SPECIAL REPORT

Expanding the Scope of Humanitarian Program Evaluation

Paul Bolton, MBBS, MPH;[1] Judith Bass, PhD;[1] Laura Murray, PhD;[1] Katharine Lee, MA, MPH;[2] William Weiss, DrPH, MA;[1] Sharon M. McDonnell, MD, MPH[3]

1. Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland USA
2. Harvard Business School, Allston, Massachusetts USA
3. Dartmouth Medical School, Hanover, New Hampshire USA

Correspondence:
Paul Bolton, MBBS, MPH
Center for Refugee and Disaster Studies
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St., Room E8646
Baltimore, Maryland 21205 USA
E-mail pbolton@jhsph.edu

The case studies described here were funded directly by the implementing partners: FilmAid International (Kenya), World Vision and War Child Holland (Northern Uganda).

Keywords: disaster; evaluation; humanitarian;impact; method; objective; program; response

Abbreviations:
None.

Web publication: 11 October 2007

Abstract

The effectiveness of humanitarian programs normally is evaluated according to a limited number of pre-defined objectives. These objectives typically represent only selected positive expected impacts of program interventions and as such, are inadequate benchmarks for understanding the overall effectiveness of aid. This is because programs also have unexpected impacts (both positive and negative) as well as expected negative impacts and expected positive impacts beyond the program objectives. The authors contend that these other categories of program impacts also should be assessed, and suggest a methodology for doing so that draws on input from the perspectives of beneficiaries. This paper includes examples of the use of this methodology in the field. Finally, the authors suggest future directions for improving this type of expanded assessment and advocate for its widespread use, both within and without the field of disaster response.

Bolton P, Bass J, Murray L, Lee K, Weiss W, McDonnell SM: Expanding the scope of humanitarian program evaluation. ***Prehospital Disast Med*** **2007;22(5):390–395.**

Introduction

Impact evaluation is a core element of humanitarian assistance programs since it provides the basis for improvements in future programs.[1–5] A positive development during the last decade has been the movement to introduce standards and measures of health status into program impact evaluations.[6,7] Examples of standards that are measured include: (1) liters of water per person per day; (2) square meters of shelter space per person; (3) measles vaccination coverage; and (4) kilocalories per person per day. Common health status measures that now are included in the evaluation of humanitarian assistance programs include: (1) number of deaths per day; (2) percent of children <5 years of age with acute malnutrition; and (3) incidence rates of measles and acute watery diarrhea.[8,9] Recently, functional status has been included as a measure of health status in the evaluations of psychosocial programs.[10] A further positive development is the effort to standardize the measurement(s) of health status in emergencies.[9] All of these measures are an improvement over the past practice of evaluation that was based on program outputs (such as the number of tons of food shipped, the number of shelter tarpaulins distributed, or the number of latrines constructed) because program outputs alone do not demonstrate the actual impact of the program on the lives of those being served.

While humanitarian programs have clearly made progress in assessing impact, such assessments typically are limited to assessing a small number of pre-defined objectives. This limited focus has several benefits including programs and associated information systems that are easier to implement, standardize, control quality, and are less resource-intensive. However, humanitarian assistance programs can have important, unexpected impacts beyond these pre-determined objectives.[1,11–14] An example of an unexpected impact emerged during a qualitative study by two of the authors (PB and WW), in Angola during the late 1990s. Interviewers were told by Angolan villagers how the glint of new metal roofs in the sun, visible from great distances, had

https://doi.org/10.1017/S1049023X00005100 Published online by Cambridge University Press

become a marker to both rebel and government forces that a village had received assistance, resulting in raids by the soldiers. Despite initial enthusiasm for the metal sheeting, it no longer was wanted or used and was discarded. In an Afghan refugee program, another author (SM) observed a medical agency initiating a health education project that included distributing soap to women. In their assessment of health risks, the agency had determined that lack of soap was a key risk factor in an outbreak of impetigo among children. The local population viewed this distribution as encouraging the women to have sexual intercourse, resulting in camp leadership shutting down the agency clinics and boycotting other agency services.

There is an ethical obligation for those who implement aid programs to do all that they can to ensure that these programs do more good than harm, and to constantly strive to maximize benefit. This requires identifying and evaluating as many of the program's major impacts as possible, both positive and negative. This cannot be achieved with the current practice of assessing changes in a small number of pre-defined objectives. Shortcomings of the current evaluation practice include the following problems:

1. *Bias toward finding positive impacts*—Program objectives usually only refer to the positive impacts of the program. If these are the only impacts that are assessed, this constitutes a bias towards positive findings, since the worst that such a program evaluation could find is a zero impact.
2. *Inability to explore the basic question of whether a program has done more good than harm*—This is a direct result of Problem 1, and of the fact that programs can achieve their objectives and still have an overall negative effect if the negative impacts outweigh the positive gains. Therefore, while program evaluations should continue to explore the achievement of predetermined objectives, evaluators and researchers also should identify and measure as many positive and negative impacts of the program as they can. Otherwise, we cannot be confident that a program has done more good than harm.
3. *Limited information on which to make program improvements*—Program evaluations that only explore (and demonstrate) success with regard to their predetermined objectives provide little information on how the project can improve in the future. Yet all programs, no matter how successful, should aim to improve continuously,[1,5,6] and program evaluations should be designed to provide as much data as possible in order to inform these improvements. This involves not only how to maximize positive impacts but also how to minimize negative impacts, therefore requiring that both types of impact are evaluated.

In this paper, we propose a methodology for humanitarian assistance program evaluation that expands the scope of current practice to examine both positive and negative types of impacts. First, we describe four broad categories of positive and negative impacts and give examples of each. Second, we describe the proposed evaluation methodology and how it can assist in exploring these impact categories. We also compare this methodology with current evaluation practice. Third, we describe the results of two case examples of this proposed evaluation methodology. Finally, we discuss the implications of our results and make suggestions for future directions.

Categories of Program Impact

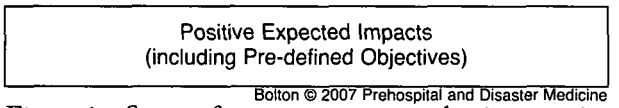
The rationale for expanding the basis of program evaluation is based on a model comprised of four broad categories of program impact. Under this model, the current basis of program evaluation (selected pre-defined objectives) forms part of a wider category titled "Positive Expected Impacts" as illustrated in Figure 1. The existence of additional categories of program impacts is represented in Figure 2. The implication of both figures is that there is much more to explore (in fact, whole categories of impacts) that currently is not well-addressed by program evaluation. Examples for each of these categories are given below:

Expected Positive Impacts—The expected positive impacts include the major objectives of the intervention (the basis of current evaluation practice), and other positive effects that may be anticipated, but are not the main purpose of the program. Common examples include: (1) economic benefits from employing local workers; (2) the entertainment value of health education programs; and (3) an enhanced sense of security (real or imagined) that the presence of outside organizations can provide among displaced populations or other populations under threat.

Expected Negative Impacts—Expected negative impacts usually are described as "tradeoffs" or opportunity costs that are known to the community and/or program staff but are considered acceptable given the anticipated benefits. A common type of expected negative impact is the opportunity cost incurred by program beneficiaries as a result of their participation in the program interventions.[15,16] For example, the time a family must spend participating in a child health program might have been spent earning wages or accumulating other benefits. Similarly, a family may have to pay for transport, meals, and healthcare visits to bring their child for vaccinations or treatment. These types of opportunity costs rarely are considered during program impact assessments.[17]

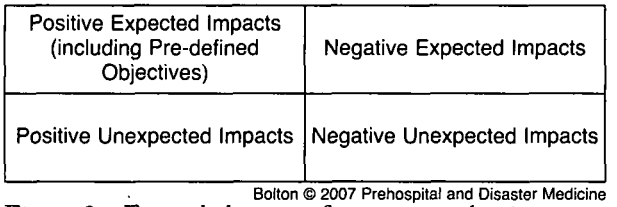
Unexpected Impacts (both positive and negative)—Programs have unexpected positive and negative impacts that, by definition, are unknown at the time the program or intervention is developed and implemented.[5,18] Yet, these unexpected impacts may be important and may be among the program's more durable effects on the lives of recipients and the wider community. For example, humanitarian assistance unexpectedly can create or aggravate religious, ethnic, or racial divisions within or between communities.[19] Increased supplies of money, resources, and even political attention, when combined with competing interests, can lead to substantial changes in the relative power of groups in a community.[20]

https://doi.org/10.1017/S1049023X00005100 Published online by Cambridge University Press



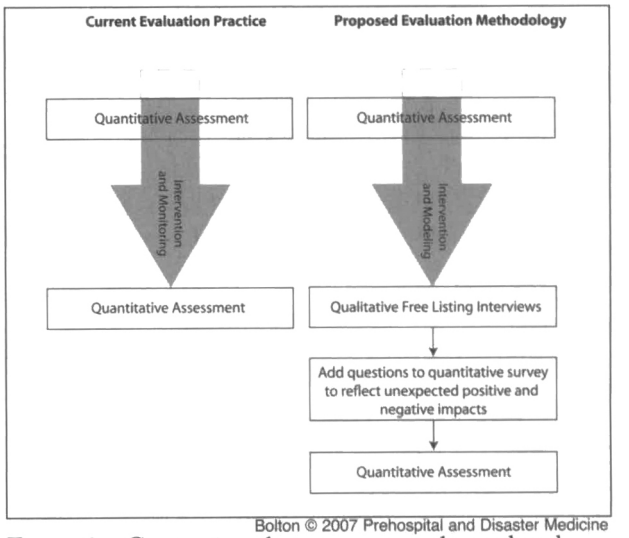
Bolton © 2007 Prehospital and Disaster Medicine

Figure 1—Scope of current program evaluation practice



Bolton © 2007 Prehospital and Disaster Medicine

Figure 2—Expanded scope of program evaluation



Bolton © 2007 Prehospital and Disaster Medicine

Figure 3—Comparison between currently used and proposed evaluation approaches

Methodology for Expanded Humanitarian Assistance Program Evaluation

Overview of Proposed Evaluation Methodology and Comparison to Current Practice

The most commonly used approach to impact evaluation is to use identical quantitative instruments (such as questionnaires) to conduct pre- and post-intervention assessments. The quantitative instruments reflect the pre-defined program objectives. The changes in the measures pre- and post-intervention are used to assess program impact. These methods, although not the subject of this paper, are well-described elsewhere.[14]

Rather than replace this approach, the authors propose an additional step with the purposes of identifying and assessing unexpected impacts and of identifying which expected impacts are likely to be significant. In designing this approach, there were two considerations. First, the approach should focus on delineating important unexpected and expected impacts (both positive and negative) from the viewpoint of those who receive the program. Prioritizing the local perspective likely will result in improvements that matter to the target populations, thereby enhancing how highly programs are valued and accepted by those they seek to serve. Second, the approach would be implemented after the intervention. Because of the large number of possible impacts and the likelihood that only a few actually will occur, an approach was selected that does not attempt to predict impacts but rather to identify those that actually have occurred.

This approach is adapted from a mixed qualitative and quantitative assessment methodology previously developed and used for needs assessments and program planning.[21–23] That methodology uses qualitative methods to develop locally appropriate instruments for use during the post-acute disaster period, or to adapt existing questionnaires to better match local needs.[10,24,25] Along the same lines, we have added a qualitative data collection method-free listing to the normal pre- and post-intervention quantitative assessment. Free listing is used to characterize unexpected and expected program impacts. Questions on these impacts are then added to the post-intervention quantitative assessment questionnaire used to reassess program participants (Figure 3).

Description of Proposed Evaluation Methodology

1. *Pre-Intervention Quantitative Assessment*—The proposed evaluation methodology begins with a pre-intervention, quantitative assessment that provides baseline measurements relevant to pre-defined program objectives. Types of quantitative assessments could include household surveys, surveillance, or vital registration systems. The quantitative instruments usually are standardized questionnaires.
2. *Post-Intervention Qualitative Free Listing Interviews*—After a program intervention has been completed—or in the case of ongoing programs, after sufficient time has elapsed for program impacts to have occurred—a convenience sample of about 20 program participants and other affected persons (such as family members) are interviewed using a free-listing qualitative interviewing method. As with other qualitative methods, interviewers rely on open-ended questions, are careful not to lead the respondent, probe the respondent for as much information as they can provide, and record all responses verbatim. This approach encourages respondents to say any and all ideas that occur to them during interview. The purpose of these interviews is to identify as many impacts as possible in all four impact categories. This approach is suited particularly to identifying impacts that are unexpected to the researchers.

   The free-list interview begins with the following primary question: "What are all the changes that have occurred in this community since the intervention/program began?" This question is designed to elicit responses in the form of a list. Participants are queried for as many responses as possible and are asked for a short description of each response. Both the responses and the short descriptions are recorded verbatim on a standard form. If the term, "since the intervention/program began," is difficult to translate, the starting date of this program may be

https://doi.org/10.1017/S1049023X00005100 Published online by Cambridge University Press

linked to a more salient event, such as the arrival of certain staff, a well-known event or holiday, or some other clearly demarcated point-in-time.

Once the first free-list is completed, interviewers immediately conduct a second free list with the same respondent as part of the same interview. The primary question for the second free list is: "What are all the changes that are the result of the intervention/program?" While the second free list question is more directly to the point, the interview always begins with the first free list question. This makes it possible to identify possible effects of program interventions that may not be recognized as such by the participants (but would be recognized as effects by the program implementers or others). Analysis consists of combining the results of each primary question into a composite list consisting of all of the responses and the number of interviewees who gave each response. The lists rank each response from highest to lowest frequency. Items that are mentioned frequently and that the evaluators consider to be important and related to the intervention (on the basis of the descriptions) are considered to be potentially significant program impacts. These items are selected for further investigation in the next stage—the post-intervention quantitative assessment.

The resources required to perform the free-listing interviews include the following: (1) 5–10 interviewers; (2) two days of training; (3) 2–4 days to conduct interviews; and (4) 1–2 days for analysis. The main requirements for interviewers are the ability to speak, read, and write in the local language. While prior interviewing experience is helpful, it is not a critical requirement.

3. *Post-Intervention Quantitative Assessment*—As usual, the pre-defined objectives are evaluated by repeating the pre-intervention quantitative assessment after the intervention has been completed and comparing the results. In this proposed expanded evaluation approach, the items identified in the free list interviews as potentially significant expected and unexpected impacts are formulated into questions. These questions can be added to the original quantitative assessment instrument used in the pre-intervention assessment. These new questions state the nature of the impacts identified from the free lists and use a Likert scale to assess the level of agreement (or disagreement) as to whether this impact occurred and was important. Analysis of the additional questions consists of measuring the prevalence and level of agreement as to whether each impact actually occurred. In this way, the impacts identified by a convenience sample of program participants during the qualitative study are quantified among a sample of program participants even though there are no pre-intervention results for comparison.

Case Examples

Some examples that illustrate the dramatic ways that humanitarian programs can have unexpected effects were introduced above. In both cases, due to their extreme nature, these effects quickly became apparent. However, unexpected effects are not always so easily identified in the absence of a formal assessment. This is illustrated by two case examples that use the proposed evaluation methodology.

Case Example 1—Kenya (Refugees)

The proposed evaluation methodology was first used in a post-disaster setting to explore expected and unexpected impacts of a program that showed films to refugees in a large refugee camp in northern Kenya. The objectives of the program were to increase knowledge and improve behavior with respect to human immunodeficiency virus (HIV) transmission, reduction of violence and conflict, and promotion of human rights issues (particularly the rights of women). The program had preceded the beginning of the evaluation, so there was no pre-intervention quantitative assessment. Instead, program evaluation began with a qualitative study to identify changes that program participants believed had occurred over the course of the program implementation. These changes were added to a draft instrument based on the program objectives, which was then used in a household survey of the camp.

Several unexpected and expected impacts were identified in the qualitative study and confirmed by significant numbers of respondents in the quantitative survey. The perceived unexpected negative impacts included (particularly during the evening film screenings) increased indiscriminate sexual behavior, sexual molestation, violence, and theft (thieves were said to loot the houses of those attending the films). There also was a perception that walking to and from the films was not safe. The desire of wives and children to attend the films was described as a source of conflict in the family because of these negative perceptions on the part of some husbands and parents, as well as safety concerns. Due to these problems and other related concerns, a significant proportion of the respondents felt that the program implementers were not respectful of local cultures and religions. Important unexpected positive effects also were identified. Specifically, many informants emphasized that the films helped them forget their problems, even if only for a short time, and that this was highly valued.

Results from the survey suggested that the original educational objectives of the program were being met. In current evaluation practice, this would have been the main message from the evaluation, and therefore, little new impact information would have been gained that could be used to evaluate and strengthen the program. The additional data on unexpected impacts enabled the identification of additional ways that the program was valued (based on the unexpected positive impacts) while at the same time suggesting ways in which the program could be improved (by addressing the unexpected negative impacts). The latter included suggestions for better security at the screening sites and on the paths to the screenings, and increased attention to issues of conflict. Other suggestions included further investigation of the image of the implementing organization among the refugees, in order to to try to improve it and thereby improve local cooperation.

https://doi.org/10.1017/S1049023X00005100 Published online by Cambridge University Press

Case Example 2—Northern Uganda (Internally Displaced Persons)
This study was part of a randomized, controlled trial of the impact of two psychosocial interventions for adolescents living in camps for internally displaced persons in Northern Uganda (unpublished data). The populations of these camps consisted of local people who had left their homes because of the ongoing conflict between the Ugandan army and the Lord's Resistance Army, a rebel group. Both interventions were for the relief of depression-like problems.

A sample of the youth who participated in the interventions and some of their caregivers were interviewed using the free listing method described previously. In this case, most of the unexpected impacts identified were positive. These included the adolescents feeling more respected and trusted by caregivers and a reduction in the incidence of eloping or getting married (which is considered undesirable at this age). The caregivers who were interviewed, indicated that the program participants were more respectful and obedient and more willing to do domestic work (e.g., cleaning the house, gathering water). Questions regarding these unexpected impacts were added to the original quantitative instrument previously used to screen participants into the interventions and establish baseline levels of symptoms. This expanded instrument was then used to reassess participants at the end of the program, which confirmed that the impacts identified in the qualitative study were experienced by (and important to) many of the program participants. This information confirmed the utility of the interventions from the local perspective, and helped the researchers to understand the variety of ways in which the program was affecting quality of life.

Discussion

Currently, many humanitarian assistance programs still do not perform even the standard impact evaluation process of pre- and post-quantitative assessment of pre-defined objectives.[2–5,26] However, it is not premature to advocate for expanding the scope of program evaluation. During humanitarian emergencies, there is intense pressure to act quickly. But, for those programs which are able to institute the standard impact assessment, the additional methods described above do not detract from the goal of a speedy response, since they occur after the intervention has been initiated. Given the resources needed for the qualitative study (see Description of Proposed Evaluation Methodology), the authors consider that the additional cost and effort to explore all four categories of impact do not need to be much greater than that needed to assess only the study objectives.

The authors consider that there is an ethical obligation for those who implement aid programs to do all that they can to ensure that these programs do more good than harm, and to constantly strive to maximize benefit and minimize harm. This cannot be achieved when the some types of impacts are not assessed and when assessment is biased to exploring only a limited number of positive, expected impacts. Whatever assessment methods are used, humanitarian workers should strive to gain as full of a picture as possible of the impacts of their programs.

The authors plan to conduct the evaluation process described here with other types of programs. We also are considering additional methods for identifying expected and unexpected impacts. The qualitative approach only explores program participant perceptions of impact. While this type of local input is essential, other perspectives, such as those of donors, field staff, and agency or government workers, also will be useful and must be tested and understood. An exploration of assessment methods other than interviews, such as observations or environmental monitoring also may identify impacts not immediately apparent to those involved. In situations in which many unexpected impacts are identified, it would be helpful to include a process of weighting these impacts from various perspectives to allow program managers to better prioritize the changes needed in future programs.

Finally, although the issues and expanded evaluation approach described here were developed with respect to humanitarian assistance programs, the experiences of the authors suggest that interventions outside of this also can have significant expected and unexpected impacts. The principles and approach presented in this paper may prove relevant to research and assessment in other areas of public health, development, and humanitarian assistance.

References

1. Taylor AJ, Cuny FC: Conference report: The evaluation of humanitarian assistance. *Disasters* 1979;3(1):37–42.
2. Roberts L, Hofmann CA: Assessing the impact of humanitarian assistance in the health sector. *Emerg Themes Epidemiol* 2004;1(1):3.
3. VanRooyen MJ, Hansch S, Curtis D, *et al*: Emerging issues and future needs in humanitarian assistance. *Prehospital Disast Med* 2001;16(4):216–222.
4. Griekspoor A, Sondorp E: Enhancing the quality of humanitarian assistance: Taking stock and future initiatives. *Prehospital Disast Med* 2001;16(4):209–215.
5. Hofmann CA, Roberts, L, Shoham, *et al*: Measuring the impact of humanitarian aid: A review of current practice. London, UK: Overseas Development Institute, Health Policy Group Report, June 2004. Available at http://www.odi.org.uk/hpg/publications_date2004.html. Accessed 11 June 2007.
6. Sphere Project: Humanitarian Charter and Minimum Standards in Disaster Response. Geneva, Switzerland: Sphere Project, 2006. Available at http://www.sphereproject.org/content/view/34/84/lang,English. Accessed 05 June 2007.
7. Bolton P: Local perceptions of the mental health effects of the Rwandan genocide. *J Nerv Ment Dis* 2001;189(4):243–248.
8. Sphere Project: Humanitarian Charter and Minimum Standards in Disaster Response: Appendix 3: Formulas for calculating rates of mortality and morbidity. Geneva, Switzerland: Sphere Project, 2006 Available at http://www.sphereproject.org/content/view/133/0/lang,English/. Accessed 07 June 2007.
9. US Agency for International Development: Standardized Monitoring and Assessment of Relief and Transitions (SMART): Protocols and Methods. Available at http://www.smartindicators.org/protocol.htm. Accessed 07 June 2007.
10. Bolton P, Bass J, Neugebauer R, *et al*: Results of a clinical trial of a group intervention for depression in rural Uganda. *JAMA* 2003;289:3117–3124.
11. Walker P, Leaning J, Minear L: Smoke and mirrors: Deficiencies in disaster funding. *BMJ* 2005;330:247–250.
12. Salama P, Spiegel P, Talley L, *et al*: Lessons learned from complex emergencies over the past decade. *Lancet* 2004;364:1801–1813.
13. Athukorala P, Resosudarmo BP: The Indian Ocean tsunami: Economic impact, disaster management and lessons. Asian Economic Papers (in press). Available at http://rspas.anu.edu.au/economics/publish/papers/wp2005/wp-econ-2005-05.pdf. Accessed 05 May 2007.
14. Baker J: *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington, DC: World Bank, 2000.

https://doi.org/10.1017/S1049023X00005100 Published online by Cambridge University Press

15. Shippee J: Appendix D: U.S. Organizational Experience with Volunteer Health Programs. In: Mullan F, Panosian C, Cuff P (eds.): *Healers Abroad: Americans Responding to the Human Resource Crisis in HIV/AIDS.* Washington, DC: National Academies Press, 2005.
16. Ogunmefun C, Schatz E: Caregivers' sacrifices: The opportunity costs of adult morbidity and mortality on female pensioners in rural South Africa. Available at http://www.colorado.edu/ibs/pubs/pop/pop2006-0011.pdf. Accessed 11 June 2007.
17. Russell LB: Opportunity costs in modern medicine. *Health Affairs* 1992;11(2):162–169.
18. White H: Challenges in Evaluating Development Effectiveness. Washington DC: World Bank, 2003. Available at http://www.worldbank.org/oed/conference2003/agenda.htm. Accessed 11 June 2007.
19. Abirafeh L: Lessons from gender-focused international aid in post-conflict Afghanistan...learned? Bonn, Germany: Friedrich-Ebert-Stiftung, 2005. Available at http://fesportal.fes.de/pls/portal30/docs/FOLDER/WORLDWIDE/ASIEN/BERICHTE/BERICHTE2005/AFGHANISTAN0905ABIRAFEHGENDER.PDF. Accessed 11 June 2007.
20. Thomas A: Linking preparedness and performance: The tsunami experience. Humanitarian Exchange 2005;32. Available at http://www.odihpn.org/documents/humanitarianexchange032.pdf#page=38. Accessed 01 June 2007.
21. Bolton P: Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *J Nerv Ment Dis* 2001;189(4):238–242.
22. Murray LK, Haworth A, Semrau K, *et al*: Violence and abuse among HIV-infected women and their children in Zambia: A qualitative study. *J Nerv Ment Dis* 2006;194(8):608–615.
23. Weiss W, Bolton P: *Training in Qualitative Research Methods for PVOs and NGOs: A Trainer's Guide to Strengthen Program Planning and Evaluation.* Baltimore, Maryland: Center for Refugee and Disaster Response, Johns Hopkins Bloomberg School of Public Health, 2000.
24. Bolton P, Neugebauer R, Ndogoni L: Prevalence of depression in rural Rwanda based on symptom and functional criteria. *J Nerv Ment Dis* 2002;190(9):631–637.
25. Bolton P, Tang A: An alternative approach to cross-cultural function assessment. *Soc Psychiatry Psychiatr Epidemiol* 2002;37(11):537–543.
26. Sundnes KO, Birnbaum ML (eds): Health disaster management: Guidelines for evaluation and research in the Utstein style. *Prehospital Disast Med* 2003;17(4,suppl 3):s31–s143.

https://doi.org/10.1017/S1049023X00005100 Published online by Cambridge University Press