically, what is evaluated and how? The survey subsequently forms the basis of an analysis of the evaluation methodologies, assessing their key strengths and limitations. The aim is to provide policy-makers, scientists and practitioners with a critical overview of the main evaluation methodologies available, outline their strengths and limitations, and set them in relation to existing and future challenges.

To the extent possible, focus will be on evaluation methodologies[5] rather than evaluation results or context. Hence, the report will not attempt to summarize or synthesize the findings of the evaluations, nor will it try to review the broader priority-setting and policy-formulation processes of the different development cooperation agencies. In terms of subject area, the survey will not include evaluations of support to higher education. Taking a long-term perspective on capacity building and researcher training, this delimitation could be argued to be artificial. Teaching and research activities are after all complementary processes at universities in most OECD countries. The situation is, however, different at most universities in developing countries where the majority of the limited funding and manpower available is reserved for teaching. As a result, the research tradition and infrastructure is very limited at developing country universities and (importantly for this survey) not (yet) related to teaching activities. Inclusion of assistance to higher education would thus complicate and expand the analysis without adding important perspectives.

Following this introduction, Chapter 2 will provide an outline of the issues related to supporting, monitoring and evaluating research activities. Special focus will be on

[5] Please note, that the survey is focused on evaluations of development research. We have only to a limited degree looked at evaluations of other types of research. This is, of course, not due to a lack of interest in these evaluations but rather an attempt to focus and make the survey manageable within a reasonable horizon.

how support to development research differs from other forms of research assistance. Subsequently, Chapter 3 contains the overview of existing evaluations, focusing on the identification of organisational, sectoral, or period specific characteristics as well as the different types of evaluation methods applied. Chapter 4 then moves on to present a methodological assessment based on the survey. What is the scope of existing evaluations? What are their strengths and limitations? How do they perform in terms of evaluating impact and what are the evaluation challenges? Finally, Chapter 5 concludes and provides perspectives for further studies and future issues.

2. Research: Policy Objectives and Evaluation Challenges

This chapter will first review the original policy rationales behind the allocation of development assistance funds to research projects. Subsequently, an overview of the difficulties associated with conducting evaluations of both research and development research[6] will be presented.

2.1 *Policy Objectives for Support to Development Research*

Based on the principle that every evaluation process has to originate from and respect the original policy rationale, it is useful to briefly summarize the policy objectives behind assistance to development research. It is possible to identify four (partly interdependent) policy objectives.

First, despite severe measurement problems there are a number of studies indicating that research has a high impact on economic development and poverty reduction in particular. A prominent example is the meta-evaluation of CGIAR impact studies. The study suggests that one of the key areas of CGIAR research, germplasm enhancement has contributed to food security and has helped lift millions out of poverty.[7] This positive interpretation of the role of research is mirrored in all the evaluations surveyed for this study. Although many of the evaluations raise the point that the methods for demonstrating the specific contribution of a research programme are partial and insufficient to overcome problems of timing and attribution, they all (explicitly or implicitly) acknowledge that the general case for the importance of research to economic growth and development is strong.

Second, the classical argument for any type of public financing of research is that under-investment in research otherwise will occur, because the social benefits from new technologies are difficult to appropriate by private firms bearing the costs of their discovery, or because imperfect capital markets inhibit firms from investing in

[6] "Development research" is here defined as research directed at understanding and/or promoting economic development and all aspects of society, science, culture and politics of poor countries.
[7] See: The World Bank Operations Evaluation Department (2004).

socially and privately valuable research projects (see: Arrow, 1962 and Stiglitz, 1989). As a consequence, there is a case for public administrations in poor countries to step in and fund research to remedy the market failure. This is especially true for research that benefits the poor as their frequent lack of both political and market power implies that they are easily overlooked. However, even if a government of a poor country acknowledges the problem, almost no developing country is able to undertake the necessary investments. This can be due to low administrative capacity, deficient funding and/or a crisis-driven agenda. If, for example, public servants are busy trying to address one acute crisis after the other, long-term decisions about funding research will be assigned a low priority.

An example of a programme which is based on this kind of argumentation is the WHO special programme for research and training in tropical diseases (see: WHO, 1998) – a programme which invests in the realisation of a global public good, the eradication of tropical diseases that are overlooked by the international pharmaceutical industry or outside the research of developing country domestic research capacity.

Third, the aforementioned introduction of knowledge-based aid assigns a central role to research as a means to enhance the effectiveness of development assistance. The envisioned quality improvements are believed to come from two different sources. First, research and researchers can help design new and improved forms of development assistance through increased interaction with the multiple stakeholders involved in designing and executing aid programmes. Second, development cooperation agencies can get better at acknowledging and working with the knowledge they already possess within their organisations. While the latter has an organisational and more internal perspective, the establishment of better and more relevant contacts between researchers and aid professionals is often of direct relevance to research funding. More specifically, development cooperation agencies may choose to allocate research funding to areas of specific relevance to their aid projects. If, for example, an agency is very active within aquaculture, it may choose to call for research to be conducted within this area.

Fourth, and frequently related to the enhancement of aid effectiveness, is the desire to strengthen the domestic (donor) resource base. The question of whether development research should be undertaken by researchers based in recipient or donor countries is a debated issue. While most development cooperation agencies increasingly demand that cooperation with developing country researchers should be established, the Netherlands is the only donor who from 1992 to 1998 broke completely with the current tradition by: (i) putting demands in developing countries before Dutch and other western universities supply, (ii) emphasizing multidisciplinary and

problem based research with increased interaction between research and policy, and (iii) giving Southern partners ownership over the research. As outlined in the recent evaluation of Dutch assistance to development research (Nuffic, 2007), a combination of lack of local (developing country) capacity to conduct research, long preparatory periods, programme management issues, and the shift towards a sector-wide approach, however, lead to a gradual revision of the Dutch approach. A key lesson emerging from the evaluation is that the quite rigid and dogmatic interpretation of demand-orientation that was applied in the first period from 1992 to 1998 in many cases led to isolation and sub-optimal performance of the funded research programme. Accordingly, the revised Dutch research policy from 2005 "*suggests ways to improve relations with the Dutch academic world*" (Nuffic, 2007).

Although few other development cooperation agencies have experienced policy shifts of this nature, the degree of mandatory involvement of domestic (donor-country) researchers still varies considerably across the different agencies. Based on the argument that the source of the research expertise is not material to the achievement of DFID's objectives, "*DFID believes that it has no responsibility to maintain UK research capability per se.*"[8] At the other end is, for example, the Danish support to research. Although Danida since the publication of the Hernes report (Danida, 2001) has worked to gradually increase the involvement of developing country researchers in all phases of the research projects, Danida still acknowledge (as do all the Scandinavian countries) that strengthening the domestic resource base also is among the policy objectives. The motivation is twofold: To strengthen the national understanding of poverty and development issues, and to safeguard the effectiveness of the other forms of development assistance of the agencies.[9]

Although not always explicitly acknowledged by the development cooperation agencies, the four policy objectives listed above are, however, not always achievable simultaneously. Possible conflicts of interest include:

- Donor country university systems increasingly emphasize the importance of academic output in peer reviewed journals. This can sometimes be difficult to combine with increased demand orientation towards developing country researchers, particularly if capacity building and training is involved. As a result developed country researchers can be reluctant to participate and/or limited capacity building take place.

- The knowledge requested for "knowledge based aid" may be of a more operational nature and/or have a shorter time horizon than the knowledge that re-

[8] See: DFID (2002) page 41.
[9] See, for example, Sida (2006e) and Sida (2006g).

searchers may be able to give.[10] Again, the result could be that researchers are reluctant to participate and/or that the knowledge generated is not sufficiently operational to be used as basis for aid programmes.

- The orientation towards recipient country needs and demands should ideally be shaped by and depend upon local researchers. However, although local ownership and demand orientation has become part of the international cooperation consensus, it is most often incompatible with the limited research capacity and infrastructure in developing countries. In the recent evaluation of the Netherlands' research policy (Nuffic, 2007) the gap between local capacity and the need to involve local researchers to find a feasible solution is termed "*the development paradox*".

- Perhaps in part as a consequence of the conflicting policy objectives, most development cooperation agencies utilize several different channels/instruments to support development research. This includes, for example, fellow-ship programmes to support capacity-building and research training in the South, grant distribution schemes to support the national resource base, and specialized research centres to support in-depth and/or field-specialized research. The various instruments do, however, not reflect a fully differentiated approach of following different policy objectives with different instruments. Several of the listed policy objectives typically enter as selection and decision criteria for each of the instruments. Applicants to national grant schemes are, for example, often requested to cooperate with researchers in developing countries as well as to describe the relevancy of the planned research for national development assistance. This is important in an evaluation context, as the underlying policy objectives and the potential tensions that exist between them both should be taken into consideration when deciding on the evaluation methodology.

2.2 *Evaluating Research*

Without a proper understanding of the challenges related to evaluating research, the analysis could end up confusing generic problems with project related obstacles and vice versa. As a consequence, the remainder of this chapter will contain a brief overview of the challenges associated with evaluating the impact of research. This includes: Quality versus impact consideration, timeframe of impact versus evaluation horizon, attribution of results, unit of analysis, and implementation context. While the first three are generic to all types of research evaluations, the challenges

[10] McGrath and King (2004) outline how the perception of knowledge embedded in knowledge-based aid at times come close to being the information and organisational practices already present in development cooperation agencies.

related to defining the unit of analysis and/or a deficient implementation context are of specific importance to evaluations of development research.

Quality is usually interpreted as the technical quality of the research conducted. Although this is often juxtaposed with the assessed policy impact of the same research, it would be a mistake to assume that quality and impact are substitutes. As Professor John Toye commented in a special survey on building research capacity in Africa (DFID, 2002): "*It is useless to produce good studies that are ignored, and it is actually counter-productive to produce bad studies that are influential.*" Scientific quality and impact must be considered in conjunction. If possible, it would, for example, make sense if an evaluation of the scientific quality of a research programme was accompanied by an assessment of the practical impact of the same. Not least if the evaluation finds that the research programme has been sub-standard, in which case the development co-operation agency (by the same logic) would have a responsibility to seek to counteract the potentially damaging effects resulting from the flawed research. Although the latter implication for many would be to take the argumentation too far, the premise that one should attempt to evaluate policy impact and scientific quality simultaneously must be maintained as a guiding principle for evaluations of development research. While the question of impact has become increasingly important and most agree that development research has the potential to contribute significantly towards overall economic growth and development, most of the connections between research and effects are diffuse and indirect.[11] Consensus among evaluation specialists[12] is that there are "*no short-circuiting approaches which can reduce complexity so as to allow the measurement of impacts to become a routine business*" (Rip, 2001).

The issue of timeframe refers to the fact that there is often a long time lag between the point in time when the research is conducted to the point in time when the impact and implications of the research is felt by the intended beneficiaries. In the absence of a striking breakthrough it will, for example, most likely be another ten to fifteen years before an AIDS vaccine will be commercially available. In addition, most research processes are open-ended and by nature unpredictable. Staying with the example of the AIDS vaccine, it is thus unlikely that the vaccine that would hit the market a decade from today would be 100% effective. As a result, efforts will undoubtedly continue over a prolonged period to identify and develop even better vaccines. The long and often open-ended processes represent a particular challenge for evaluations. How can one identify the specific contributions of providing short term funding if the process is nowhere near completion?

[11] See Scott (2003) for a review of the links between research and policy
[12] See, for example, Feller (2001), Georghiou (2001) and Rip (2001).

According to some critics, the problems associated with showing measurable result in the short-term has lead some development cooperation agencies to prioritise research more likely to demonstrate short- to medium-term results. Former Director General of the International Food Policy Research Institute (IFPRI), Professor Per Pinstrup-Andersen has detected a tendency towards this type of behaviour from the development cooperation agencies and warns that it may increase as the 2015 deadline for the Millennium Development Goals approaches without the poor countries making sufficient progress towards achieving them (see Mariegaard, 2006). The response to this criticism by the development cooperation agencies is twofold. First, it is noticed that the distinction between short- and long-term perspectives is both problematic and difficult to make operational. Second, the allocation of funds to research is based on other criteria than the time horizon of the research. More specifically, assistance to development research is based on the above-mentioned policy objectives; not least the extent to which the research supports the achievement of other developmental goals such as economic growth, gender equality, environmental sustainability, etc. Whether or not the increased emphasis on impact and measurable outcomes over the past decade has resulted in a shift of priorities towards favouring short-term research is unfortunately impossible to determine without a special analysis of research allocations.

The problems related to attribution are in part related to the open-ended, long-term nature of research, and in part to the trend towards research increasingly being undertaken in international networks and teams. The related problem of identifying the specific output associated with a particular allocation of support to a research programme is called the "project fallacy" (see, for example, Georghiou 2001). The fallacy in question is committed by the evaluators (and policymakers), who wrongly assume that their contribution to a larger programme will (or has) result in uniquely attributable outcomes. Separating the effects of funding part of a research project quickly turns into a complex exercise that inevitably demands that the evaluator assesses the relative weights of the contributions of different organizations and activities. This is, of course, a problem which is shared with evaluations within other fields, but the often global and dispersed nature of many research project imply that it is particularly acute here. Finally, the problem of additionality needs to be addressed. Would the research have been undertaken irrespective of the support (input additionality), would it have been conducted in the same manner (behavioural additionality), and would it have provided the same results (output additionality)?

While the globalisation of research is a shared trend for all types of research, development research differs in an important respect from other types of research. More specifically, the intended beneficiaries of the research frequently live in other countries and/or far from the universities hosting the researchers. Of course, researchers

conduct trials and fieldwork among the poor to collect data/information to better target and include the intended recipients, but in the end, the resulting innovation and/or new knowledge need to be disseminated to the poor – most often without the assistance of the researchers. By implication, measuring the impact among the end beneficiaries represents a special challenge that needs to be considered in the scope of the evaluation.

Evaluating impact in resource poor settings is, however, beset with additional hurdles as necessary conditions such as physical infrastructure, access to financial markets, distribution systems and complementary inputs frequently are deficient or missing in developing countries. It would, for example, be unfair to conclude that a research project to develop a new drug against malaria is unsuccessful because an inadequate drug delivery infrastructure fails to distribute an otherwise affordable and effective new drug. As observed by the World Bank "*the absence of evidence of impact does not imply the absence of impact*."[13]

In conclusion, this chapter has presented a number of specific challenges to the evaluations of research – some of which represent a particular challenge for evaluations of development research. The following two chapters will first uncover how existing evaluations have met these challenges (Chapter 3) and subsequently discuss the strengths and limitations of the different approaches (Chapter 4).

[13] See: The World Bank (2004a)

3. Overview of Existing Evaluations of Development Research

The search for the relevant evaluations for this study was guided by three criteria. First, the included evaluations should be broad-based and comprehensive in order to avoid having to generalize from too many specialized sector/field specific analyses. Second, the date of publication should be within the past decade to ensure that their findings and the context in which they were conducted are still relevant. Third, the evaluations should be of support to research and development related to understanding and/or promoting economic development in poor countries. This means that all evaluations related to support to higher education are excluded.[14]

3.1 *Overview of Included Evaluations*

In total the search resulted in 23 evaluations that meet the three criteria. The 23 evaluations originate from seven different development cooperation agencies (five bilateral and two multilateral) and span the entire period from 1997 to 2007. Some of the evaluations were based on a number of sub-evaluations/case-studies, which for the most part are not included in this survey. The same can be said about a number of evaluations of research support to a single country, which were excluded with reference to the decision to focus on comprehensive and broad-based evaluations (the first of the three selection criteria).

Given the comprehensive nature of the search, the presented 23 evaluations are believed to provide a fairly accurate and up-to-date overview of the present state of evaluations of development research. On the basis of the policy objectives defined in Chapter 2.1, it is possible to identify the following three categories of programme targets: (i) support to programmes directed at supporting the generation of new knowledge and evidence, (ii) support to research capacity building in developing

[14] Appendix B contains a more detailed account of the search criteria as well as of the evaluations excluded because they did not meet the criteria.

countries[15], and (iii) support to programmes directed at strengthening the framework for research.

Not surprisingly, the majority of the evaluated programmes (18 of the 23) contain elements of research funding. Little more than half (12 of the 23) of the evaluated programmes have elements of capacity building, whilst only a minority (4 of the 23) of the evaluated programmes include a focus on research infrastructure. The table below depicts the distribution across donors, periods covered in terms of publication dates of the evaluations, and whether the focus of an evaluation is comprehensive (i.e. covering all research support activities by a particular development cooperation agency) or it is focused on the activities within a particular sector/field:

Donor	Number of evaluations (period covered)	Comprehensive or focus on a sector
Danida (DK)	4 (1996 - 2007)	Both
DFID (GBR)	4 (1998 - 2005)	Both
Norad (NOR)	4 (1998 - 2005)	Both
Nuffic (NLD)	1 (2007)	Comprehensive
Sida (SWE)	7 (2006)	Comprehensive
WHO	2 (1998 - 2006)	Sector
World Bank	2 (2004)	Both

Table 1. Evaluations in survey according to donor and focus

Turning next to the methods used in the 23 evaluations, it is possible to identify the following three main tools/approaches: Expert reviews, case studies and bibliometric studies. A rate of return approach was used as part of the meta-evaluation of the CGIAR institutes (demonstrating extraordinary high rates of return to germplasm research) but otherwise not. Whether the limited use of rate of return methodologies can be attributed to lack of data and/or the validity and sensitivity problems that surround rate of return formulations is not clear. DFID (2005c) provides an overview of what is known about rates of return to research and assess key evidence that has been presented on agricultural and health research in particular (none of the references for the DFID study are, however, formal evaluations).

[15] Support given with the objective of strengthening the domestic resource base of development cooperation agencies is not categorized as capacity building but as direct research support.

3.2 *Evaluation Methodologies in the Survey*

Expert review is the most widely applied approach used by both government and industry evaluators throughout all fields of research and science in both developed and developing countries. According to a recent comprehensive survey of research and technological evaluations in the European Union (European Commission, 2002) "*current EU evaluation practices strongly focus on monitoring and less on impact assessment, and rely mainly on expert panels*." The definition of expert review applied here is quite broad, as the "experts" can include: academics, consultants, administrators, potential end-users of the research, and representatives of the agency commissioning the evaluation. The common denominator is that they rely on a combination of their expertise and the already available material. The latter can include prior evaluations, project documents and interviews with internal and external stakeholders.

As described here and in the absence of complementary analyses, expert reviews are to some extent measurement without theory. This is, however, a simplification as considerable differences exist in both the composition of the panel and how systematic and rigorous the analysis of the existing material is. This is the topic of Section 4.1, which looks at expert reviews in more detail.

Expert reviews are also assumed to be different from the traditional peer review conducted by scholars in a field. The latter is dominated by researchers (peers) and usually only focus on assessing the scientific quality of the research conducted. An expert review can (and typically will) have much broader terms of reference, looking at programme management and/or the economic and social impact of the research.[16] In conclusion, the key characteristic of expert reviews is the formation of a group of demonstrably independent and knowledgeable persons, who consider and analyze the available information.

Case studies are in-depth studies of specific projects, programmes or outcomes. They typically involve direct observations and/or field visits, and can be made up of many potential methods based on both quantitative and qualitative information. Often case studies are applied as exploratory and descriptive means of investigation, frequently to supplement expert reviews of the available evidence. Case studies are, as a consequence, an option in settings where pre-existing, theorized understandings are not available. Although case studies in most evaluations are used as a complementary technique used to supplement an expert review, there are also examples of

[16] Other analyses use the terms "expert review" and "peer review" interchangeably – hence, the definition used in this paper is not universal.

evaluations where case studies form the main body of analysis (see, for example, the World Bank, 2004 and Nuffic, 2007).

In comparison, the use of bibliometric indicators is a more rigorous and well-defined technique focusing principally on constructing indicators based on data/information about the publication and subsequent uptake of scientific publications. This could, for example, include indicators of science and technology activity like the number of citations or relational indicators like the co-occurrence of words. Following a modest beginning at the start of the last century, bibliometrics has been established as a distinct field of study and found widespread use in evaluations of research programmes across a wide range of scientific fields.[17] This is in line with the trend towards basing opinions and subsequent decisions on criteria that lend themselves to quantitative evaluation. Bibliometric indicators are also increasingly used as an input for researchers and research departments for job application/promotion procedures. Bibliometric indicators have thus been applied as an input for the evaluation of everything from individual researchers[18] to national research programmes.

As noted by Godin (1996) bibliometric indicators do "*have limits because they normally include only the natural sciences, engineering and the biomedical sciences*". Most observers also agree that comparisons across scientific fields (who may have different publication traditions and practices) should be avoided. In addition, the definitions of scientific quality (whether it is measured by the number of citations, ranking of the journal, etc.) and impact measurements have been disputed and much discussed. As a result, Gauthier (1998) concludes that bibliometric indicators should be used as a complementary input to more traditional methods such as expert and/or peer reviews. The perception of bibliometric indicators as a valuable way of getting a systematic and new perspective on research outcomes, which should be used in combination with in depth analyses of the research and researchers in question appears to be the predominant view.[19] To quote Boyack and Börner (2005) no one suggests "*replacing experts by automated techniques*".

This is far from the case looking at the existing evaluations of development research as all 23 evaluations (see Table 2 below) rely on expert reviews. It is important to note that the reliance is not only related to summarizing or synthesizing the evidence available from, for example, case studies. The experts do in all cases perform

[17] See Okubo (1997) for a presentation of the history and development of Bibliometrics.
[18] Please note that Gauthier (1998) concludes that bibliometric tools are not considered a valid method of measuring the productivity of individual researchers. This is, however, a contested issue, as others such as the US Committee on Science, Engineering, and Public Policy (2001) believe that the indicators are better suited for individuals than for institutes, universities or even countries.
[19] See, for example, Okubo (1997), Boyack and Börner (2003) and Leimu and Koricheva (2005).

independent analysis and draw conclusions which go beyond what can be concluded on the basis of other techniques applied. In sum, expert review is standard although it is only allowed to stand alone in a minority (6) of the 23 evaluations – many of which are smaller evaluations that are part of a comprehensive evaluation (the Sida evaluations) or a special case (the Norad working paper).

	Peer review	Bibliometrics	Case studies	Research funding	Capacity building	Infrastructure	Strategic evaluation
Danida (1996)	*		*	*			
Danida (2001)	*		*	*	*	*	*
Danida (2006)	*					*	
Danida (2007)	*	*	*	*	*		
DFID (1998)	*		*	*	*		
DFID (2002)	*		*	*	*	*	*
DFID (2005a)	*		*	*			
DFID (2005b)	*		*	*			
Norad (1998)	*		*		*		
Norad (2000)	*		*		*		
Norad (2005a)	*			*			*
Norad (2005b)	*		*		*		
Nuffic (2007)	*		*	*	*		*
Sida (2006a)	*					*	
Sida (2006b)	*		*	*	*		
Sida (2006c)	*						
Sida (2006d)	*			*	*		
Sida (2006e)	*			*			
Sida (2006f)	*		*	*			
Sida (2006g)	*		*	*			*
WHO (1998)	*	*	*	*	*		
WHO (2006)	*		(*)[20]	*	*		*
World Bank (2004a)	*		*	*	*		*
World Bank (2004b)	*		*	*			

Table 2. Evaluations in survey according to methodology and policy objective

[20] The TDR programme had just prior to the 4th external evaluation been evaluated as part of the World Bank evaluation of Global Programmes (World Bank, 2004a).

The reliance on expert reviews should, however, not come as a surprise given that this is also the dominant and preferred evaluation technique in evaluations of all types of research and development (see European Commission, 2002), and in all evaluations of development assistance (see Hoebink, 1996).

About a third (7 of the 23) of the evaluations are strategic in nature. All seven evaluations state that they deliberately adopt a forward-looking and strategic perspective in order to understand future challenges and inform strategic planning. This is in accordance with one of the recent trends within research evaluations noted by Rip (2001): the interest in strategy and using evaluations as decision support. This gives a more direct political dimension to the evaluations and introduces a possible inherent conflict between looking backwards (assessing the impact and relevance of past actions) and looking forward (orientating programmes towards new policy directions and trends).

As can be seen from Table 2, three quarters (18 of the 23) of the evaluations supplement the expert review with in depth case studies and two evaluations (WHO 1998 and Danida 2007) include a complementary bibliometric analysis to support the expert review – both of these evaluations are within the biomedical/health sector.

The overview of methodology and policy objectives reveals only limited correlations between the two dimensions. The correlations that do exist are not particularly surprising: Comprehensive evaluations only to a limited degree rely on case studies and evaluations of research infrastructure support do not employ bibliometric methods. Neither should come as a surprise given that it would be prohibitively expensive to conduct sufficient case studies for a comprehensive evaluation, and assistance directed at research infrastructure (ITC technology and/or the formation of networks) rarely include scientific publications as an (immediate) objective.

A systematic difference, however, appears when it comes to the period covered by the sector/field evaluations and comprehensive evaluations, respectively. Here, it is found that the comprehensive evaluations appear to cover a shorter period of time (typically three to five years) than the sector/field evaluations (which typically cover a period of five to ten years). While it from an evaluative perspective appears illogical that the comprehensive evaluations do not cover at least a period of similar length as those of sector/field programmes, it is possible to identify a number of plausible explanations of this difference. First, a resource constraint may kick in, as the amount of material necessary for comprehensive evaluations covering a long period may be both too expensive and unmanageable. Second, the complexity of a comprehensive evaluation, having to analyze (and perhaps compare) activities across

scientific fields, projects and countries is already considerable. Third and related, the sector/field specific programmes may allow for a greater degree of traceability, enabling a longer evaluation horizon.

In their review of the feasibility of and approach to impact evaluation of the WHO special programme for research and training in tropical diseases, Michaud and Reich (2005) call for an extension of the timeframe from five to ten years. This would in their view "*increase the likelihood of observing a measurable impact, because drugs, diagnostics and intervention strategies often take years if not decades to influence population health (in developing countries).*" Most likely in recognition of the resource and traceability constraints involved, Michaud and Reich however go on to propose a differentiated timeframe that is adapted to the nature of the research. This would, for example, entail having a longer time frame for fundamental research than for research specifically directed at improving government policies of a special area/sector.

The time period covered by the evaluations is interesting, not least due to the aforementioned considerable lag between time of initial research and time of impact (see chapter ›). It is, however, important to keep in mind that the link between research and policy or evidence and practice only in the rarest occasions (if ever) is a linear process. The work conducted by the RAPID (Research And Policy In Development) programme at the UK Overseas Development Institute[21] is uncovering more dynamic and complex perceptions of the process. The RAPID research points to a two-way process between research and policy, shaped by multiple relations and reservoirs of knowledge. It is, however, still an emerging field, and it is often difficult to explain why some policies are open to research and evidence while others remain evidence-averse. A similar conclusion is reached by Diane Stone[22] who provide a complex perspective upon the connections between research and policy, indicating that there are many possible routes to 'bridging' research and policy. In addition, the research conducted by Diane Stone also emphasizes that although research and knowledge is part of the solution to many development problems it is not of itself a panacea.

The uncertainty and lack of data characterizing this field is also reflected in the evaluations, which almost all call for better indicators and/or a more quantitative basis for evaluation – a recommendation which should be contrasted with the relatively limited use of bibliometric indicators.

Despite the lack of concrete evidence, all evaluations explicitly or implicitly acknowledge that the general case for research having a positive effect upon economic

[21] See www.odi.org.uk/rapid
[22] See, for example, Stone (2002) and (2004) as well as Stone, Maxwell and Keating (2001).

growth and development is strong.[23] At the same time, a number of the evaluations also raise the point that the methods for demonstrating the specific contribution of a research programme are partial and insufficient to overcome problems of timing and attribution. Appropriately, the strengths and limitations of the methods used in the evaluations is the topic of the next chapter.

[23] The studies that attempt to estimate rates of return of research (see: DFID, 2005c) find a robust and positive relationship between spending on research and development and economic growth. In addition, both the theoretical models and cross-country empirical studies indicate a significant and positive effect of research and development. See, for example, Romer, 1990; Aghion and Howitt, 1996 and Coe *et al*, 1997.

4. Assessment of Evaluation Methodologies

The assessment of the three evaluation methods (expert reviews, case studies and bibliometric indicators) will look at two main characteristics. First, how credible is the evaluation methodology? And second, how operational is it? In this context "credibility" refers to the methods used to secure the quality and objectivity of the evaluations, whilst "operationalibility" is a reference to the information and data required for the method to work as well as the level to which the method can produce relevant and timely information for the policymakers that commissioned the evaluation. The latter dimension is included in acknowledgement of the increasingly strategic and forward-looking nature of evaluations (see Chapter 3). The methods are considered in turn.

4.1 *Expert Reviews*

As mentioned earlier, all evaluations in the survey rely on expert reviews. This should come as no surprise given that expert review is a well-recognized and widely used technique in the evaluation of research and development support.

4.1.1 Credibility of Expert Reviews

The credibility of expert reviews rests on two pillars: the perceived fairness, standing and competence of the "experts" and the amount and level of detail of the material they look at. Turning first to the issue of selecting the experts, the strength of expert reviews is, of course, that the experts can be chosen to be uniquely familiar with the standards, context, history, and trends of a research area. As a consequence, they are well qualified to both assess the quality of the research conducted and to recommend improvements.

There is a trend towards having more academics and fewer consultants on the strategic and comprehensive evaluations. This is, for example, the case with the Surr report (DFID, 2002), the Hernes report (Danida, 2001), and the 3rd and 4th external reviews of the TDR (WHO, 1998 and WHO, 2006). In all cases, a group of promi-

nent academics are given the dual task of evaluating and suggesting future policy/programme directions. The reasoning behind this choice of evaluators is most likely that the credibility of an evaluation rests on the perceived competence of the panel members. As a result, participation and endorsement(s) from credible and trusted public researchers will enhance the overall credibility and reception of the evaluation.

Georghiou (1995) describes the process undertaken by the European Commission to improve evaluation procedures and outcomes. Given the extensive reliance on expert reviews, special consideration has been given to this method. Here it was found that "*heterogeneous panels were considered more successful, consisting of experts in the technical field under consideration, experts from different fields, users of the results and people with management and science policy experience. On some occasions a specialist in evaluation would also be selected.*" This was based on a recognition that the tasks required by a member of an expert panel are both many and complex. The tasks can, for example, range from assessing the quality of the research conducted to estimating its impact in developing countries and suggesting improvements in the management of the programme. Few people will be able to undertake all such tasks at an international level. The recommendation to compose a group of heterogeneous experts is thus a response to this challenge.

This leads directly to the discussion of the limitations of expert reviews vis-à-vis evaluation credibility. Apart from being able to undertake the evaluation in a satisfactory manner, the key issue for the selection of experts is that participants must be demonstrably independent. Expert reviews should in other words be conducted by persons who have technical expertise in the subject being reviewed but who are professionally independent of the program under review. However, none of the evaluations in the survey reveal how the independence of the reviewers is established (or for that matter how they were selected). For some of the strategic and forward-looking evaluations, the problem is exacerbated by the participation of persons from the evaluating agency (see, for example DFID, 2002 and Danida 2001). Although the people employed by the evaluating agency most likely were included to provide the necessary inside knowledge to the strategic recommendations, it is still a cause of concern as the strategic evaluations also are based on an assessment of past programmes and policies. In other cases, the experts are external consultants, which given the sometimes quite small international communities of specialized experts does not automatically remove the problem. It is in all cases unclear whether the validation of expert independence has taken place (but not communicated) or whether it was not made. Overall, it would enhance the credibility of the expert reviews if the selection criteria and process for the group of experts were explicitly and publicly articulated as part of the evaluation report.

4.1.2 Operationalibility of Expert Reviews

The strong side of the expert review is clearly that the experts can analyze all forms of data/information. They are, moreover, able to include policy recommendations and can, provided that the selection of experts is well thought out, foresee and seek to accommodate possible conflicts of interest. The amount of material taken into account by the experts obviously varies with the size of the programme under evaluation. The World Bank evaluation of the CGIAR centres (World Bank, 2004b) reviewed over 700 documents (of which more than 200 were academic articles and publications). In addition, interviews were conducted with more than 100 stakeholders and more than 200 filled out a questionnaire for the same evaluation. At the other end, looking at a considerably smaller programme, the Sida evaluation of bilateral research support (Sida, 2006b) was based on a team of experts updating four previously undertaken case studies. This is mentioned only to illustrate the versatility of the method and not in any way to indicate that the magnitude of material considered for either of the two evaluations was "inappropriate".

The ability of the expert review to accommodate almost any type of incomplete and/or missing information can, however, also be a cause of concern. Provided one can get the experts to agree to undertake the task, they can conduct the evaluation almost regardless of data, resource and time constraints. This will, however, decrease the quality of the evaluation. In the World Bank evaluation of the Global Development Network (Sarna, 2004) it is, for example, stated: "*Basic information on financing arrangements and the roles and responsibilities of the Bank as a partner is not always clear or easily available. Hence, this OED evaluation has had to explore new and often untested ground attempting to clarify concepts and then apply them.*" Similarly, the preface of the evaluation of the Danish Research Networks (Danida, 2006) states that: "*while the terms of reference were comprehensive, the time available was hardly commensurate with the task. The review of six research networks was made within the time span of merely three weeks.*" If these (and similar) disclaimers related to time and resource constraints (and the resulting negative effects on evaluation quality) are taken into account by the end-users of the evaluation – the ability to overcome both information and resource constraints is a strength. Otherwise, it must be regarded as a limitation.

Among the more clear-cut limitations of the expert review in terms of operationalibility is the danger that extensive reliance on interviews and questionnaires can cause a tendency to adopt and endorse ideas already in circulation among programme managers and other stakeholders. This is especially a problem for evaluations that rely solely on existing documents and interviews. Missing the opportunity of having new and external information and faced with possible peer pressure and/or the threat of not getting further contracts to perform evaluations, the ex-

perts might choose to adopt and endorse the prevailing wisdom. Two measures can mitigate this tendency. First, the list of interviewees should be balanced and reflect all interest groups and political views. Second, the expert review should be complemented with other forms of analysis (case studies, background surveys, bibliometric indicators, and/or other forms of empirical analysis). Whilst the latter measure is adopted by three quarters of the evaluations in this survey, it is not possible to gauge the extent to which the former measure is taken into account. Again, lack of information and transparency surrounding the evaluation set-up preclude an assessment of whether this is the case.

A final limitation of the expert review in terms of operationalibility is the lack of guidelines and/or information about aggregations. The final evaluation reports are often based on a multitude of different sources, but little (if any) guidance is provided about the relative weights assigned to the individual factors/studies. This makes the underlying dynamics unclear and may leave the commissioning agency with a very difficult task of prioritizing between suggested policy options. This is not mentioned to imply that the experts should define the weights and causal relationships between all relevant factors. This would be neither possible nor desirable. Indeed, a high degree of aggregation may be chosen because individual program activities are not easily linked to budgetary items or because specific mechanisms of decision are too numerous to discern. It should, however, still be possible to provide general guidelines and assessments – both of which would improve the transparency and potential usefulness of the evaluations.

4.2 *Case Studies*

As mentioned in Chapter 3, three quarters of the evaluations in this survey conduct case studies, most of them with the objective of assessing the impact of the research support.

4.2.1 Credibility of Case Studies

As is the case for expert reviews, conducting case studies is an adaptable method, which can be applied to all forms of research support from research networks as in the Norad evaluation of regional research networks (Norad, 1998) to the Sida evaluation of ITC support to research institutions in developing countries (Sida, 2006a). Case studies can also be conducted almost regardless of the quantity and quality of the information available.

Similarly, the key determinant of the credibility of this method lies in the appropriateness and relevance of the cases selected. In other words, if the results from the specific cases appear relevant and transferable to other settings and environments, the results can convincingly be generalized to apply for the entire programme. The

suitability and not least the number of case studies, of course, depend upon the heterogeneity of the individual components of a programme and the context in which they have been implemented. In some cases, where the different projects/programmes subject to evaluation are very different in character and context, the sample may be very large. This was the case with the World Bank evaluation of the Bank's approach to global programmes[24] (World Bank, 2004a) – here more than a third (24 of 70) of the global programmes where selected for a more detailed study.

Another strength associated with the use of case studies is the potential to uncover in greater detail the effect of the support given. Although descriptive in nature and consequently generally not conducive to the testing of causal relationships, case studies do provide a closer look and offer potential explanations to the relative success/failure of individual components of a project. In addition, case studies are a means to assessing the impact upon the intended beneficiaries of the research, namely the poor people living in developing countries. The lack of data and indicators, moreover imply that case studies, despite their imperfections and the difficulties associated with both uncovering causal relationships and generalizing results, can be the only method available to evaluate the impact of the research. Add to this the signal value of leaving your desk to actually look at and assess the results, and you have an explanation of why three quarters of the evaluations in this survey rely on case studies.

The key limitation of case studies with regards to credibility has already been mentioned: How general are the results from the specific case(s)? This problem in part falls back on the selection of the cases. Whether a proper sampling strategy and/or sampling criteria have been applied in the selection of cases is, however, not possible to tell as none of the evaluations mention it. Some, like the Norad evaluation of research networks (Norad, 1998), mention that a resource constraint prevented the evaluators from visiting all – but no description is given of how the resulting selection of cases was made.

The resource and time constraint is another obvious limitation of the method. Conducting case studies is expensive, and it might not be feasible to get a representative selection within the evaluation budget. In this context, it is unfortunately often the end-recipients that are deselected as the costs of conducting visits to developing countries are prohibitively higher. This could point towards conducting fewer evaluations with a better and more representative coverage of cases.

[24] Please note that far from all of the global programmes were directed towards funding development research.

A related issue is the question of who should carry out the case studies. Whilst the majority of the evaluations in this survey lets the experts conduct the case studies, this might not always be the best idea. As pointed out by Georghiou (1995) expert evaluators are not necessarily experts in all aspects. If, for example, a group of experts are chosen on the basis of their scientific credentials, it does not necessarily imply that they are able to ask the right questions when they, for example, have to visit the intended beneficiaries of the research. To counter such problems, some agencies outside the area of development cooperation[25] have designed set of guidelines for the implementation of evaluations. This can include both evaluation planning, choice of evaluators, content/scope/range of the evaluation, the methods and indicators, and the final dissemination and use of the results.

4.2.2 Operationalibility of Case Studies

The strengths and limitations of case studies with respect to operationalibility are similar in nature to those that were identified looking at the credibility of this type of study. On the positive side, case studies have, depending on the level of detail, the potential to indicate (but not test) causal relationships and effects. In addition, the studies can take the political context and issues into account.

On the negative side, case studies are to some extent also measurement without theory. The quality and relevance of the output relies almost solely on evaluator abilities. Furthermore, it can be difficult to generalize from a selected number of cases to identify the most optimal programme and/or policy changes.

4.3 *Bibliometric Indicators*

As mentioned previously, only two of the 23 evaluations have included an analysis of bibliometric indicators as part of the basis for the evaluation – both of them within the field of health assistance.

4.3.1 Credibility of Bibliometric Indicators

The strengths of bibliometric indicators are quite straightforward: In an area which lacks both data and measurable results, bibliometric indicators offer a quantitative and precise way of measuring scientific impact. Given the increased use of these indicators at universities and other centres of higher learning and research, they are moreover recognized by most researchers. In addition, bibliometric indicators offer the evaluators the possibility of benchmarking with other programmes and/or gen-

[25] This includes the EU Commission as described in Krull (1995) and the Federal German Government as described in Kuhlmann (1995).

eral research within the field as well as an opportunity to see how departments/research groups develop over time.

The limitations are, however, equally easy to identify. Bibliometric indicators can, of course, only address the scientific impact of research, and only the formalized part of this. In addition, capacity building and teaching will not show up in the bibliometric indicators (see chapter 2.1). The question of comparability across scientific fields and even across institutions within the same field is also contested. Some believe it is possible to construct field normalization factors that facilitate comparison, whilst others strongly object to any attempt to compare. Differences in publication traditions and channels moreover imply that some fields are not suited for bibliometric analyses. This includes the evaluation of research within humanities, social science, IT and communication. Finally and more fundamentally, researchers may choose to cite the works of their colleagues for at least three different reasons: to acknowledge the scientific quality of the research presented, to differentiate the qualities of their own research, or to forward objectives not related to the scientific dimensions of their research (e.g. their career, their relationship with colleagues, their political beliefs). Only the first of the three motifs would be in accordance with the standard interpretation of bibliometric indicators.

As a consequence of the shortcomings listed above, bibliometric indicators should not be allowed to stand-alone and any form of comparison should be treated with caution. Under these circumstances, the bibliometric indicators, however, provide valuable insights into the scientific impact of research support and not least how it develops over time. It is in this capacity that the method is used in the Danida evaluation of health research (Danida, 2007) and the third external review of the TDR (WHO, 1998).

4.3.2 Operationalibility of Bibliometric Indicators

The strengths and limitations of bibliometric indicators in terms of operationalibility are quite similar to those outlined above relating to the credibility of the method. It is, of course, a strong point that the method provides a very precise indication of the areas of strength. In addition, the possible inclusion of a dynamic perspective, which most often is lacking, is another strong point.

The limitations are shared with all other types of quantitative indicators: They cannot be used in isolation and should always be accompanied by a description of context and interpretations of observed differences and/or special circumstances. In addition, the indicators cannot be taken as indicative of causal relationships and problems of additionality in terms of input, output, and behaviour are unresolved using this method.

Overall, using bibliometric indicators is a useful way of meeting the demand for objective indicators to complement the dominant expert review process. In addition, to providing alternative and complementary information to overcome the reliance on existing information and prevailing wisdom, bibliometric indicators can provide a systematic and potentially dynamic perspective, which is sometimes missing from the other methods.

Given the widespread use of bibliometric indicators by both scientists and evaluators in other fields and the aforementioned frequent call for a more quantitative basis for evaluations of development research, the limited reliance on bibliometric indicators is perhaps a bit surprising. It is, however, not possible to provide anything but a purely speculative answer to this question, as none of the evaluations have mentioned the methodological considerations and/or analyses that underlie their choice of method.

5. Conclusion

Based on a survey of 23 evaluations of assistance to development research, this analysis has provided an overview of the state of affairs when it comes to evaluations of support to development research. Not surprisingly, all the evaluations in the survey rely on the expert review method, although a significant proportion complement the analysis of existing documentation and interviews with case studies. Only a minority relies on bibliometric indicators to assess the scientific impact of the assistance to development research.

The choice of evaluation methodology appears to be motivated by the flexibility and adaptability of both expert reviews and case studies – both in terms of data/informational requirements for the analysis and in terms of the ease at which policy and decision support can be covered. Many of the evaluations, however, implicitly or explicitly acknowledge the need for better data and monitoring to form a more solid basis for future evaluation – not least to accommodate the increased demand for impact assessments. The challenges identified for future evaluations can be categorized into the (relatively) easy challenges and the (very) difficult challenges.

The (relatively) easy challenges all relate to adopting a more structured approach and being more transparent about this and other aspects of evaluation methodology. More specifically, a more structured approach is called for in terms of:

- The selection criteria for the panel of experts for the expert review. What criteria were used in the selection of experts? How are different views and stakeholders represented? And not least: How was the independence of the participants established and can it be verified by outsiders?
- The selection of case studies. What criteria were used and what aspects are the different cases meant to cover? And how was the number of cases decided and are end-user perspectives represented?
- The balance and weight assigned to different inputs in the final evaluation, in particular how were interview responses evaluated and possibly cross checked?

The latter should preferably be backed by a systematic and well documented approach to both the selection of interviewees and the subsequent use of their answers.

Common to all these recommendations is that they can be implemented at a relatively low cost and will assist in making the final evaluation reports both more credible and more operational.

The (very) difficult challenges all relate to the unresolved and complex issues regarding the evaluation of research. Although the introduction of indicators will neither solve the problem nor bring evaluators any closer to uncovering the unresolved questions about the impact of research, the introduction of more systematic attempts to monitor the research project will almost surely improve the quality and relevance of future evaluations. This is, of course, a long-term project but does not need to incur substantial cost upon the development cooperation agencies as it could be part of the requirements to researchers that receive support that they attempt to document and monitor progress and impact.

Another issue is developing the evaluation methodology for research assistance. The introduction of a more systematic approach to critical selection issues and a higher degree of transparency about this and other methods will obviously not solve the problems. Here it has to be acknowledged that the methods for demonstrating the specific contribution and impact of a research programme are likely to remain insufficient to overcome the joint barriers of a complex interaction, long time lags, attribution, and deficient or missing necessary factors. The potential importance and impact that high quality research can have upon the livelihood and well-being of millions of poor people, however, make it imperative that we keep on trying.

6. **References**

Aghion, P and Howitt, P (1996) Research and Development in the Growth Process, *Journal of Economic Growth*, Vol. 1 (March), pp 49-73.

Arrow, K (1962) Economic Welfare and the Allocation of Resources for Invention in *National Bureau of Economic Research, The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton, N.J.: Princeton University Press.

Boyack, KW and Börner, K (2003) Indicator-Assisted Evaluation and Funding of Research: Visualizing the Influence of Grants on the Number and Citation Counts of Research Papers, *Journal of the American Society for Information Science and Technology*, 54(4), pp. 447-461.

Coe, DT; Helpman, E; and Hoffmaister AW (1997) North-South R&D Spillovers, The Economic Journal, Vol. 107(January), pp. 134-149.

Committee on Science, Engineering, and Public Policy (2001) *Implementing the Government Performance and Results Act for Research: A Status Report*, National Academy of Sciences, National Academy Press, Washington, D.C., USA

DFID (2005c) *Rates of Return to Research: Literature Review and Critique*, 29 p, Enterplan, Department for International Development, London, England.

European Commission (2002) *RTD Evaluation Toolbox Assessing the Socio-Economic Impact of RTD Policies*, 295 p. European Commission, Joint Research Centre and Institute for Prospective Technological Studies, Sevilla, Spain.

European Research Council (2003) *The European Research Council: A Cornerstone in the European Research Area*, Ministry of Science, Technology and Innovation, Copenhagen, Denmark.

Feller, I. (2001) *The Academic Policy Analyst: The Who, What and How of Evaluating Science and Technology Programs* in: Proceedings from the 2000 Workshop on Learning from Science and Technology Policy Evaluation, by Shapira, P and Kuhlmann, S (eds.), p. 1- 17, School of Public Policy, Georgia Institute of Technology, Atlanta,

USA and the Fraunhofer Institute for Systems and Innovations Research, Karlsruhe, Germany.

Georghiou, L (1995) Assessing the Framework Programmes: A Meta-Evaluation, *Evaluation*, Vol. 1, No. 2, pp. 171-188.

Georghiou, L (2001) *Evaluation of Research and Innovation Policy in Europe – New Policies, new Frameworks?* in: Proceedings from the 2000 Workshop on Learning from Science and Technology Policy Evaluation, by Shapira, P and Kuhlmann, S (eds.), p 42 - 53, School of Public Policy, Georgia Institute of Technology, Atlanta, USA and the Fraunhofer Institute for Systems and Innovations Research, Karlsruhe, Germany.

Gauthier, E (1998) Bibliometric Analysis of Scientific and Technological Research: A User's Guide to the Methodology, Statistics Canada, Ottawa, Canada.

Griliches, Z (1998) R&D and Productivity: The Econometric Evidence, University of Chicago Press, Chicago, USA.

Godin, B (1996) The State of Science and Technology Indicators in the OECD Countries, Statistics Canada, Ottawa, Canada.

Hoebink, P (1996) *The Comparative Effectiveness and the Evaluation Efforts of EU Donors*, 9 p., The Dutch National Advisory Council for Development Cooperation, The Hague, The Netherlands.

IPTS (2002) *Research and Technology Development (RTD) toolbox: assessing the socio-economic impact of RTD polices*, IPTS Technical Paper, 269 p, Seville, Spain.

Kuhlmann, S (1995) German Government Department's experience of RT & D programme evaluation and methodology, *Scientometrics*, Vol. 34(3), pp. 461-471.

Krull, W (1995) *Guidelines for the Evaluation of European Community R&D Programmes*, European Commission, Brussels, Belgium.

Leimu, R and Koricheva, J (2005) What Determines the Citation Frequency of Ecological Papers? *Trends in Ecology and Evolution*, 20(1), pp. 28-32.

Mariegaard, M (2006) Forskere til Danida: Støt langsigtet forskning (Eng. translation Researchers to Danida: Support long-term research), *Udvikling*, Vol. 5, p. 12.

McGrath S and King K (2004) Knowledge-Based Aid: A Four Agency Comparative Study, *International Journal of Educational Development*, vol. 24, pp. 167-181.

Michaud, CM and Reich, MR (2005) *Impact Evaluation for TDR: Feasibility and Approaches*, Background paper for the 4th external review of the TDR, 15 p. Geneva, Switzerland.

Okubu, Y (1997) *Bibliometric indicators and Analysis of Research Systems: Methods and Examples*, OECD Science, Technology and Industry Working Papers 1997/1, OECD Publishing, Paris, France.

Rip, A (2001) *Societal Challenges for R&D Evaluation*, in: Proceedings from the 2000 Workshop on Learning from Science and Technology Policy Evaluation, by Shapira, P and Kuhlmann, S (eds.), p 18- 41, School of Public Policy, Georgia Institute of Technology, Atlanta, USA and the Fraunhofer Institute for Systems and Innovations Research, Karlsruhe, Germany.

Romer, PM (1990) Endogenous Technological Change, J*ournal of Political Economy*, Vol. 98(5), pp. S71-S102.

Scott, A. (2003) *A Review of the Links between Research and Policy*, Draft Working Paper, University of Sussex, Sussex, England.

Stiglitz, J (1989) Markets, Market Failures, and Development, *American Economic Review*, Vol. 79(2), pp.197-203.

Stone, D (2002) Using Knowledge: The Dilemmas of Bridging Research and Policy, Compare, Vol. 32(3), pp. 285-296.

Stone, D (2004) Transfer Agents and Global Networks in the Transnationalization of Policy, *Journal of European Public Policy*, Vol 11(3), pp. 545-566.

Stone, D, Maxwell, S and Keating M (2001) *Bridging Research and Policy*, Paper prepared for the International Workshop at Warwick University 16-17 July, 2001.

6.1 *Evaluations of Development Research*

Australian Department of Education, Science and Training (2006) *Review of the National Research Flagships – an initiative of CSIRO*, 25p, Commonwealth Scientific and Industrial Research Organization, Canberra, Australia.

Danida (2007) Review of Danida-supported health research in developing *countries* (draft version), 104 p, Danish Ministry of Foreign Affairs, Copenhagen, Denmark

Danida (2006) *Bridging Research and Development Assistance: a Review of Danish research Networks*, 28 p, Christian Michelsen Institute, Oslo, Norway.

Danida (2001) Partnerships at the leading edge: A Danish vision for knowledge, research and development (The Hernes report), Danish Ministry of Foreign Affairs, Copenhagen, Denmark.

DFID (1998) *Natural Resources Research Evaluation Synthesis*, DFID Evaluation report EV604, 46 p, Department for International Development, London, England.

DFID (2002) *Research for poverty reduction: DFID research policy paper (The Surr Report)*, DFID Research Policy Paper, 91 p, Department for International Development, London, England.

DFID (2005a) *Evaluation of DFID's Renewable Natural Resources Research Strategy*, LTS International, Department for International Development, 517 p, Edinburgh, UK.

DFID (2005b) *Evaluation of DFID's Engineering Knowledge and Research (EngKaR) Programme*, Technopolis and Overseas Development Institute, 83 p, Department for International Development, London, England.

European Commission (2004) *Five-Year Assessment of the European Union Research Framework Programmes*, 1999-2003, 36 p, Directorate General for Research, Brussels, Belgium.

Norad (2000) Evaluation of the Norwegian Council of Universities' Programme for Development Research and Education, 70 p, Norad Evaluation Report 5/2000, Norwegian Agency for Development Cooperation, Oslo, Norway.

Norad (2005a) *Making support to higher education and research more effective*, 21 p, Norad Discussion Report 1/05, Norwegian Agency for Development Cooperation, Oslo.

Norad (2005b) *Evaluation of the Norad Fellowship Programme*, 163 p, Norad Evaluation Report 1/2005, Norwegian Agency for Development Cooperation, Oslo, Norway.

Nuffic (2001) *Comparative study of the impacts of donor-initiated programmes on research capacity in the south*, 160 p, The Ministry of Foreign Affairs, The Hague, the Netherlands.

Nuffic (2007) *Evaluation of the Netherlands' Research Policy*, 1992-2005, IOB Evaluation No. 304, English Summary, 98 p. The Hague, the Netherlands.

Sarna, N. (2004) *The Global Development Network. Addressing the Challenges of Globalization: An Independent Evaluation of the World Bank's Approach to Global Programs*, The World Bank Operations Evaluation Department, Washington D.C., USA

Sida (2006a) *Evaluation of Sida Information and Communications Technologies Support to Universities*, Sida Evaluation 06/13, 54 p, Swedish International Development Cooperation Agency, Stockholm, Sweden.

Sida (2006b) *Sida/SAREC Bilateral Research Cooperation: Lessons Learned*, Sida Evaluation 06/17, 56 p, Swedish International Development Cooperation Agency, Stockholm, Sweden.

Sida (2006c) *Organisationsstudie av SAREC,* 06/22, 86 p, Swedish International Development Cooperation Agency, Stockholm, Sweden.

Sida (2006d) *Sidas U-landsforskningsråd*, Sida Evaluation 06/24, 68 p, Swedish International Development Cooperation Agency, Stockholm, Sweden.

Sida (2006e) *SAREC:s stöd till svensk u-landsforskning*, Sida Evaluation 06/27, 56 p (in Swedish), Swedish International Development Cooperation Agency, Stockholm, Sweden.

Sida (2006f) *SAREC Support to International and Regional Thematic Research Programs*, Sida Evaluation 06/40, 56 p, Swedish International Development Cooperation Agency, Stockholm, Sweden.

Sida (2006g) Översyn av verksamhetsgrenen forskning, Sida Evaluation 06/41, 42 p, Swedish International Development Cooperation Agency, Stockholm, Sweden.

The Research Council of Norway (1998) *Building Research Capacity in Africa: a Review of NORAD's Assistance to Regional Research Organizations*, 155 p, Oslo, Norway.

The Research Council of Norway (1998) Norwegian Research Support to Developing Countries: The Cases of Uganda and Zimbabwe, 118 p, Oslo, Norway

The Royal Society of New Zealand (2004) *the Impact of Marsden Funded Research: a Bibliometric Assessment of Marsden Funded Publications*, 1997-2001, 45 p, Wellington, New Zealand.

WHO (1998) *Third External Review of the TDR: Final Report*, the Special Programme for Research and Training in Tropical Diseases (TDR), the World Health Organisation, Geneva, Switzerland.

WHO (2006) *Fourth External Review of the TDR: Towards Evolution and Growth*, The Special Programme for Research and Training in Tropical Diseases (TDR), The World Health Organisation, Geneva, Switzerland.

The World Bank (1999) K*nowledge for Development: World Development Report 1999/1998*, World Bank, Washington DC, USA.

The World Bank (2004a) *Evaluating the World Bank's Approach to Global Programs: Addressing the Challenges of Globalization,* World Bank, Washington DC, USA.

The World Bank (2004b) *The CGIAR at 31: An independent meta-evaluation of the consultative group on international agricultural research*, World Bank, Washington DC, USA.

Appendix A: Terms of Reference

Background

As an input to a planned evaluation of Danida's support for development research, the Evaluation Department has decided to elaborate a synthesis of a number of available evaluations of support for development research. The evaluations, which will be covered by the synthesis, will be identified by the consultant.

Objective

The objective of the synthesis can be split into two. First, an overview of existing evaluations of support for development research will be provided. Second, a critical analysis and evaluation of the applied evaluation methods will be conducted. The first part of the synthesis is thus instrumental to the second, which will be the main part of the analysis.

Output

The final report will be a short synthesis report of not more than 20 pages. The final synthesis report will be made available on the Internet by the Evaluation Department.

In addition the intention is to present the synthesis report on a workshop organised by the Evaluation Department.

Scope of work

The synthesis will be based on a desk review of a selected list of references. The synthesis report will not contain assessments of individual research projects and programmes and only secondary sources will be used.

The synthesis report will cover two main areas:

1. An outline of the already existing evaluations of assistance to development research. This includes a summary of :

The key findings of other evaluations. What are, for example, the conclusions regarding the relationship between development research and economic development?

The scope of other evaluations. Are they, for example, restricted to certain sectors, what is the timeframe considered, what is the geographical focus, and does the evaluations consider spill-over effects outside the research environment?

The methods applied in other evaluations. More specifically, do they, for example, apply a rate of return approach, what type of output indicators is considered, and how is the balance between quantitative and qualitative assessments?

The overall scope and distribution of development assistance to research. This will be instrumental and essential for the efforts to qualify the assessment of other evaluations. A set of tables describing the size and distribution of this type of assistance will be provided as an appendix to the report.

2. A methodological assessment of the strength and weaknesses of different approaches to evaluating assistance to development research. This includes an assessment of:

   The scope of the evaluations. What is the implication of choosing different sectoral, temporal and geographical cut-offs?

   The methods applied in the evaluations. How can an evaluation cover different types of research? What are the issues regarding the use of rate of return and other quantifiable methods? What type of data is typically available? For all methods considered the robustness and the general applicability of the results will be considered.

Please note, that the analysis will not have a separate assessment of the effectiveness of assistance to research vis-à-vis other forms of development assistance. In addition, the discussion of which methods will be best suited for the Danish context will not be initiated in the synthesis report.

Evaluation Department 23.3.2007

Appendix B: Identification of Evaluations for the Survey

A desk-based review of the relevant evaluations was conducted, applying the following three search criteria:

1. **Scope of evaluations**. The programmes that were subject to evaluation should be as broad-based and comprehensive as possible. The rationale behind this criterion is to limit the inclusion of evaluations of single-field programmes as it can be difficult to generalize and learn from these.

2. **Time of publication**. The included evaluations should be as recent as possible. An expiry period of ten years was imposed, resulting in the exclusion of all evaluations conducted before 1997.

3. **Topic of evaluation.** The majority of the included evaluations should be of support to research and development related to understanding and/or promoting economic development in poor countries.

The application of these selection criteria resulted in the elimination of a number of evaluations from consideration. More specifically:

Ad 1. A number of sub-evaluations of specific sectors were not considered in details. This includes, for example, the special surveys made for the Surr report (DFID 2002) and the Hernes report (Danida 2001). In addition, a donor like USAID who only conducts sector/programme specific evaluations of research support is not represented in the survey. Finally, a number of evaluations of research support and cooperation with individual countries were also excluded from the survey.

Ad 2. Only a limited number of evaluations were eliminated from the survey due to this constraint. This includes, for example, the first two evaluations of the special programme for research and training in tropical diseases (TDR) and a number of early, country-specific evaluations of research support conducted by Sida, SAREC and Norad.

Ad 3. All evaluations related to support to higher education and general capacity building at universities in developing countries are deselected because of this criteria.

The relevant evaluations were identified through a net-based search and direct consultations with a number of organisations. In this context, I would like to thank the following people for their kind assistance: Arne Tostensen of the Christian Michelsen Institute in Norway, Atsuko Shintani of UNESCO, Marit Kuyper of the Dutch Agency for International Development, Pierre-Joseph Kingbo of The World Bank, Stephan Dalgren of SAREC and Jim Harold of USAID.

Appendix C: Overview of Assistance to Development Research

Please note that the sector classification does not refer to the type of goods or services provided by the donor. Sector specific education or research activities (e.g. agricultural research) or construction of research infrastructure (e.g. agricultural research lab) is reported under the sector to which they are directed, not under, for example, consultancy or construction.

CRS code	Description	Additional notes on coverage
11182	Educational research	Research and studies on education effectiveness, relevance and quality; systematic evaluation and monitoring.
12182	Medical research	General medical research (excluding basic health research).
23082	Energy research	Including general inventories, surveys.
31182	Agricultural research	Plant breeding, physiology, genetic resources, ecology, taxonomy, disease control, agricultural bio-technology; including livestock research (animal health, breeding and genetics, nutrition, physiology).
31282	Forestry research	Including artificial regeneration, genetic improvement, production methods, fertilizer, harvesting.
31382	Fishery research	Pilot fish culture; marine/freshwater biological research.
32182	Technological research and development	Including industrial standards; quality management; metrology; testing; accreditation; certification.
41082	Environmental research	Including establishment of databases, inventories/accounts of physical and natural resources; environmental profiles and impact studies if not sector specific.

43082	Research/scientific institutions	When sector cannot be identified.

Please note, that each activity can be assigned only one purpose code (this is to avoid double-counting when summing up activities in different ways). For activities cutting across several sectors, either a multi-sector code or the code corresponding to the largest component of the activity is used. The data obtained using the method of a single purpose code may thus differ slightly from those provided by donors' internal systems that allow an activity to be assigned to more than one sector. However, at present it is the only practical method of standardising reporting on a basis that permits valid donor comparisons. It is not likely to bias analyses of trends and orders of magnitude.

Total Development Assistance

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Total (Million USD)	42.410	51.760	46.535	41.333	43.031	47.812	49.912	43.949	49.560	55.789	56.436	55.364	64.769	90.569
Bilateral (Million USD)	32.355	41.004	34.724	31.115	30.945	36.831	37.476	31.510	33.718	41.004	41.478	38.897	46.415	67.768
Australia	369	540	323	659	726	270	1.478	722	566	806	1.027	710	656	1.101
Austria	122	482	195	176	166	162	529	339	515	625	359	401	458	282
Belgium	93	222	129	71	356	447	582	457	522	466	503	557	1.083	1.547
Canada	1.713	920	1.050	960	1.063	1.360	1.256	951	1.084	952	1.216	986	1.595	1.576
Denmark	410	419	573	638	503	889	1.442	808	406	641	940	931	872	677
Finland	510	678	365	208	149	207	198	175	215	214	199	280	298	386
France	2.504	3.103	2.694	2.444	2.524	2.453	2.773	2.254	3.344	3.026	2.529	2.892	4.164	6.532
Germany	3.938	3.321	4.284	2.644	3.148	3.999	4.495	3.167	2.403	3.972	3.128	3.370	4.584	5.642
Ireland	..	..	..	..	..	..	..	..	..	..	138	201	267	325
Japan	8.486	11.150	9.387	12.463	10.856	14.090	13.083	11.308	10.769	10.410	10.171	8.655	6.587	14.442
Netherlands	1.086	1.008	1.512	1.727	1.321	2.132	2.189	1.983	2.028	1.851	2.551	2.392	4.471	2.439
Norway	477	555	409	536	707	899	857	679	700	1.155	795	1.085	1.102	1.434
Portugal	..	15	204	178	229	90	81	125	51	246	322	197	184	188
Spain	525	1.037	885	435	265	323	389	739	1.056	1.026	950	1.230	1.165	1.488
Sweden	2.378	1.719	2.130	1.007	1.163	1.121	1.006	1.104	1.059	978	989	938	1.088	2.025
Switzerland	532	693	565	401	934	758	562	509	500	867	677	736	785	918
United King-dom	1.063	1.147	2.826	811	936	1.278	1.907	1.620	3.175	2.927	4.226	2.739	3.574	3.956
United States	6.552	11.749	4.830	4.656	4.657	5.364	3.958	4.053	4.649	10.228	10.030	9.879	11.950	20.934
Other	1.598	2.247	2.364	1.102	1.244	989	691	519	678	615	729	719	1.532	1.875
Multilateral (Million USD)	10.055	10.756	11.812	10.218	12.085	10.981	12.435	12.439	15.842	14.785	14.958	16.467	18.354	22.801

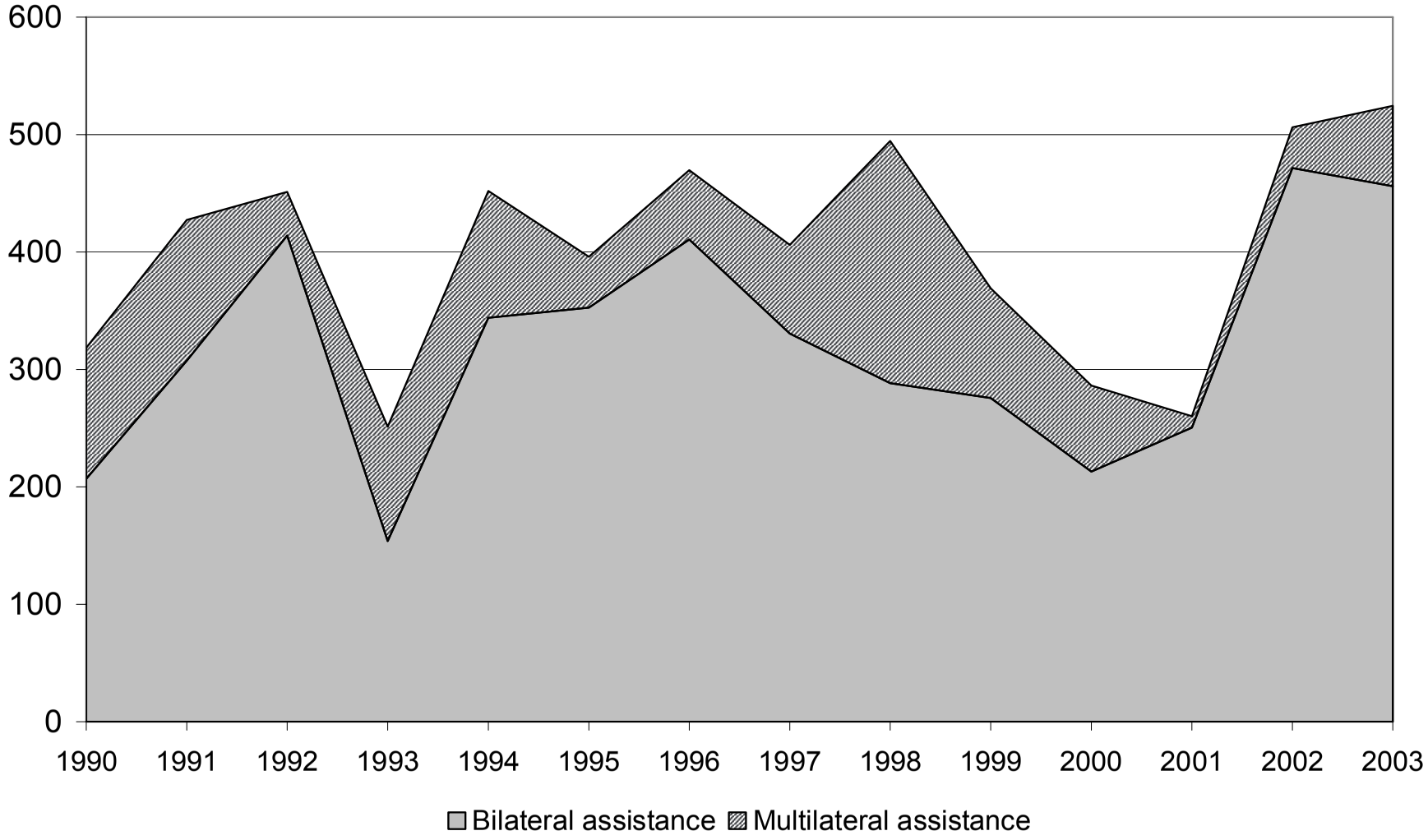
Source: OECD CRS database (accessed 12-Sep-07)

Total Development Assistance to Research

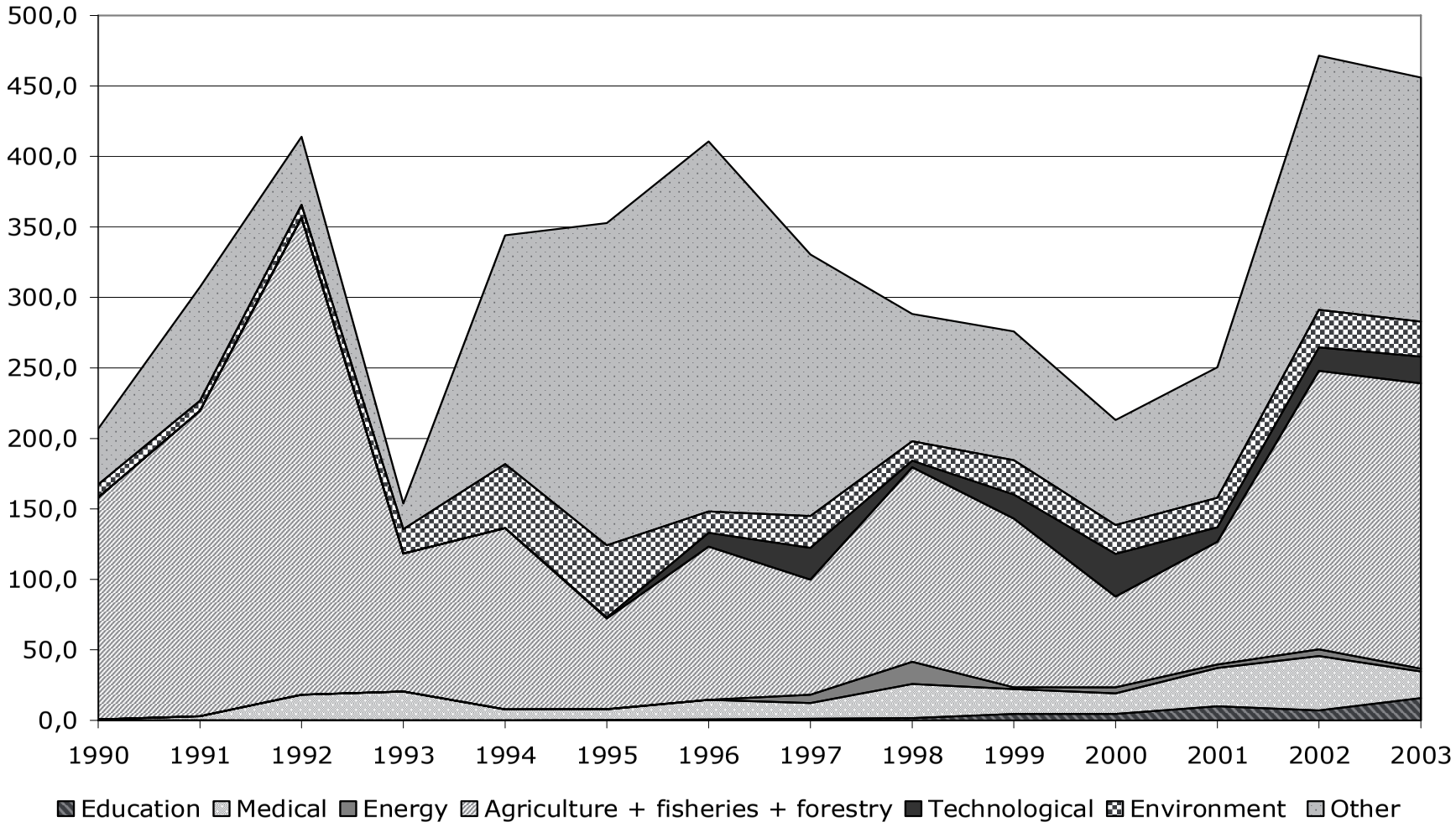
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Total (Million USD)	318	427	451	251	452	396	470	406	495	369	286	260	506	524
Bilateral (Million USD)	207	308	414	154	344	353	411	331	288	276	213	251	472	456
Australia	18	8	1	3	5	1	14	2	12	11	15	12	16	8
Austria	0	3	0	0	0	0	0	0	2	2	0	0	1	1
Belgium	0	0	0	0	9	11	15	9	10	6	13	6	6	12
Canada	30	28	6	11	3	4	3	6	5	3	6	11	11	48
Denmark	4	7	26	8	5	6	2	1	8	1	3	6	23	19
Finland	4	13	3	2	6	2	3	3	2	6	5	4	6	8
France	10	15	9	9	140	175	178	140	25	20	16	20	22	29
Germany	0	11	0	0	3	0	0	17	3	32	18	18	21	26
Greece	0	0	0	0	0	0	0	0	0	0	0	0	2	0
Ireland	0	0	0	0	0	0	0	0	0	0	2	2	2	5
Italy	19	11	16	9	1	5	4	5	4	0	0	0	1	2
Japan	15	77	17	20	47	7	0	18	18	15	20	8	6	19
Luxembourg	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Netherlands	9	10	50	40	74	48	36	60	42	43	20	39	52	19
New Zealand	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Norway	1	11	3	7	3	29	60	7	7	23	12	27	27	38
Portugal	0	0	0	0	0	0	0	0	0	2	2	3	5	5
Spain	0	0	0	0	0	0	1	9	20	7	8	7	10	45
Sweden	1	5	8	1	0	3	4	10	29	28	30	30	83	98
Switzerland	9	9	9	2	13	12	2	6	11	0	0	0	10	0
United Kingdom	24	24	191	9	36	14	40	10	43	70	43	51	96	11
United States	64	75	77	33	0	37	48	28	48	6	2	5	71	63
Multilateral (Million USD)	112	120	37	97	108	43	59	76	206	93	73	10	35	68

Source: OECD CRS database (accessed 12-Sep-07)

Total Assistance to Research: Bilateral and Multilateral



Bilateral Assistance to Research: By Type of Research



Published by:
Ministry of Foreign Affairs of Denmark
Evaluation Department
Asiatisk plads 2
1448 Copenhagen K
Denmark
E-mail: eval@um.dk

The publication can be downloaded from:
www.evaluation.dk

ISBN: 978-87-7667-792-3 (print version)
ISBN: 978-87-7667-793-0 (internet version)