

# Lost Opportunities and Constraints in Producing Rigorous Evaluations of USAID Health Projects, 2004–7

Charles H. Teller\*

## 1 Introduction and the problem

Impact evaluation should have a central role in evidence-based policymaking. But, whilst the USAID Health Bureau has prided itself on taking international leadership in evidence-based results monitoring and comprehensive evaluations, attention to evaluation has declined in recent years.

The earlier technical excellence of the USAID Health Bureau in evaluation methods was demonstrated by the development of the LOG FRAME in the late 1970s and subsequent development documents (TIPS: <http://evalweb.usaid.gov/resources/tipsseries.cfm>) and operational guidance for Agency operations (Center for Development, Information and Evaluation (CDIE), USAID, [www.USAID.gov/us](http://www.USAID.gov/us); Automated Directives System (ADS), USAID, [www.USAID.gov/policy](http://www.USAID.gov/policy)). But since 1994 there has been a steep decline in the quantity and quality of evaluation; the number of evaluation studies falling from a peak of 497 a year to a low of 104 in 1998 (Clapp-Wincek and Blue 2001). This decline has been attributed to many factors, including Albert Gore's policy change on 'reinventing government', a sharp reduction in technical staff, and a change in USAID Guidance from requiring every project to be evaluated to recommending that evaluations *only be done in response to management need*.

Since 2000, there have been several Agency-wide and Bureau-specific reviews of evaluation (e.g. Clapp-Wincek and Blue 2001; and Weber 2004), which have demonstrated the loss of institutional learning and best practices. While the country USAID missions depended on evaluations, their greatest concern was the very limited number of in-depth programme evaluations. Moreover, while the

partners did many of the evaluations, USAID did not (Clapp-Wincek and Blue 2001: iii), and those few evaluations supported by the missions were not being submitted to the Development Exchange Clearinghouse (DEC).

Evaluation quality has long been an issue with USAID, both Agency-wide (e.g., Hopstock *et al.* 1989) and in the technical bureaux (e.g. Adamchak *et al.* 2004 for Health; Bollen *et al.* 2005 for Democracy and Governance). Many evaluation reports do not include more than a few paragraphs on method, and many were qualitative and unsystematic: the expatriate 'fly-in' assessment where a team comes for 2–3 weeks and bases its findings only on qualitative interviews with key informants and stakeholders. Most reviews of USAID evaluations found weak methods employed, even with external, professional evaluators whose objectivity was often compromised by their desire to please the managers and continue to be hired. This article documents these trends in recent years, analysing the factors behind them and the steps required to ensure the production of more and better impact studies.

The focus of this study is on the methodological strength and design rigour of evaluations of outcomes, effects, and impacts. White (2006) defines impact as the counterfactual analysis of how an intervention affects final welfare outcomes. In that sense, we want to know if the donor-funded activities are attaining their expected results as set out in the project paper and the results framework or the monitoring and evaluation (M&E) design. However, since donors fund too few real impact evaluations of project attribution and the counterfactual, evaluators have often been limited to considering whether the project achieved

---

**Table 1 Classification system for evaluation designs**

---

**Revised classification of USAID health project evaluation design**

---

**A. Programme interventions and project components:**

1. *Post hoc* cross-sectional
2. Baseline & post-intervention
3. Baseline & post-intervention with comparison/control area/group (QED-light)

**B. Child Survival projects:**

1. *Post hoc* cross-sectional
2. Baseline and post-intervention

---

its intended outcome in its intervention areas or groups, preferably as compared to control groups. Moreover, in order to inform future programming, evaluation must be transparent and externally credible to decision-makers. However, without the rigour of an impact evaluation these attributes are harder to achieve.

**2 Methodology**

There are three main approaches taken in this article:

- 1 A document review of evaluation design, methods, training, and technical leadership in M&E was carried out. These included activities of the flagship USAID/Global Health (GH) evaluation-related projects: MEASURE/Evaluation (Monitoring and Evaluation to Assess and Use Results); Frontiers and Horizons Operations Research; Data for Decision-making, including: State of the Art papers, M&E Reference and Working Groups, Task Forces; International M&E and Operations Research Regional Training and Capacity-building; M&E Working group survey and needs assessment (2005); and finally, M&E manuals, guidelines, indicators (see website at: [www.cpc.edu/measure/evaluation](http://www.cpc.edu/measure/evaluation)).
- 2 A meta-analysis review of Global Health evaluation documents and their design in the last few years. These included: Strategic Planning and Results/Portfolio Reviews in 2005 and 2006, which were a stocktake of the number and types of evaluations per technical office; and a systematic meta-analysis of Global Health evaluations of designs and methodological approaches of 93 GH-funded projects, sub-projects and interventions from 2004–6. Table 1 shows the revised classification during the content analysis of evaluation designs.

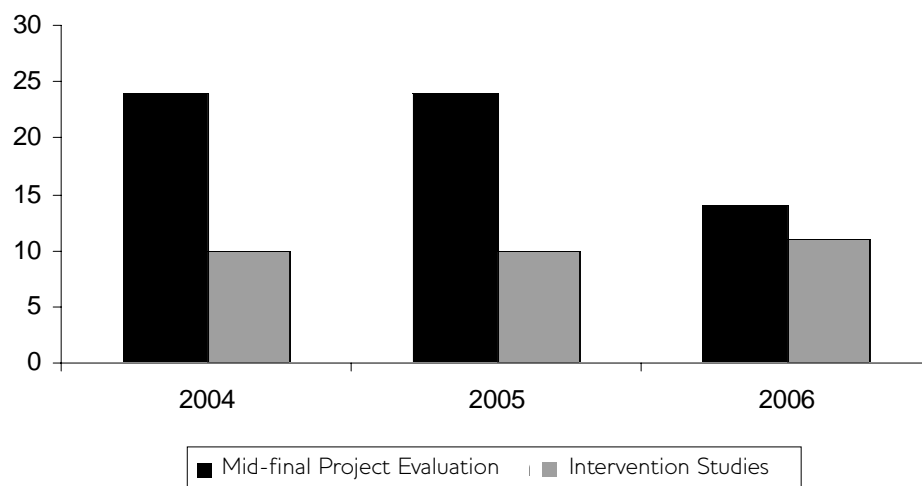
- 3 An opinion survey and follow-up interviews with 47 experienced USAID evaluators. These surveys covered ten content areas, soliciting their experience about methodological rigour; the quality of the team; the issues raised; technical, operational, and behavioural constraints; and the use of the findings for policy and programme change. There was dissemination of the findings, much discussion, and further in-depth investigation.

**3 Findings and results****3.1 Agency-wide and global health planning and leadership in evaluation**

Given the deterioration in quality and quantity of evaluations in the 1994–2004 period, and the highly critical 2004 assessment of evaluations (Weber 2004), the USAID Administrator, Andrew Natsios, decided to revitalise evaluation, with the active implementation support of CDIE. Following direct consultations with field missions and central bureaux (Adams 2005; Kerley and Croake 2005), the Administrator sent a directive to the field outlining the strategy and actions needed, along with a much-needed training package (Natsios 2005). This process was cut short after less than a year by his replacement with a new Federal Director of Foreign Assistance.

Under Ambassador Tobias' new administration, CDIE, with its experienced evaluators, was disbanded, and its mother bureau moved over to the State Department, where all M&E was to be coordinated. The consequent Federal/State office dropped the revitalisation plans for evaluation, in favour of strict activity and output reporting (Operational Plans) for budgeting purposes. Moreover, the quantitative and analytical Results Frameworks and Packages, Strategic Objectives and Intermediate Results, and Performance Monitoring Plans (PMPs) were down-played, and

**Figure 1 Annual number of global health evaluations in the USAID clearinghouse: 2004–6**



outcome-oriented annual reports were reduced to annual operational plans for output reporting.

In 2003–4, the Global Health Bureau was entering into a paradigm shift whereby the Emergency Presidential Initiatives in HIV/AIDS (PEFPAR), in malaria (PMI), and Avian Influenza were also shifting away from evaluation and towards strategic information systems and output monitoring. Federal/State required the development of common indicators under eight common objectives or ‘elements’ and cross-cutting sub-elements. Whilst defining common indicators assists in the alignment of monitoring indicators with Agency objectives and the aggregation of results, there is a tension with the need for M&E systems to be context-specific. More importantly, outcome monitoring is not the same as evaluation since it says nothing about attribution.<sup>1</sup>

However, the Family Planning/Reproductive Health and Maternal and Child Health and Nutrition office management and technical advisors agreed to maintain their technical leadership in evaluation, and developed several strategies to strengthen M&E. They:

- 1 Supported their own Assessment of Programme M&E (Adamchak *et al.* 2004) which recommended less monitoring and more rigorous evaluation;
- 2 Formed an M&E working group among the USAID partners and collaborating agencies, which

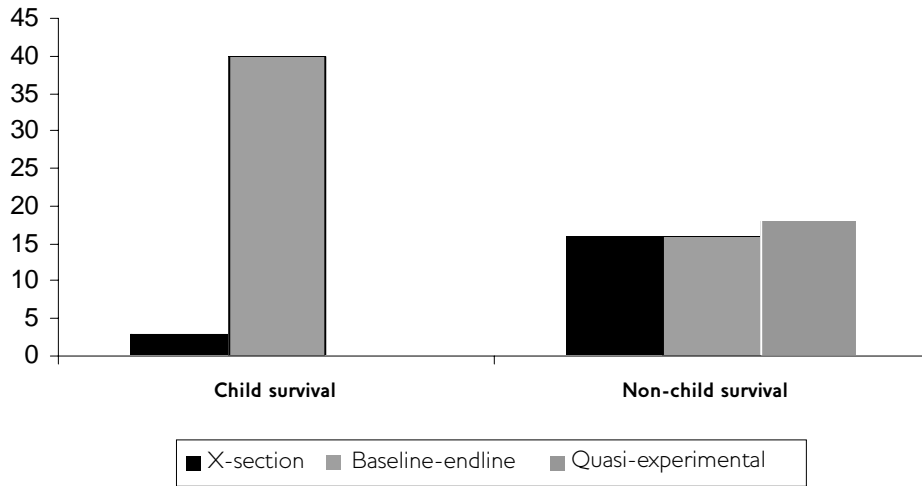
recommended more training and sharing of lessons learned (Teller and Pandit 2004);

- 3 Developed a new conceptual framework and indicators, designed programme ‘pathways’ and coordinated an annual results review;
- 4 Used the showcase MEASURE, and sub-cooperating agencies, to refine the tools and methods necessary to generate, analyse, coordinate, and disseminate information for evaluation, programme learning, and capacity-building, and to develop a programme results system.

### 3.2 Findings of the meta-analysis of methodological rigour

Here we highlight those findings relevant to the rigour of the design of outcome and impact evaluations, both for projects and sub-project components. In the introduction to this issue, White and Bamberger define the stronger ‘real world’ evaluations as those with a quasi-experimental design, particularly those which have baseline and end-line data and both treatment and control groups (see also Bamberger *et al.* 2006). The USAID clearinghouse contains only 31 evaluation studies a year for the period 2004–6 (Figure 1, a total of 93), though as already noted, not all documents are submitted and others are restricted as procurement-sensitive.

**Figure 2 Evaluation design classifications: 2004–6**



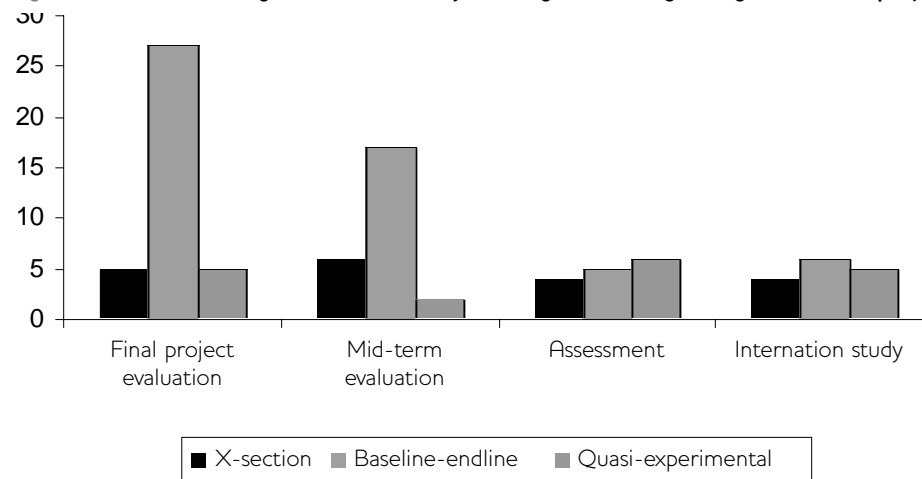
Most of the 93 assessments on evaluation design were pre-post child survival project designs (Figure 2), as required by the USAID Child Survival Grants Program, while other studies had *post hoc* end-line designs, which are often unable to deliver reliable impact estimates. The non-child survival evaluations had stronger designs: quasi-experimental with comparison groups (Figure 2). In the DEC classification scheme, most of those classified as assessments or intervention studies were more rigorous than those classified as evaluations (Figure 3).

### 3.3 Expert survey and in-depth interviews

From the list of Team Leaders, frequent evaluator practitioners, and USAID evaluations advisors, over 60 experts were selected. Most were sent semi-structured surveys and did respond, sometimes in person; over 90 per cent were followed up by telephone or inter-personal interviews. The main findings are:

- **Evaluation planning and design:** there is a lack of adequate project needs assessments prior to baselines; an over-reliance on outputs and

**Figure 3 Evaluation design classifications by clearinghouse categories global health projects 2004–6**



**Table 2 What and when is it best to evaluate?**

What	When
<ul style="list-style-type: none"> <li>● M&amp;E systems built into project proposal before award decision</li> <li>● Critical interventions need strong evaluations</li> <li>● Innovative activities and interventions whose practical effectiveness not yet rigorously evaluated</li> <li>● Multidisciplinary and contextual factors (the counter-factual; value-added)</li> </ul>	<ul style="list-style-type: none"> <li>● Decisions on replication, extension, scaling up</li> <li>● At end of 4th year of most five-year projects</li> <li>● Monitor extensively, evaluate selectively'</li> <li>● Major challenges need in-depth analysis into the causes</li> <li>● Demographic and Health Survey (DHS) data show unexpected changes</li> </ul>

milestones, instead of on project outcomes or attribution; and project designs are often driven by externalities and 'fashions of the month'.

- **Evaluation implementation:** team members, stakeholders, and counterparts were often inexperienced and untrained in evaluation methods, e.g. former USAID Mission directors hired as Team Leaders despite little hands-on experience in evaluation methods. Also, USAID managers were critical of external evaluators coming in with their own 'agendas,' lacking enough knowledge of project context, and seeking to please senior administration officials. We also found a lack of an enabling environment from senior management and an erosion of the culture of evaluation in the field.
- **Follow-up and use of evaluations:** a common criticism was the undue influence of key stakeholders on final recommendations. Most frustrating to the evaluation experts was that critical project decisions were made before evaluation finished. Evaluators rarely found out whether or not their key recommendations were carried out.

A key question raised by USAID strategic health planning, other technical, and programme offices in preparation for designing a new evaluation agenda, given the shortage of time and resources, is what and when is it best to evaluate. Table 2 shows a summary of expert opinion on the subject.

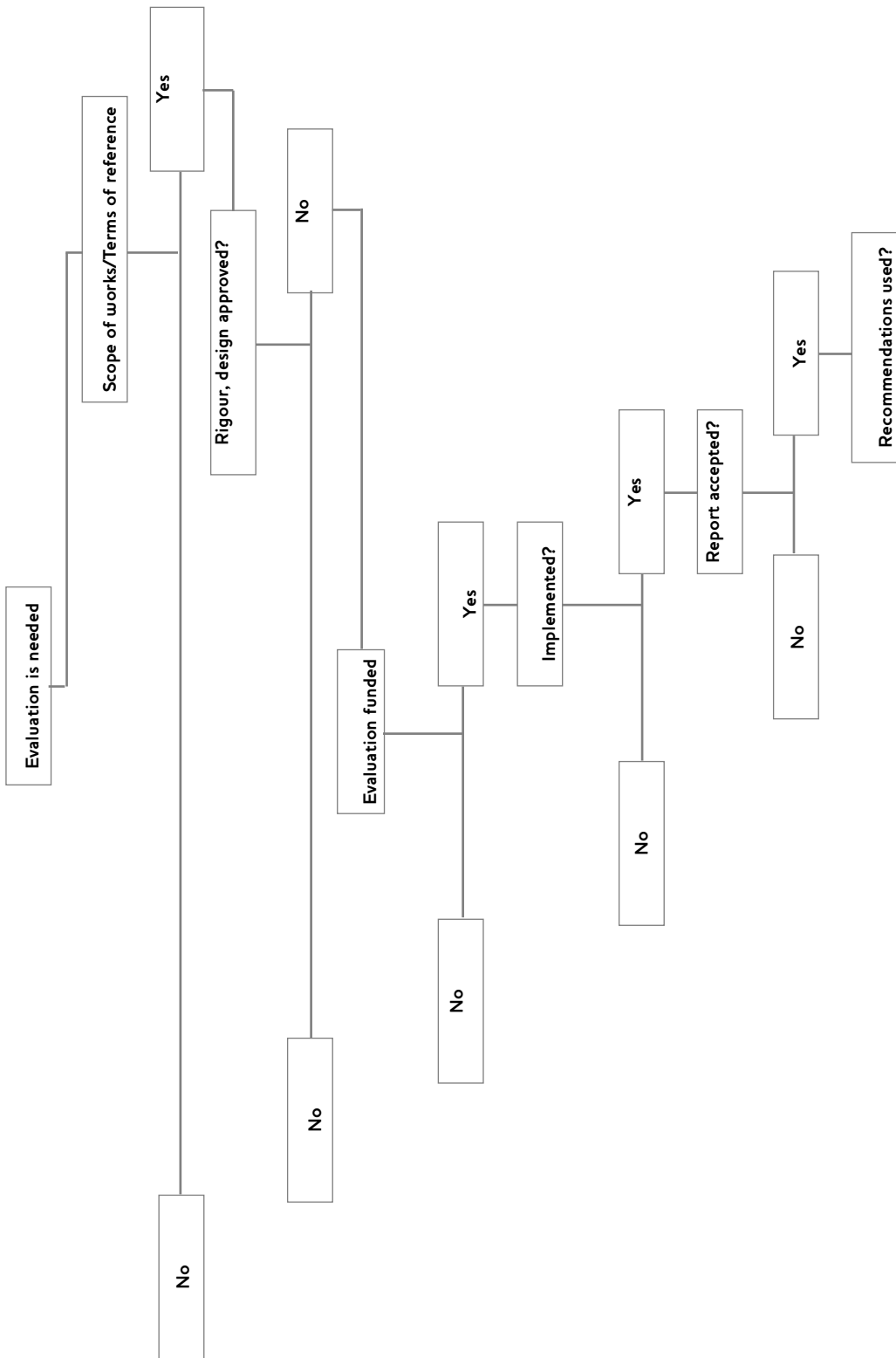
**4 Discussion: constraints to producing and using rigorous evaluations**

There are no doubt many constraints to designing and implementing strong, rigorous health

evaluations, as noted in the newer articles, including White and Bamberger in this issue (see also Bamberger *et al.* 2006; Victora *et al.* 2004). From my experience as evaluation advisor within the Global Health Bureau, inside USAID country missions, while on secondment to developing country governments, and as a consultant in M&E for health sector NGOs and private sector partners, I would list the main constraints recorded from the experts inside and outside USAID/GH in four categories:

- 1 **Technical:** Contemporary impact evaluation designs need to be rigorous, requiring multivariate analytical skills. Training opportunities for these skills are limited, and experience in their application is even more limited. Moreover, some of the content of these training courses is not that useful in the real world of health and development (Reynolds *et al.* 2006) or in poor and unstable countries. There is concern amongst donors over lack of usefulness of some past evaluations, especially when these have been superficial or carried out at times when they were not needed. Moreover, M&E experts are located more within institutional partners and independent consultants than within the donor.
- 2 **Organisational:** There is common knowledge of hiding evaluations: many evaluations are not made public, either because they are procurement-sensitive and then never released, too critical, or poorly done. Organisations tend to self-promote and apply a positive 'spin' in their own self-interest, advertising competencies for new proposal competition. Moreover, there has been a decline in the number and continuity of USAID technical advisors and contract officers, and a lack

Figure 4 Missed opportunities for rigorous evaluation – a tree of leakages from assessed need to unused recommendations



of senior M&E positions in many country missions and Washington GH offices. As in most international organisations, there is tension over local autonomy of programme resource use between central and country USAID offices.

- 3 **Political:** The Agency's effort in 2005 to revitalise evaluation demonstrated the inadequate political will to make sufficient resources available for its implementation. What took priority in 2006 was the explicit absorption of USAID into the new State Department policy goal of Transformational Diplomacy and a focus on politically important countries (e.g. Egypt, Israel, Iraq, Afghanistan, Pakistan). The fear that negative evaluations will play into the hands of the foreign aid critics in Congress and the State Department produces fear of the visibility of failures and mistakes.
- 4 **Behavioural:** These restraints include the career ambitions, roles, and interests of individuals and affinity groups to act for their own benefit. Careers can be threatened, given the lack of a reward system for strong evaluations in order to learn from mistakes. There is a lack of managerial motivation, and a lack of professional incentives for technical advisors, for rigorous evaluations. Finally, the time pressures of administrative tasks often crowd out the institutional learning tasks.

Figure 4 represents in schematic form the frustrating experience of what has often happened to the proposed or actual designs of rigorous evaluations in the last four years. It is estimated that few (less than 20 per cent) of these designs were either approved for funding, or implemented with rigour. All the real world constraints came to bear in creating this major leakage of credibility.

#### 4 Conclusions and recommendations for institutional learning

This article has documented the precipitous decline in the number and quality of USAID-wide evaluation designs accessed through the USAID Document Exchange Clearinghouse (DEC) between 1994 and 2006. We used three methods to analyse this decline and the factors behind it:

- 1 Results and Portfolio Review assessment of evaluation of over 100 ongoing centrally funded projects in Global Health in fiscal year (FY) 2005–6;

- 2 Meta-analysis of 93 project and sub-project/ intervention evaluations done in 2004–6 accessed in the DEC; and
- 3 Opinion survey and interviews of 47 experienced USAID evaluation experts, both internal and external, about their issues and suggestions.

The findings point to the legitimacy of the concern of senior management about the weaknesses of the Results Review component of the Portfolio Review Process. There are too many missed opportunities as well as institutionalised constraints to learning more about what works and what needs improvement in the GH portfolio. Results reporting cannot take the place of impact studies to examine attribution. Thus transparent and evidenced-based policy and strategic planning decisions have been compromised in the process. The constraints to stronger evaluation discussed above – technical, organisational, political, and behavioural – are formidable, and will require a concerted effort to overcome.

#### 4.1 Implications and recommendations

After the aborted attempt of the Natsios administration to revitalise evaluation, and the Federal/State focus on short-term operational plans and budget-related output monitoring, the GH Bureau has been considering measures to revitalise evaluation. The demise of CDIE, the lack of an autonomous M&E unit in Global Health, and the complicated US Government inter-agency evaluation structures of the Presidential Health Initiatives all work against a credible, flexible, and transparent approach to programme evaluations.

The development of a new evaluation agenda during FY 08 is an important first step for USAID, though it must be supported by the political will to allocate sufficient priority to and resources for both central and mission learning needs. This could be considered as moving forwards in a more coordinated manner with other donors on more rigorous impact and programme evaluation. A minimum set of recommendations for this new evaluation agenda is suggested below.

- A high-level, autonomous (from the technical offices) evaluation group (AEG) or division, including strategic information and analysis, within the Programme Office. It should be able to design rigorous impact and outcome evaluations,

including cost-effectiveness; complemented by a stable consultant group of independent health evaluation experts;

- An evaluation agenda to be developed by the AEG in consultation with the technical offices and given a proper place within the new Health Strategy;
- Rigorous programme evaluations, guided by the Bureau-wide evaluation, every five years before the next project approval cycle; needs assessments, baselines, and mid-term M&E systems in most projects;

### Notes

\* Formerly Senior Technical Advisor in Monitoring/Evaluation, Global Health Bureau, USAID, Washington, DC.

- 1 Reference is made here to the triple-A requirement for Agency-wide monitoring

### References

- Adams, D. (2005) 'Initiative to Revitalize the Evaluation System', memo to missions, July, USAID
- Adamchak, S.; Reynolds, J. and Henn, J. (2004) *Assessment of the M&E in Projects Managed by the BGH, OPRH; LTG Inc. and SSS Inc.*, Washington DC: USAID
- Bamberger, M.; Rugh, J. and Mabry, L. (2006) *Real World Evaluation: Working Under Budget, Time, Data and Political Constraints*, California: Sage Publications
- Bollen, K.; Paxton, P. and Morishima, R. (2005) 'Assessing International Evaluations: An Example from USAID's Democracy and Governance Program', *American Journal of Evaluation* 26.2
- CGD (2006) *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, Washington DC: Center for Global Development
- Clapp-Wincek, C. and Blue, R. (2001) *Evaluation of Recent USAID Evaluation Experience*, Working Paper 320, Washington DC: CDIE, USAID
- Hopstock, P.; Kellum, A. and Young, M. (1989) *Review of the Quality of A.I.D. Evaluations FY 1987 and FY 1988*, A.I.D. Evaluation Occasional Paper 19, Washington DC: Agency for International Development
- Kerley, J. and Croake, K. (2005) *Report of the Results of the Survey on Evaluation Revitalization*, Washington

- Capacity-building, professional development, and training, guided by best practices in M&E, to be supported for M&E and programme design staff – both within USAID and for partners;
- A responsive knowledge and documentation centre reorganised to serve the information, analysis, evaluation, and decision-making needs of the Global Health Bureau. This should be organised in coordination with the Agency-wide centre.

systems: aggregation (being able to aggregate results across interventions), attribution (being able to link changes in outcomes to the intervention), and alignment (if indicators of project success indicate achievement of the agency's overall goals); see White (2004).

- DC: Policy and Program Coordination, Center for Development Information
- Natsios, A. (2005) 'Actions Required to Implement the Initiative to Revitalize Evaluation in the Agency', USAID directive cable, 17 June
- Reynolds, J.; Webb, L.; Augustin, R. and Kindsfather-Lopez, M. (2005) *Evaluating Short-term Training in Health Program Evaluation*, Washington DC: USAID
- Teller, C. and Pandit, T. (2004) 'M&E Opinion Survey Results', draft, GH/PRH2004, (Bureau of Global Health/Office of Population and Reproductive Health)
- Victoria, C.G.; Habicht, J.-P. and Bryce, J. (2004) 'Evidence-based Public Health. Moving Beyond Randomized Trials', *American Journal of Public Health* 94
- Weber, J. (2004) *An Evaluation of USAID's Evaluation Function: Recommendations for Reinvigorating the Evaluation Culture within the Agency*, September, Washington DC: PPC, USAID
- White, H. (2006) *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*, Washington DC: IEG, World Bank
- White, H. (2004) 'Using the MDGs to Measure Agency Performance' in R. Black and H. White (eds), *Targeting Development: Critical Perspectives of the Millennium Development Goals*, London: Routledge