



THE MEGA 2004 EVALUATION

**(Meta-Evaluation of Goal Achievement
in CARE Projects)**

**A Review of Findings and Methodological Lessons from
CARE Evaluation Reports, 2003-2004**

SUMMARY REPORT

Craig Russon, Ph.D.
Programme Planning & Evaluation, Inc.

April 2005

Table of Contents

Acronyms Used in this Report	2
Executive Summary	3
The MEGA 2004 Evaluation	6
Present Terms of Reference	6
Method	6
Results	6
General Information	7
Goals and Objectives	7
SII-related Themes	9
Evaluation Approach, Design, and Methods	9
Methodological Rigor	11
Terms of Reference for conducting program/project evaluations	12
CARE DME Standards	13
Accessibility of Evaluation Reports	17
Conclusion	18
References	19

Acronyms Used in this Report

AKAP	Awareness, Knowledge, Attitude, Practice (survey)
ANR	Agriculture and Natural Resources (sector)
BGE	Basic and Girls; Education (sector)
CCS	Civil Society Strengthening (sector)
C-PIN	CARE International Program Information Network
CO	CARE Country Office
DME	Design, Monitoring, and Evaluation
EeL	Evaluation Electronic Library
ER	Emergency Response (sector)
GOV	Governance, Municipal Development (sector)
HH	Household
HIV	HIV/AIDS (sector)
HLS	Household Livelihood Security
IMLT	Impact Measurement and Learning Team
INF	Infrastructure (sector)
MEGA	Meta-Evaluation of Goal Achievement
MDG	Millennium Development Goals
NUT	Nutritional Health (sector)
PNGO	Partner Non-Governmental Organization
RBA	Rights Based Approach
RH	Reproductive Health (sector)
SEAD	Small Economic Activity Development (sector)
SEG	Sustainable Economic Growth
SII	Strategic Impact Inquiry
SWOT	Strengths, Weaknesses, Opportunities, and Threats (analysis)
TOR	Terms of Reference
USAID	United States Agency for International Development
WATSAN	Water, Sanitation, and Environmental Health (sector)

Executive Summary

Are CARE projects having impact on the lives of their intended beneficiaries? One way to answer that question is to examine available reports of the evaluations of CARE projects. Such a synthesis and metaevaluation is conducted every two years. This is the report of the third such study, covering project evaluations conducted during calendar years 2003 and 2004.

MEGA 2004 utilized the same methodology that was employed in 2000 and 2002, though it was conducted by a different external evaluator than were the first two MEGAs. While reviewing each document, the evaluator recorded general information and key characteristics on goals and objectives, themes related to the Strategic Impact Inquiry on women's empowerment, evaluation methodology, how well the project met the CARE International DME standards, and accessibility of the evaluation report.

The results of the review found that, according to those conducting their evaluations, 74 percent of these CARE projects fully or partially achieved their final goals. CARE is indeed having an influence on the lives of people in poor communities in areas as diverse as disaster relief, refugee resettlement, food programs, water and sanitation, economic development, health, education, peace building, natural resources management, and the list goes on.

Only 11 reports specifically incorporated Household Livelihood Security indicators and/or Millennium Development Goal indicators. Unfortunately, the projects and programs that did use these indicator sets did not use them in a very effective manner. For example, some projects proposed HLS and/or MDG indicators, but collected no data on them.

A few lessons emerged regarding the goals at which projects and programs aim and how well they achieve them.

- Goals and timelines that are overly ambitious are difficult to attain
- Expanding a successful strategy too quickly can have a negative influence on goal attainment
- Sometimes it is necessary to manage expectations about goals to avoid disappointment among beneficiaries
- Tangible goals (e.g., building infrastructure) are easier to attain and measure than are social goals
- Emergency assistance provides short-term relief but can leave communities without capacity to sustain facilities

Evaluators who wrote the reports often determined success by the completion of project activities associated with the objectives of the project or program. Impacts may or may not have been measured.

The recent emphasis on logical frameworks and HLS/MDG indicators suggests that CARE may be making a shift to a global results-based approach. If so, this may enable CARE to better aggregate the impact of its projects and programs in the future

According to what could be ascertained from the evaluation reports, almost all of the projects and programs were consistent with at least one of CARE International's Programming Principles. If the proposed CARE Evaluation Policy is adopted, in the future project and program evaluation TORs should more specifically include examining how well the project/program was consistent with these Principles and the DME Standards.

The MEGA'04 assessed how well the evaluated projects met the CARE DME Standards. The following paragraphs summarize the findings relative to each of those Standards:

About 32 percent of the reports established a link between the project and the Country Office and/or long-term goal. There were primarily two ways in which this appeared to happen. There were several pilot projects whose explicit purpose was to help inform future Country Office programming. Also, many projects were follow-ons to previous work that had been accomplished. Both of these approaches demonstrate how CARE is using its programming to promote organizational learning in its Country Offices.

Many CARE projects attempted to engage main stakeholders in participatory design, implementation, and M&E processes, at least in non-emergency projects.

About 42 percent of the reports described a holistic analysis of the needs and rights of target populations on which the project was based. There were almost as many ways to conduct the analysis as there were projects.

About 31 percent of the reports provided evidence of a logical framework. However, based on his review, it is the consultant's opinion that many of those projects did not use logical framework to guide the evaluation as effectively as they might have.

Only about 19 percent of the reports reiterated a final goal that met criteria for well-defined goals.

About 40 percent of the reports contained evidence of technical, social, **and** environmental appropriateness. However, it would take sectoral experts to determine the answers to these questions with more authority, considering accepted current best practice guidelines.

Thirty-one percent of the reviewed reports provided information on the appropriateness of project costs, making the justification that they were worth the value in terms of outputs.

Though not necessarily based on the logical framework, about 46 percent of the evaluation reports included documentation of the project's M&E plan and/or system.

About 28 percent of the reports provided evidence that the project budget included adequate amounts for implementing their M&E plan.

About 51 percent of these evaluation reports mentioned that some form of baseline study had been conducted. Several projects identified flaws in the baselines. Some baselines were conducted well after project activities had begun. In many cases, the baseline was not comparable with the evaluation in terms of indicators measured and methodologies used.

Seventy-five percent of the evaluations used posttest only evaluation designs – i.e. no comparison with a baseline study, nor with a comparison group that was not reached by the project's interventions. It appears that even in those projects where baseline studies had been conducted, in many cases they were not utilized by the final evaluators to conduct evaluations with designs that would have been more rigorous (e.g., pre-posttest or a before-and-after comparison). Establishing a baseline and then not using it in subsequent evaluation efforts is a waste of resources.

As mentioned above, although about 47 percent of these evaluation reports included indicators, few used globally comparable HLS and/or MDG indicators. There does not appear to be much consensus among projects on this issue. Most of these indicator sets are relevant only to their particular project.

About 29 percent of the reports noted a more or less equal balance of quantitative and qualitative evaluation methodologies.

CARE projects have demonstrated a strong desire to contribute to the creation of a learning organization. About 54 percent of the evaluation reports included lessons learned of some form or other.

The evaluation challenges confronting CARE seem to be enormous. Evaluation is being planned and implemented in different ways in different countries. The result is that the quality of project and program evaluations is uneven. However, the future is bright. CARE's commitment to the Programming Principles, the DME Standards and the new Evaluation Policy positions the organization to take the quality of its evaluation work to the next level. Strategies must be developed to promote the DME Standards in the Country Offices and to facilitate their application. These steps may be the key to transforming CARE into a truly results-based, learning organization.

The MEGA 2004 Evaluation

The first global synthesis of lessons learned from CARE project evaluation reports was conducted in late 2000. Entitled "The MEGA Evaluation," the study examined 104 CARE project and program evaluations conducted during the period of 1994 through 2000. That report was well received by senior-level stakeholders in CARE. The second MEGA examined 65 evaluations conducted during the period of 2001 through 2002. It was completed in February 2003 and distributed widely in CARE (and beyond).

Present Terms of Reference

The present TOR called for a review of evaluation reports that can be found on the CARE Evaluation Electronic Library (EeL) and that were conducted since the second MEGA study was completed (i.e. during CY 2003-2004). Using the same criteria and methodology, as well as the same high professional standards, a short final report (this one) summarizes the results CARE projects achieved during the past two years as reported by available project and program evaluations.¹

Method

In close cooperation with the CARE Impact Measurement and Learning Team (IMLT), the consultant prepared a data collection form. The form was used to review 74 of CARE's 2003-2004 mid-term and final evaluation reports. Only reports written in English and Spanish were reviewed. Those written in French were not reviewed, due to the consultant's inability to read French, and lack of time and resources for them to be translated into English. The evaluation reports came from three sources: the CARE EeL, two IMLT CDs, and reports that were emailed to the consultant. Quantitative data was analyzed using SPSS and qualitative data was analyzed using MS Word.

Results

MEGA 2004 utilized the same methodology that was employed in 2000 and 2002, though it was conducted by a different external evaluator than were the first two MEGAs. While reviewing each document, the evaluator recorded general information and key characteristics on goals and objectives, SII-related themes, evaluation methodology, how well the project met the CARE International DME standards, and accessibility of the evaluation report.

¹ A more detailed, 75-page full technical report is also available to interested readers.

General Information

Table 1 illustrates the distribution of the projects and programs whose reports were reviewed. Reports were received from more countries than in the past (26 in 2004 vs. 19 in 2002). Also, the regional distribution of reports has changed. The number of reports coming from Asia is still high, but now almost equal numbers come from East/Central and South/West Africa. The number of evaluation reports received from Latin America declined 43 percent from 2002.

Table 1. MEGA 2004 Evaluation Reports Reviewed — CARE Regions and Countries

Region	Country	Number
Asia		18
	Afghanistan	3
	Bangladesh	4
	Indonesia	2
	Cambodia	3
	Sri Lanka	4
	Tajikistan	1
	Vietnam	1
East/Central Africa		15
	Eritrea	5
	Ethiopia	1
	Tanzania	9
South/West Africa		22
	Democratic Republic of Congo	2
	Ghana	7
	Mali	2
	Malawi	2
	Madagascar	1
	Mozambique	3
	Togo	1
	South Africa	1
	Zimbabwe	3
Middle East/Europe		11
	Egypt	3
	Kosovo	3
	Macedonia	3
	West Bank-Gaza	2
Latin America		8
	Bolivia	7
	El Salvador	1
Total		74

Goals and Objectives

According to the evaluators who wrote the reviewed reports, 74 percent of CARE's projects and programs fully or partially achieved their final goals. Review of project and program achievement summaries leads to the inescapable conclusion that CARE is indeed having a significant influence on the lives of people in poor communities in areas

as diverse as disaster relief, refugee resettlement, food programs, water and sanitation, economic development, health, education, peace building, and natural resource management; and the list goes on.

Table 2 shows the number of projects and programs that fully or partially achieved their final goals, disaggregated by region and sector. Because of the missing data and the low numbers in each cell of the matrix, no valid conclusions can be reached regarding the relative effectiveness of regions or sectors.

Table 2. Number of Projects Achieving Full or Partial Success in Meeting Final Goals -- by Region and Sector

Sector	Region					Total that met goals	Total n
	Asia	E/C Africa	S/W Africa	ME/Europe	Latin Am.		
ANR	1	3	2	2	2	10	14
BGE	-	1	2	1	-	4	5
CSS	2	-	-	1	-	3	3
ER	0	-	-	1	-	1	3
GOV	-	1	-	-	-	1	1
HIV	-	1	1	-	-	2	2
INF	2	-	-	-	-	2	2
NUT	-	1	1	-	-	2	2
RH	-	0	2	-	-	2	3
SEAD	1	1	3	-	-	5	6
WATSAN	1	-	1	-	1	3	3
Multiple	4	1	2	2	2	11	17
Total that met goals	11	9	14	7	5	46	61 ²
Total n	18	15	22	11	8		74

A few lessons emerged regarding the goals at which projects and programs aim and how well they achieve them.

- Goals and timelines that are overly ambitious are difficult to attain.
- Lack of shared understanding of goals makes them difficult to attain.
- Tangible goals (e.g., building infrastructure) are easier to attain and measure than are social goals.
- Government institutionalization helps promote long-term goals.
- Expanding a successful strategy too quickly can have a negative influence on goal attainment.
- Sometimes it is necessary to manage expectations about goals to avoid disappointment among beneficiaries.
- Emergency assistance provides short-term relief but leaves communities without capacity to sustain facilities.
- Conflicting objectives have a negative influence on final goal attainment.

If there is any bad news, it is that it is difficult to aggregate the impact of CARE's projects and programs. Only 11 reports incorporated Household Livelihood Security

² Project sector identification not clear on all evaluation reports.

Indicators and/or Millennium Development Goal Indicators. Unfortunately, the projects and programs that did use these indicator sets did not use them in a very effective manner. For example, some projects proposed HLS and/or MDG indicators, but collected no data on them.

SII-related Themes

In order to gain additional insights on CARE’s goals and objectives, the consultant looked for gender and empowerment objectives and analysis. Twenty-one of the projects and programs whose reports were reviewed had gender objectives. Forty (54%) of the reports contained evidence of some kind of gender analysis.³

Twenty-one of the projects and programs whose reports were reviewed also had empowerment objectives. Often the objectives were expressed in capacity building language. Eighteen of the reports contained descriptions of how well the project promoted empowerment (see Table 3).

Table 3. Gender and Empowerment Objectives and Analysis

	Frequency	Percent
Gender Objective	21	28.4
Gender Analysis	40	54.1
Empower Objective	21	28.4
Empower Analysis	18	24.3

Note: n=74

Evaluation Approach, Design, and Methods

No systematic information was collected on evaluation approaches. However, to this outside observer, it would appear from the heavy emphasis on goals and objectives that CARE has used a goal-based approach to evaluation in the past. Evaluators who wrote the reports often determined success by the completion of project activities associated with the objectives of the project or program. Impacts may or may not have been measured.

The recent emphasis on logical frameworks and HLS/MDG indicators suggests that CARE may be making a shift to a global results-based approach. If so, this may enable CARE to better aggregate the impact of its projects and programs in the future.

Table 4 shows the designs that were used to evaluate the projects and programs whose reports were reviewed. The design that was used most often was the posttest-only analysis of the project group (i.e. without baseline or comparison group). According to

³ It should be noted that analysis of gender-related issues was not necessarily included in the ToRs of individual project evaluations. A part of the current SII is to retrospectively identify which project evaluations might have addressed gender empowerment in the recent past.

Campbell & Stanley (1963), this would be among the least rigorous designs. However, Stake (1995) would probably disagree with that assessment.

Table 4. Evaluation Designs

	Frequency	Valid Percent	Cumulative Percent
Longitudinal	1	1.5	1.5
Pre- Posttest Project and Control	1	1.5	3.0
Truncated Longitudinal	1	1.5	4.5
Pretest Project, Posttest Project and Control	2	3.0	7.5
Posttest Project and Control	4	6.0	13.4
Pre- Posttest Project	8	11.9	25.4
Posttest Project	50	74.6	100.0
Total	67	100.0	
Missing Data	7		

This finding was somewhat surprising to the consultant because a baseline had been established for 19 of the projects and programs that conducted a posttest-only analysis of the project group. It appears that these baseline studies were not utilized by the final evaluators to conduct evaluations with designs that would have been more rigorous (e.g., pre-posttest). There could be any number of reasons for this including accessibility, comparability (in terms of indicators and methodologies) and/or quality of the baseline studies. Or, there could also be reasons associated with the final evaluations. Whatever the reason, establishing a baseline and then not using it in subsequent evaluation efforts is a waste of resources.

Posttest analysis of the project group often employs qualitative methods to collect data. The qualitative methods that were used commonly included document analysis, individual and group interviews, focus groups, site visits, and analysis workshops. Rapid Rural Appraisal methods included transect walks, social mapping, problem identification and prioritization, mobility charts, wealth ranking, problem trees, income-expense trees, scoring games, Venn diagrams, coping strategies indices, seasonalities, pocket charts, suggestion boxes, and SWOT analyses. To enhance qualitative rigor, evaluators often triangulated data from different methods and sources, questionnaires were translated into local languages, advisory groups were established, and teams were formed with international and national members who had complementary skills.

As Table 5 shows, evaluators commonly used surveys to gather baseline and/or final data on CARE projects and programs. Household and community surveys were the most common.

Table 5. Survey Methodology

		Frequency	Valid Percent	Cumulative Percent
Valid	Household Survey	19	42.2	42.2
	Community Survey	15	33.3	75.6
	Institutional Survey	6	13.3	88.9
	Household and Community	2	4.4	93.3
	Household and Institutional	1	2.2	95.6
	Community and Institutional	1	2.2	97.8
	All	1	2.2	100.0
	Total	45	100.0	
Missing Data		29		

Methodological Rigor

A definition of methodological rigor was adopted for MEGA 2004 that is more in line with Campbell and Stanley’s (1963) factors jeopardizing validity. These factors include history (specific events between measurements), testing (changes in procedures and tools), bias in selection of groups, and experimental mortality (changes in subjects). Table 6 shows that 51 factors that jeopardized validity were identified in the reports.

Table 6: Factors Jeopardizing Validity identified in evaluation reports

		Responses		Percent of Cases
		N	Percent	
Rigor	History	17	33.3%	53.1%
	Testing	15	29.4%	46.9%
	Bias	13	25.5%	40.6%
	Mortality	6	11.8%	18.8%
Total		51	100.0%	

Note: To compute percent of responses each N was divided by the total (51). To compute percent of cases each N was divided by the number of reports in which at least one factor jeopardizing validity was identified (32).

- Historical factors included changes to an area’s security status between evaluator visits, civil unrest, invasions, de-escalation of conflict, weather, changes in a country’s economic status, elections, and international terrorism.
- Testing factors included changing evaluation questionnaires in mid-project, changing indicators between baseline and final surveys, delays in establishing a baseline, poor record keeping, and badly formed questions.

- Bias factors in Muslim countries included Western male evaluators not having access to female translators so they were unable to interview female subjects, and the unavailability of subject groups.
- Experimental mortality factors included CARE and government staff turnover, out-migration in response to crises, sites being replaced because wells went dry, and budget cuts that closed project sites.

The above findings suggest that CARE evaluators conduct their work under very difficult circumstances and sometimes at great personal risk. Considering all these challenges, it is often amazing that any evaluation could be conducted at all!

Terms of Reference for conducting program/project evaluations

CARE appears to follow the injunction that evaluations should be conducted as economically as possible (JCSEE, 1994). Analysis of the reports that contained information about the Terms of Reference (TOR) showed that the evaluations ranged between 10 days (JCBCP, Tanzania) to 2 ½ months (CCPF, Indonesia). However, CARE’s desire to economize may be having unintended negative effects on the rigor of the evaluation designs and methods that are employed by some evaluators. (This may help to explain the previously presented findings about evaluation design.)

Some CARE TORs did not allow enough time to do the work required. (The two text boxes below are examples.) Due to the resulting press for time, “compromises were made.” Designs and methods were used that were less rigorous, but which were not as time intensive.

Comments on TOR from a WATSAN in Mozambique:

The TOR required a final review of a three year project over a three week period, with 10 days in the field with the project staff and one week report writing. The TOR assumed easy access to necessary data, but in the event it has taken more than four weeks to obtain full details of communities benefiting and construction details, and also budget and expenditure details. There was no mention of the lack of any surveys or regular monitoring of impact, which made it necessary to undertake a full End of Project Survey during the period of the evaluation. This is normally a separate exercise taking several weeks to design, test, train enumerators, collect field data and analyze it, and as a result it took up a large proportion of the evaluation time, but impact could not have been assessed without it.

A quote from the evaluation of an educational project in Egypt:

Regarding the methodology employed, the team adopted an approach that can be characterized as “more qualitative and anecdotal and less quantitative.” This decision was made early on due to the limited time available (2 weeks for all data-gathering) and

the geographic spread, size and diversity of the target population. A random or weighted sampling that would lead to statistically -significant or representative findings was not realistic given time and other limitations. Instead, the team employed other techniques to ensure that the information gathered was comprehensive, thorough and based on respected approaches to classroom and teacher observation.”

The above findings suggest that CARE may need to reexamine some of its policies to determine if the desire to economize is working at cross-purposes with the level of rigor it hopes to achieve.

CARE DME Standards

Over the past decade, CARE has placed increased emphasis on improving program quality. One of the ways in which it has tried to accomplish this goal is through the introduction of the DME Standards (Rugh, 2005). The DME Standards were endorsed by the CARE International Board in May 2002 and so were in effect during the time when most of the evaluation reports reviewed in MEGA 2004 were written. Despite this, only one report mentioned the DME Standards as guiding the evaluation. (European Union evaluation standards were mentioned more often than the CARE DME standards!) Table 7 shows how well the DME standards were addressed in the evaluation reports that were reviewed. (One caveat: the results presented below are not conclusive. Projects may in fact have met CARE’s DME standards to a greater extent than was conveyed in the evaluation reports.)

Table 7. Project Compliance with CARE’s DME Standards

CARE DME Standards	Responses		Percent of Cases
	N	Percent	
Programming Principles	69	16.2%	95.8%
Link to CO strategy	23	5.4%	31.9%
Stakeholder Participation	36	8.5%	50.0%
Design Based on Holistic Analysis	30	7.0%	41.7%
Logical Framework	22	5.2%	30.6%
Achievable, Measurable Final Goal	14	3.3%	19.4%
Technical, Environmental, Social	29	6.8%	40.3%
Project Costs Addressed	22	5.2%	30.6%
Logframe-based M&E Plan	33	7.7%	45.8%
Adequate M&E Budget	2	.5%	2.8%
Changes Due to M&E Process	15	3.5%	20.8%
Baseline Established	37	8.7%	51.4%
Measurable and Reliable Indicators	34	8.0%	47.2%

Balanced Methods	21	4.9%	29.2%
Organizational Learning	39	9.2%	54.2%
Total	426	100.0%	

Note: To compute percent of responses each N was divided by the total (426). To compute percent of cases each N was divided by the number of projects (72).

Was the project consistent with the CARE International Programming Principles?

According to what could be ascertained from the evaluation reports, almost all of the projects and programs were consistent with at least one of CARE’s Programming Principles. If the proposed CARE Evaluation Policy is adopted, in the future project and program evaluation TORs should more specifically include examining how well the project/program was consistent with these Principles and Standards.

Was the project clearly linked to a Country Office and/or long-term program goals?

Table 7 shows that about 32 percent of the reports established a link between the project and the Country Office and/or long-term goal. There were primarily two ways in which this appeared to happen. There were several pilot projects (e.g., TBA/midwife, Cambodia; CBS CAP, Eritrea) whose explicit purpose was to help inform future Country Office programming. Also, many projects were follow-ons (e.g., WWRH, Ghana; CFP, Indonesia) to previous work that had been accomplished. Both of these approaches demonstrate how CARE is using its programming to promote organizational learning in its Country Offices.

Did the project ensure the active participation and influence of stakeholders in its analysis, design, implementation, monitoring, and evaluation processes?

The degree to which projects ensure the active participation and influence of stakeholders is contingent upon several factors. During emergency or high-risk interventions, CARE appeared to be less participatory. Participation takes time and in an emergency time can mean lives. Table 7 shows that CARE usually attempted to engage main stakeholders in participatory design, implementation, and M&E processes in non-emergency projects. There were some projects (e.g., DZADP, Sri Lanka) in which there were no partners with whom to participate. In such cases, CARE invested resources into organizational development in order to create viable partners.

Did the project have a design that is based on a holistic analysis of the needs and rights of the target populations and the underlying causes of their conditions of poverty and social injustice?

Table 7 shows that about 42 percent of the reports described a holistic analysis of the needs and rights of target populations on which the project was based. There were almost as many ways to conduct the analysis as there were projects. The list includes baseline needs assessments, surveys, feasibility studies, participatory assessments, field surveys, pilot phase of projects, viability studies, rapid food and HLS assessments, awareness consultations, project design workshops, context analyses, and AKAP surveys. Projects

that did not meet this standard either did not conduct a holistic analysis or their analysis was flawed. One project (RAWR, Sri Lanka) reported, “The initial design and then startup was undertaken with limited time and resources [so] an all-inclusive participatory process was not done.”

Did the project use a logical framework that explained how the project was to ultimately impact upon the lives of members of a defined target population?

About 31 percent of the reports provided evidence of a logical framework. Based on his review, it is the consultant’s **opinion** that many of those projects did not use logical frameworks as effectively as they might have. A logical framework can be used to focus an evaluation. Then, in consultation with primary stakeholders, evaluation questions are determined. Afterwards, indicator sets are developed to answer the questions. There was little evidence to suggest that many CARE projects had gone through such a process.

Did the project set a significant yet achievable and measurable final goal?

A set of criteria adapted from Wilson, Robeck, and Michael (1974) was used to determine if the CARE final goals were well-formulated. The criteria included:

- Statement of conditions (context)
- Designation of the beneficiary group
- Use of action verbs that indicate observable activities
- Specification of an outcome
- Specification of the criterion of an acceptable level of performance

As can be seen in Table 7, only about 19 percent of the reports reiterated a final goal that met the above-stated criteria. The best goal statement, which was from a project in Afghanistan (SOLAR II), read:

By December 2002, key livelihood indicators (food production, household income, and health) of 30,000 households in villages of Ghazni, Wordak, Kabul, Logar, and Paktia will have been enhanced to a level that permits sustained resettlement of displaced households

Was the project technically, environmentally, and socially appropriate?

Table 7 shows that about 40 percent of the reports contained evidence of technical, social, AND environmental appropriateness. However, it would take sectoral experts to determine the answers to these questions with more authority, considering accepted current best practice guidelines. It was interesting to note that some CARE projects have made a link between topics as diverse as poverty and environmental degradation and refugee assistance and the environment.

Did the project evaluation address the appropriateness of project costs in light of the selected project strategies and expected outputs and outcomes?

About 31 percent of the reviewed reports provided information on the appropriateness of project costs, making the justification that they were worth the value in terms of outputs.

Did the project develop and implement a monitoring and evaluation plan and system based on the logical framework that ensured the collection of baseline, monitoring, and final evaluation data?

Though not necessarily based on the logical framework, about 46 percent of the reports documented a M&E plan and/or system. The reasons given by projects for NOT doing so include:

- Deficiencies in project design
- No formal M&E unit was established
- M&E unit not properly maintained
- No one was charged with the M&E task
- Vacancies in M&E positions
- Unclear lines of authority and responsibility
- Poor data management
- Visits by external groups were seen as a substitute for evaluation

Did the budget include an adequate amount for implementing the monitoring and evaluation plan?

Only about 28 percent of the reports provided evidence that the project budget included adequate amounts for implementing the M&E plan. One of those projects (HLSP, Macedonia) reduced the international consultant line in order to hire a good evaluation team. An example of inadequate resources was a project (DPAP, Cambodia) with four staff members to monitor more than 400 sub-projects dispersed over a wide geographic area.

Did the M&E process result in changes in project plans, approaches, or strategies?

According to what could be ascertained from the evaluation reports, the M&E process resulted in changes in about 21 percent of the projects. Content analysis revealed that projects most often used evaluation results by implementing the recommendations. For example, one project (FSNC, Ghana) reported that “The process allowed for continuous fine-tuning of project implementation.” However, some projects used evaluation results to modify the mental models that guided their work. For example, one project (CSBESP, Malawi) reported that evaluation “informed a number of project decisions such as reviewing issues on components of the logframe.”

Did the project establish a baseline that could be used for measuring change in indicators of impact and effect (by conducting a study or survey prior to implementation of project activities)?

About 51 percent of the reports mentioned some form of baseline study. The issue of projects not using the baselines to improve the designs of the final evaluations has

already been noted. Findings from this part of the review provide insights into that problem. Several projects identified flaws in the baselines that were established. The most common flaw mentioned was that some baselines were conducted well after project activities had begun. For example, one project (WEDLAN, Ghana) reported that the baseline was conducted nine months after the project had started. Another (IFSP, Bangladesh) reported that the mid-term review was its baseline. In many cases, the baseline was not comparable with the evaluation in terms of indicators measured and methodologies used.

Did the project use indicators that are relevant, measurable, verifiable, and reliable?

In a previous section it was reported that few projects used HLS and/or MDG indicators. This does not necessarily mean that projects are not using indicators. Table 7 shows that at least 47 percent of them do. However, there does not appear to be much consensus among projects on this issue. The following types of indicators are currently being used by projects: conceptual, operational, performance, impact, logframe, progress, result level, efficiency, success, effect, positive, output, process, USAID SEG, and monitoring. Most of these indicator sets are relevant only to their particular project.

Did the project employ a balance of evaluation methodologies, assure an appropriate level of rigor, and adhere to recognized ethical standards?

Table 7 shows that about 29 percent of the reports noted a more or less equal balance of evaluation methodologies. Of the projects that did NOT use an equal balance, most primarily used qualitative methods. A list of the types of evaluation designs and methods used was presented earlier.

Was there evidence documented in the evaluation report that the project design was informed by, and that the findings contributed to, ongoing learning within and outside CARE?

CARE projects have demonstrated a strong desire to contribute to the creation of a learning organization. About 54 percent of them included lessons learned of some form or other. One report from Ethiopia was a compilation of lessons learned over a five-year period.

Accessibility of Evaluation Reports

CARE evaluation reports are supposed to be systematically submitted to the Evaluation Electronic Library (EeL) either directly or via C-PIN. Two main purposes of the EeL are to make evaluation reports available to the bi-annual MEGA synthesis and to make them available for sharing with others around CARE (and beyond). Though it is known there were many more program and project evaluations conducted by CARE during the past two years, only 74 evaluation reports were made accessible for the MEGA 2004 review. Table 8 shows the sources from which the reports were accessed by the reviewer. It should be noted that nine reports were added to the EeL in 2003. However, further

investigation revealed that only one of the nine was completed in 2003. Thus, only one report from the EeL was included in the MEGA 2003-2004 study. Also, there was some duplication between the reports that were sent by email and those on the CDs sent to the consultant by IMLT.

Table 8. Evaluation Report Accessibility

	Numbers	Valid Percent	Cumulative Percent
On-line EeL	1	1.4	1.4
IMLT CDs	71	95.9	97.3
Email attachments	2	2.7	100
Total	74	100.0	

As noted above, CARE Country Offices are learning important lessons from their projects and programs. The EeL could be a way of sharing those lessons across Country Offices. Sharing lessons among Country Offices could help transform CARE into an international learning community (Kim, 2001). Unfortunately, the current state of the EeL does not permit this to happen. Updating the EeL might be worth the investment.

Conclusion

The evaluation challenges confronting CARE seem to be enormous. Evaluation is being planned and implemented in different ways in different countries. The result is that the quality of project and program evaluations is uneven. However, the future is bright. CARE's commitment to the DME standards positions the organization to take the quality of its evaluation work to the next level (Rugh, 2005). Strategies must be developed to promote the DME standards in the Country Offices and to facilitate their application. The DME standards may be the key to transforming CARE into a results-based, learning organization.

References

- Campbell, D. & Stanley, J. (1963). *Experimental and Quasi-experimental Designs for Research*. Boston: Houghton Mifflin.
- Joint Committee for Standards on Educational Evaluation (JCSEE). (1994). *The Program Evaluation Standards* (2nd ed.). Thousand Oakes, CA: Sage Publications
- Kim, D. (2001). *Organizing for Learning*. Waltham, MA: Pegasus Communications.
- Rugh, J. (2005). The CARE International Evaluation Standards. In C. Russon & G. Russon (Eds.), International Perspectives on Evaluation Standards. *New Directions in Evaluation*, Number 104, pgs. 89-94. Jossey-Bass/AEA.
- Stake, R. (1995). *The Art of Case Study Research*. Thousand Oakes, CA: Sage Publications.
- Wilson, J., Robeck, M., & Michael, W. (1974). *Psychological Foundations of Learning and Teaching* (2nd ed.). New York: McGraw-Hill.