LINKING COMPLEX EMERGENCY RESPONSE AND TRANSITION INITIATIVE

CERTI Crisis and Transition Tool Kit

Cross-Cultural Assessment Of Trauma-Related Mental Illness

Research Report Summary

Paul Bolton MD MPH ScM
The Johns Hopkins University

Lincoln Ndogoni
World Vision International

August 2000

This project was made possible through Cooperative Agreement Number HRN-A-00-96-9006 between the US Agency for International Development and Tulane University. This publication was made possible through support provided by The Africa Bureau's Office of Sustainable Development, United States Agency for International Development; in support of USAID Strategic Objective 22. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of USAID.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

1.1 Project Objectives

1. To create an instrument adaptation and validation process which can be used by non-governmental organizations (NGOs) and others to quantitatively assess the mental health burden of trauma at the population level across cultures and situations.

2. To use this instrument and process to assess part of the mental health burden of trauma on a civilian population in Rwanda.

3. To use the resulting data to assess the need for interventions, form the baseline for an intervention process, and (at a future date) to plan the form of such an intervention.

4. Current methods to assess mental health across cultures require resources and time not available to NGOs and many of the populations they serve, and are therefore research tools only. In this study we have attempted to develop a method useful for NGOs because it requires only training and existing resources.

1.2 Methods

1.2.1 Overview

The method we tested in Rwanda is designed for use with various clinical mental health indicators. For this first trial we chose to study Depression as an indicator of the effects of mental trauma, and to study only adults. We assessed Depression using a version of the well-known Hopkins Symptom Checklist (HSCL), which has been used among many populations. We chose Depression because there is agreement on its basic form, it represents the most severely affected persons, and because it has been studied in a variety of situations and cultures. We chose it over Posttraumatic Stress Disorder (the other major clinical indicator of the severe effects of trauma) because it occurs in situations other than trauma, and so our findings would have implications beyond populations affected by trauma. Controversy continues about the use of these and other Western clinical mental health indicators in non-Western cultures. This is partly due to the failure of workers using these indicators to adequately validate them prior to use, thereby leaving the question of their appropriateness unanswered. Therefore, a vital component of this method is preliminary investigation of the local validity of these indicators.

The method we developed has 8 main stages:

1. Collecting ethnographic data on local perceptions of mental health.
2. Analyzing these data for evidence that Western indicators of mental problems are appropriate.
3. If so, using these data to adapt existing questionnaires that measure these indicators.
4. Testing the validity of these questionnaires and these indicators.

5. Using the resulting instrument in a community-based survey of a random population sample.
6. Analyzing the survey data to
    (a) assess the local prevalence and characteristics of the selected mental health problems;
    (b) further test the validity and reliability of the instruments and indicators.
7. Using the ethnographic and survey data to design appropriate interventions.
8. Repeating the survey after the intervention to measure impact.

This report describes the first 6 stages in detail, and our experience in fielding testing these stages among a population in Rwanda.

1.2.2 Ethnographic Study

The purpose of the ethnographic study was to determine if local people experienced Depression as a result of trauma, and to learn the names and symptoms of comparable 'Depression-like' illnesses. If local people experience the symptoms of Depression as a result of trauma, this would be evidence that Depression occurs among this population. Knowledge about local perceptions of Depression would enable us to work with local people to adapt and validate existing Depression instruments (in this case the HSCL). Data gathered in the ethnographic study is also useful in designing interventions. We trained PSSP staff in three ethnographic methods - free listing, key informant interviewing and pile sorts - which they then used to conduct the study.

PSSP staff used free listing to develop a list of problems resulting from the genocide in 1994, according to local informants. This list contained the local names and short descriptions on significant mental health issues - as perceived by local people - and the importance of these issues relative to other problems. From this list we identified Depression-like illnesses for detailed investigation using key informant interviews. These are longer interviews with local people that provide more information than the shorter free listing. The results of these interviews enabled us to determine whether Depression occurred in this population, and how it is understood.

We anticipated that respondents might link classical Depression symptoms with other symptoms. We used pile sorts to determine if these symptoms might comprise important local variations in Depression symptomatology, or were merely individual opinions. The pile sort results enabled us to determine which symptoms were potentially important and should be added to the HSCL.

Interviewers also conducted a free list exercise asking respondents to list the important tasks that local people do regularly. We used the results to create sets of questions (separate for men and women) on functional disability. These questions were added to the HSCL to explore the relationship with Depression, which is known to be associated with reduced function in other populations.

1.2.3 Translation

After adaptation using the ethnographic results, we had an assessment instrument consisting of basic demographic data, the adapted HSCL, and a community-specific questionnaire on function. The instrument was translated from English to Kinyarwanda using a combination of

group and translation-backtranslation methods. During the translation we constantly cross-checked the results with the ethnographic data. This was to ensure that the translation used terms which were truly part of the local vocabulary, since word usage frequently varies across regions.

1.2.4 Validity Study

Prior to the main survey we conducted a mini-survey to study the validity of the instrument. We reasoned that the local syndrome identified in the ethnographic study as most similar to Depression should be highly correlated with Depression itself if Depression occurs in this population and is accurately diagnosed by the instrument, and if the local syndrome is accurately diagnosed by local people. We asked key informants in the community to identify local people who have and who don't have the local syndrome. Blinded interviewers then interviewed these people using the instrument, and also asked if the respondents felt they had the syndrome. We compared diagnoses of the local syndrome by key informants and the respondents themselves with Depression diagnoses using the instrument.

1.2.5 Survey and Analysis

The instrument was then used in a random survey of the adult population. We selected 5 'Sectors' from each of the rural Communes of Kinzenze and Butamwa, based on geographical spread. Within these Sectors we then selected a simple random sample for interview. *Therefore, the survey results do not refer to the Communes overall, but only to the selected Sectors.*

Interviewers came from Kigali because local people with a high school education who could read and write were not available. PSSP staff who had already received training and data gathering experience in the ethnographic study acted as supervisors. As well as assisting the interviewers they checked on all refusals and re-interviewed 10% of respondents. This provided a quality control measure, and data to test the reliability of the instrument over time (test-retest reliability).

We used the data to conduct further analyses of the validity of the Depression concept and the instrument, and to test the instrument reliability. We calculated the prevalence of Depression among the population as an indicator of need for specific interventions. We explored the relationships between ability to perform specific tasks and Depression and other respondent characteristics, using correlation analyses. Using a summary scale of functional disability we also explored the relationship between Depression and scores on this scale, using logistic regression. Finally, we developed a Depression scale and calculated a cut-off score that can be used for diagnosing Depression (instead of using the DSM criteria) and is most sensitive and specific for detecting associated functional disability.

1.3 Results

The ethnographic study showed that local people experience all the DSM[1] diagnostic symptoms of Depression as a result of the 1994 genocide. This supports the occurrence of Depression among this population. They also experience other symptoms associated with Depression,

---

1 Diagnostic and Statistical Manual of the American Psychiatric Association, 4th edition 1994.

three of which were confirmed by the pile sort exercise and incorporated into the Depression questionnaire. Local people do not organize symptoms into an entity similar to Depression that we could use for direct comparison. The closest syndrome to Depression is *Agahinda gakabije*, or severe grief. In the validity study we therefore compared local diagnosis of this syndrome (by informants and respondents) with Depression diagnosis by the instrument. We found that people with Depression form a sub-group of those with severe grief. This is consistent with experience elsewhere that grief triggers Depression in a proportion of cases (Craig, 1996). It provides some additional evidence that Depression occurs among this population and can be diagnosed by the instrument.

We used the questionnaire to survey 368 and 72 adults in selected Sectors of Kinzenze and Butamwa Communes respectively. 66% of respondents were female in Kinzenze and 79% in Butamwa (other sources have found that Rwanda is now 70% female (Dabelstein, 1996)). In the Kinzenze Sectors 17.9% met the DSM criteria for Depression (see Appendix E for criteria) and 41.8% described themselves as having severe grief. In the Butamwa Sectors the corresponding figures were 5.6% and 31.9%. Studies in other parts of Africa and the world have found prevalences of Depression between 0.8-5.8% (Weissman, et al, 1996; Bhagwanjee et al, 1998). This suggests that the Butamwa data may be close to the background level of Depression in this population, whereas the higher rate in the Kinzenze Sectors may reflect the more severe genocide experience of that Commune.

Internal consistency reliability (agreement between similar questions) was very good for all the Depression questions (Cronbach's alpha =0.87), both the original HSCL questions and the local symptoms identified in the pile sorts. This supported their significance as part of the local expression of Depression. The internal consistency of the male and female function questions was also good (Cronbach's alpha = 0.82 and 0.82 respectively).

Overall, Depression is strongly associated with reduced function. Grief is not associated with reduced function when the effects of Depression are accounted for. When we studied these relationships between Depression, grief and function among various subgroups we found that the association between Depression and function is greatest in those with more education and who are older. These associations are not significantly different for men and women.

Among the non-depressed both sexes most frequently experience difficulties with tasks requiring more energy. This may reflect physical ailments and under-nutrition (during the study there was a drought and food shortage). Depression is particularly associated with increased difficulties in those tasks necessary for the family's well-being: labor and earning among men; and washing clothes, cleaning house, and caring for children among women. For each of these tasks most of the cases among the depressed group are associated with the Depression itself, and Depression accounts for 17-48% of all significant difficulty in the population.

We created a simple score of Depression for each respondent by adding their responses to the Depression questions. We used the validity and reliability results to decide which Depression questions were useful and should be included in the scale, and which should not be used. In the same way we also created a function score for each respondent. Since both Depression and function questions use a Likert scale of responses with higher numbers representing increasing severity, in both scales higher scores represent increasing Depression or dysfunction. We measured the reliability of these scales over time by comparing scores on the first and second interviews for 37 respondents who were re-interviewed. The Pearson correlations between the first and second interviews were 0.671 for the Depression scale and

0.574 for the function questions, which is adequate. Using ROC analysis we calculated a cut-off score of 30.5 on the Depression scale. At this score Sensitivity = 68.3% and Specificity = 71.8% for detecting significant functional disability and 98.5% and 79.1%, respectively for detecting Depression based on the DSM criteria. This scale and cut-off score can be used in future to assess Depression and its effects on function among this population.

1.4 Discussion

This study was the first field trial of a method to assess the burden of mental problems across cultures. Existing methods require time and resources beyond the means of most NGOs, and in many cases the necessary gold standards are just not available. *Our approach, was to understand how local people view mental health so that we could enlist their assistance in answering these questions.*

For this first trial we worked in Rwanda because of our interest in transitional populations, and because World Vision has a psychosocial program there. For transitional populations affected by war (and genocide) the two most severe mental health problems are Depression and PTSD, which commonly occur together (Engdahl et al, 1998. Shalev et al, 1998. Peltzer, 1998). Resources prevented us from investigating more than one disorder so we focused on Depression, although we could have investigated PTSD or any mental illness or health issue using this method. We chose Depression because - unlike PTSD - it is also common in countries without experience of war (Weissman et al, 1996). Therefore our results could have relevance beyond the scope of transitional populations.

The field trial proceeded smoothly and we believe demonstrated that it is possible to train local staff and conduct a rapid ethnographic and quantitative study in a very short time with resources currently available to many NGOs and other organizations - the only additional requirements are training and a commitment to understanding local communities. Studies like this one are best done at the beginning of programs, even before the interventions have been decided: Ethnographic methods provide a lot of general information about these communities which can be used to plan a program, as well as an effective way to meet local people and build trust. For example, the first free listing exercise provides information on all the community's problems that can guide an NGO in setting up all programs, not just programs for mental health.

We trained local staff in instrument preparation, data collection and qualitative data analysis. We did not train them in quantitative data analysis and interpretation, because of time and resource limitations. In future, this training can be provided to selected staff with computer skills. Eventually the quantitative analysis section of this method will consist of the following:

- Calculation of correlations between Depression and local illness
- Generating and interpreting Cronbach's alpha scores and item analysis
- Calculation of prevalences
- Creation of simple function and Depression scales
- Calculating cut-off scores on the Depression scale most appropriate for detecting functional disability.

Our results suggest that Depression occurs among this population even though it is not recognized locally as a distinct syndrome. The validity and internal reliability of the instrument were also good although the test-retest reliability were adequate only. This combination suggests that moods may vary between tests, or that there is reactivity to the survey which is

expressed in the second interview. However, this was not severe enough to invalidate using the instrument.

Depression rates were much lower than grief, and there was a large difference between the two areas: 17.9% of adults depressed in the Kinzenze Sectors and 5.6% in Butamwa. A study of Depression in 10 countries (not including Africa) found prevalences ranging between 0.8-5.8% (Weissman, et al, 1996), which was consistent with results from other parts of Africa (Bhagwanjee et al, 1998). This suggests that the Butamwa data may be close to the background level of Depression in this population, whereas the higher rate in Kinzenze may reflect the more severe genocide experience in that Commune.

Even people with severe grief, in the absence of depression do not suffer significantly reduced function. The more general implication may be that even severe levels of distress in those without mental illness do not greatly affect function. To test this hypothesis these methods should be used in similar research among other populations, and to examine other mental illnesses (such as PTSD). If confirmed, it would support a shift from current approaches which broadly address trauma experience and grief, to focusing on those who have developed mental illness as a result

Among both men and women depression is associated with dysfunction in tasks important to the care of the family, including children. Among those depressed most of the difficulty with these tasks was associated with the depression itself, and with a significant proportion of the difficulty reported by the entire sample. This is remarkable, considering the many other problems and diseases that cause dysfunction in this part of Africa. It also carries implications for social and economic development, since the earning power and well-being of the family are compromised. If Depression is a direct cause, then Depression in an adult will significantly affect the whole family, and treatment will be required before the family can fully take advantage of available social and economic resources to improve their situation.

Since this was a cross-sectional study we could not prove that Depression causes dysfunction. It is possible that primary dysfunction could produce Depression, or that there are mixed effects. Research is required to sort this out. The most efficient way to do this would be a study of effective interventions specifically for depression. Concomitant improvement in depression and function with depression treatments would then demonstrate the link. This type of research would also demonstrate whether function can be improved by treating depression, and could be used to test new interventions appropriate for areas with limited resources. If Depression is shown to be the cause of dysfunction, to be treatable, and that function improves as a result, then the prevalence of Depression becomes a developmental issue and should be assessed whenever high rates are suspected, such as after a war or disaster. We consider that the measurement method we have described here provides a means of conducting this type of research.

During the survey interviewers noted that those subsequently diagnosed as depressed were obviously distressed, and very keen to talk about their experiences. Interviewers reported that these respondents were grateful for the opportunity to speak about their problems, and many said that this was the first time they had discussed them openly with anyone. PSSP staff also noted that these persons were not picked up by their existing outreach program, despite the program's good coverage. The same was true of persons diagnosed with Depression during the validity study. This may be due to the reclusive behavior of the depressed, as expressed by the respondents and noted by the interviewers. Whatever the cause, it suggests that current

programs are not reaching those most in need of assistance. These are the people whose mental distress is the most severe, and whose reduced function is most likely to affect their own well-being and that of the community.

1.5 Recommendations

For further assessment

1. World Vision and Johns Hopkins University should repeat this assessment procedure with other populations; to assess need and build up a composite picture of mental illness across Africa, and to continue to refine and simplify the method. Future changes should include assessing exposure to events and their relationship to mental illness, assessing personal and community factors which may mediate reactions to trauma, and including in the ethnographic study investigating community suggestions for addressing these issues. Improvements should also include assessments for other mental problems (our study found good evidence for frequent occurrence of PTSD), and for children. Future field trials should include training in quantitative analysis.

2. Current programs do not appear to reach those most in need of mental health assistance. World Vision should form a technical advisory group composed of World Vision staff and experts in the field of mental health issues. This group should review these results and consider how WV and other NGOs can best help persons with Depression, given existing NGO resources. Communities should contribute to this process, and particularly to the review of proposals for feasibility, acceptability and implementation.

3. World Vision Rwanda should use the recommendations in 2. to assist those identified in the survey as depressed, and screen for others with Depression. The survey should also be repeated in other areas in which WV Rwanda suspects a need and is planning a mental health intervention, to assess the level of need.

4. Finally, we must emphasize that the free listing and validity study both revealed that poverty and lack of people are more pressing problems for most people than mental and emotional issues. One of the positive features of this assessment method is that it should enable the NGO to put mental health in the context of other problems. WV Rwanda should use information to prioritize all their interventions for Kinzenze and Butamwa Communes, and repeat the free listing in other areas prior to planning interventions.

For new research

1. Likely interventions resulting from the technical advisory group (and other sources) should be tested using standard research protocols. Such protocols could be designed and executed which could investigate three outstanding issues:
    - To find effective treatment for depression, given limited resources and large group of affected persons.
    - The nature of the cause-effect relationship between depression (and other mental illness) and function.
    - To determine if improvement in depression results in improvement in function and how much.

2. To repeat (or expand) these protocols to other mental illnesses associated with significant dysfunction.

2. STATEMENT OF SUPPORT

This research was funded in part by the US Agency for International Development (USAID) under Cooperative Agreement HRN-A-00-96-9006 with The Johns Hopkins University (JHU). These funds support research and consultancy costs. World Vision also provided financial support, as well as staff and logistic support.

3. ACKNOWLEDGMENTS

We would like to acknowledge the following organizations and individuals for their contributions to the research: Bill Lyerly of USAID and Tom Ventimiglia of World Vision, without whose support this project would not have been possible, the staff of the World Vision Rwanda Psychosocial Support Program and our interviewers and interviewees. Also Bill Eaton and Bill Weiss of Johns Hopkins University, and Nancy Mock of Tulane University, all of whom gave invaluable advice on the design and execution of this project.

4. INTRODUCTION

4.1 Background to Johns Hopkins University-World Vision Collaboration

This project is part of the Complex Emergency Response and Transition Initiative (CERTI). This is an interagency initiative, partly funded by USAID. The objectives of CERTI are to develop a consensus on best practices during and after a complex humanitarian emergency (CHE), and to strengthen the capacity of organizations providing assistance. The focus is on Sub-Saharan Africa.

Johns Hopkins University (JHU) received funding through CERTI to improve the capacity of NGOs to assist transitional populations. These are post-disaster populations currently between the emergency phase and resumption of community development. JHU faculty chose to assist NGOs to enhance their ability to do formative and operational research, both in-house and in cooperation with academic institutions like JHU. The reasons for boosting research capacity are:

- There is a lack of information about the health of populations in transition, and therefore no firm basis on which to design programs and assess program effectiveness.
- NGOs work at the community level, where standardized interventions require local adaptation. To do this for each new situation requires a field research capacity. To not do it risks a waste of resources on an unsuccessful project.
- Many NGOs now employ technically qualified persons capable of acting as the focal points for enhanced research capability.

In recent years the mental health effects of CHEs have gained significant attention. NGOs and other humanitarian assistance organizations have begun programs to reduce these effects. However, little exists in the way of science-based methods to assess the need for, and impact of, these interventions that does not require time and resources beyond NGO capacities. Therefore, program impacts are assumed rather than measured, and aid agencies can only report outputs and anecdotal improvements.

As part of their CERTI activities, JHU faculty developed an approach to assessing the mental health impact of traumatic events, and the impact of interventions. JHU and World Vision agreed to field test and further develop this method in Rwanda. We chose Rwanda because of the continuing impact of the 1994 genocide and war, and because WV currently has a local psycho-social program.

We conducted the field trial in selected areas of Kinzenze and Butamwa Communes. These Communes are rural administrative areas near Kigali, each containing 100,000-160,000 people. We chose these Communes because World Vision is planning psychosocial activities there, so the study can help assess the local need, and provide baseline figures. Kinzenze is south of Kigali. It is relatively dry and flat. In the past sleeping sickness was endemic there and the land was not highly valued. Kinzenze has traditionally had a large number of Tutsi. During the genocide many were killed or fled and their homes destroyed. Now the region is an irregular patchwork of mud houses on small plots, although in some areas the government has moved people into newly built villages with streets and houses set out on grids. This process has been

referred to as 'villagization' and is not popular. Butamwa has richer soil and is traditionally Hutu. The region is hilly, and many houses are difficult to reach except by foot. Villagization has also occured in some areas of Butamwa.

4.2 Background to WV's Interventions in Rwanda

In 1994 Rwanda endured 100 days of genocide during which the ethnic majority Hutus - urged on and organized by the Hutu-dominated government - targeted the Tutsi minority and moderate Hutus. By the time the government was overthrown by an invading Tutsi-dominated army between eight hundred thousand and one million people had been killed. In 1996 WV set up the Psychosocial Support Program (PSSP) to address the perceived psychological aftermath of the genocide. The PSSP has two interventions:

- Community-based mental health support
- Personal Development Workshops.

The community-based intervention aims to improve the local capacity of the people in the Kigali Rural Prefecture to deal with mental health problems resulting from trauma. This is done by training local people - teachers, community health workers and special community trauma monitors and counselors - to recognize, assist and refer persons with these problems. The personal development workshops assist individuals to understand and deal with their reactions to the genocide. Participants discuss bereavement, emotional effects, healing and forgiveness.

A recent project evaluation praised the approach and methods of the PSSP[2], and reactions of the staff and participants have been very positive. However, the WV staff have recognized that they lack a means to scientifically assess the program's impact. In both progress reports and in the formal evaluation, staff and the evaluators were limited to assessing outputs and relying on anecdotal reports. WV staff were therefore interested in developing methods of objective measurement of need and impact.

4.3 Background and Overview of Project Design

Progress in improving mental health among trauma affected populations depends on accurate measurement of the problems being addressed. Without quantitative measurement of mental health indicators there is no basis for allocating resources, for determining whether an intervention has been successful, or for choosing between multiple competing interventions. Most methods to assess the burden of mental health problems were developed in western countries. They are based on concepts of mental illness developed among western populations and defined in the widely used Diagnostic and Statistical Manual (DSM-IV) of the American Psychiatric Association. There is evidence that at least some of these concepts (including Depression) are valid across many cultures, and therefore instruments based on them have been adapted for use among different groups. Nevertheless, in each new culture the validity and reliability of these concepts and their instruments must be tested anew: validity in some cultures does not prove validity in every culture.

---

2 Robertson M, and Hajiyiannis H. Evaluation of the Psycho-social Support Programme, World Vision Rwanda. October 1999.

Existing validation methods require highly trained health workers very familiar with the local culture and language. These workers diagnose persons using alternative methods (usually psychiatric examination), and validation consists of comparing these diagnoses with those made using the instrument. These workers are also key to adapting and translating the instrument, using their knowledge of the local language and local perceptions of mental illness. They are expensive and in many areas are unavailable.

Faculty at JHU developed an alternative method to adapt and validate instruments that would not require these highly trained workers. Our aim is to produce a method appropriate for use by (NGOs) and the populations they work with. Following theoretical development, JHU faculty approached World Vision to test and further develop this method by field trial. World Vision Rwanda agreed to collaborate and field work was conducted in October-December of 1999 among a Rwandan population.

This report describes the method we developed, our experiences with its use in Rwanda, and the data we collected using this process. We make recommendations for next steps by JHU, World Vision and/or others to continue to develop this process. We also make recommendations based on the Rwanda data. We intend that when fully developed this method could be used by World Vision and other NGOs in other parts of Africa and beyond. The process is ideally suited to use by NGOs, since it focuses on community participation in the entire process of design, execution and analysis. It would be used where there is suspicion that mental health issues were significant and/or interventions are planned. By empowering communities and NGOs to assess these needs and the effect of interventions, it enables these same groups to target needs, leverage donor support for interventions and to test their effectiveness by means of repeated measurements before and after implementation.

5. PROJECT OBJECTIVES

1. To create an instrument adaptation and validation process which can be used by NGOs and others to quantitatively assess the mental health burden of trauma at the population level across cultures and situations.

2. To use this instrument and process to assess part of the mental health burden of trauma on a civilian population in Rwanda.

3. To use the resulting data to assess the need for interventions, form the baseline for an intervention process, and (at a future date) to plan the form of such an intervention.

6. METHODS

6.1 Overview

6.1.1 Stages

The method we propose has 8 main stages:

tures since its development in the 1950s.

The HSCL was developed before the modern criteria for Depression were published in the DSM. It lacks a question on psychomotor agitation (one of the diagnostic criteria). We added a question, giving an initial assessment instrument with 16 questions. In the study we used a previously tested algorithm based on DSM criteria to diagnose Depression with these questions (Mollica et a, 1999) (see Appendix F). The HSCL also contains a question referring to 'feeling trapped', that does not appear to fit any of the DSM criteria. This question was retained for the survey, although it was not used in the analysis.

Part of the burden of mental health problems includes their impact on function. Like Depression, most instruments measuring function were created in developed countries and are not appropriate to other situations. Moreover, many questions focus on ability to execute component activities (such as walking) rather than on the ability to complete specific tasks important to the life of the person and those around them. Since the important tasks will vary in different situations, we developed an approach to creating a functional disability assessment instrument for each community. We created a template of the form shown in the English version of the questionnaire (Appendix A) in which the actual tasks in the table would vary with each community. Which tasks to include depends on a community-based free list exercise; literally asking each new community (see below). The free listing asks for tasks under the categories of self-care, caring for family, and community duties, to elicit tasks covering the major areas of human responsibility. The additional category of 'other' was added to the template to allow individual respondents to include tasks important to them but not found in the free listing exercise.

The template includes 5 response categories (to facilitate data analysis). This may be a lot for a person to remember, so we created a non-verbal response card. This provided a visual metaphor for each of the categories in terms of a person carrying an increasingly heavy load (for higher levels of dysfunction). Beneath each picture was written the category of response it represented, for literate respondents. Since functional ability varies with age, sex and the presence of other problems, persons were asked to assess their ability to function compared with others their age and sex who are problem-free.

6.2 Ethnographic Mini-Study

6.2.1 Overview

The purposes of this phase are to provide data to:

1. Test independently (of the survey) the validity of applying western syndromes of mental illness to the local situation.

2. To provide data to adapt and translate existing instruments measuring those syndromes.

3. To validate the instruments.

4. To provide information (in the long term) for designing appropriate interventions.

This section describes how free listing, key informant interviewing, and pile sorting were used in this study to achieve these purposes. More detailed descriptions of these methods can be found in Appendix B.

The first activity was a short training in the principles of research - particularly ethnographic research - for the five PSSP staff who were to be the interviewers. Emphasis was on the importance of not leading respondents or reacting to answers - skills that are key to the success of these methods. When done correctly responses reflect only the respondent's thinking because he/she has no idea what the interviewer might expect or want to hear. In this way ethnographic data can provide independent evidence of validity if responses agree with existing concepts.

6.2.2 Free Listing

Training in general principles was followed by a short training in Free Listing, the first data gathering method. The purpose of this activity was to learn about the more important local mental health issues - local terms and descriptions - as well as how highly these problems were ranked by the community. Based on their prior experience working in the community, PSSP staff created a list of local persons knowledgeable about the community's problems. Staff then went to the communities and interviewed these people using the free listing method. Respondents were asked to provide lists in answer to the following question:

'What are the main problems that affect the people of this community as a result of the genocide in 1994?'

Each problem was recorded using the words of the respondent, who was then asked to provide a short description. For problems that sounded mental or emotional, respondents were also asked who people consulted about these problems; their role, name and contact information.

The resulting data provided a composite list of the major mental and emotional problems from the community's perspective, as well as how highly the community prioritizes them compared with other issues.

After completing this free list exercise a second was conducted. The second exercise was not directed at particular informants, but rather a convenience sample of adults from across the study area. The purpose was to learn about the tasks most important to people, for use in the function questionnaire. Informants were asked to provide three lists of the major tasks and duties a person must perform regularly to care for themselves, their families and their community. Men were asked to describe the tasks specific for men, and women to describe the tasks for women. As before, responses were recorded using the language of the informants and a short description of each task was obtained. At the end of the exercise each respondent was asked for contact information about people who were consulted for the mental and emotional problems mentioned in the first free list exercise. This was combined with the same data from the first free list to give a list of key informants for the following activity.

6.2.3 Key Informant Interviews

The mental and emotional problems identified in the first free list were the basis for the next exercise. This was key informant interviews conducted with the persons identified by the community as dealing with these problems. The purpose of the key informant interviews was to confirm the local words for important Depression-like illnesses and symptoms that emerged from the first free lists (since we were investigating the validity of the Depression concept), and to develop more detailed descriptions of them. We also wanted to determine if there were any other important terms possibly related to Depression that were missed in the first exercise, and to obtain descriptions of them as well.

Interviewers began the first interview by describing a hypothetical person with many of the Depression-like symptoms named in the free lists. This description included only symptoms mentioned in the free lists and included in the DSM criteria. Interviewers then asked the informant to name and describe all the problems that these symptoms might represent. At the second interview with the same informants (the next day) interviewers repeated the same question, to gather any additional thoughts the person may have had since the first interview. Plans for further interviews were dropped because of the limited amount of additional information obtained from these repeat interviews.

These data were reviewed for evidence that Depression occurs in this population. We were primarily interested in how many of the DSM symptoms of Depression were also mentioned by the respondents as the results of trauma. Mention of most of the diagnostic symptoms would be independent evidence that Depression occurs among this population, even if they are not classified into a Depression-like syndrome.[3] However, understanding how these symptoms are classified was important for the next activity.

6.2.4 Pile Sorts

We anticipated that local people would describe symptoms related to Depression which are not part of the DSM definition. These could be important local symptoms of these disorders, or just the opinion of individuals. To test their significance in regard to Depression, we conducted a pile sort exercise: Each of the local symptoms was recorded on a separate card, as were several of the major Depression symptoms listed in the DSM and also described by local people. A convenience sample from across the community was then asked to create piles of cards based on which symptoms 'go together.' This information was summarized in a table that shows the number of times each item was paired with each other item. Reviewing this table demonstrated which local symptoms were most often linked with Depression symptoms, and were therefore potentially part of a local Depression-like syndrome. This marked the end of the ethnographic mini-study.

6.3 Finalizing the English version of the Instrument

In the pile sort we identified potential local symptoms of Depression. We inserted questions on each of these symptoms into the HSCL, using the same response format. This completed the Depression assessment section of the instrument.

---

[3] In other situations we have used this approach to confirm the presence of physical illnesses. For example in Angola we found that people describe the symptoms of malaria and pneumonia in areas where these occur, yet do not combine these symptoms into these diseases. The presence of the symptoms is stronger evidence for the occurrence of a disease than whether local people have arrived at the same classification that we use.

To complete the functional disability assessment section the results of the second free list were reviewed. The most frequent responses in each category (self, family and community tasks) were studied. Tasks which did not affect other people or were not really tasks were removed. For example, many respondents listed prayer as an important task in caring for themselves. However this was removed because it was not clear that inability to do this would affect others. Many men listed 'sending children to school' as an important task to care for the family. But their descriptions made it clear that their task in this regard was to earn enough money to pay the fees. It was removed since it was covered under the 'earning money' task. The remaining tasks were then inserted into the function questionnaire.

For each task we included a question on the causes of disability (see Appendix A). Interviewers asked this question if a respondent reported significant disability (moderate amount or more of increased difficulty) and all health-related causes were recorded. At the end of the questionnaire we included questions on whether the respondent thought they had the local syndrome we had identified as most similar to Depression, and whether other people had told them that they had it. This was done to assess validity (see explanation below). We included questions on the duration of the local syndrome and on alcohol use - to assess its potential impact on function.

6.4 Translation

Four translators were hired in Kigali because none were available in the Communes. Three worked together to translate the English version of the instrument into Kinyarwanda. This was then back-translated by the fourth translator working in isolation. All four translators then met together with the project director to reconcile the differences between the translation and back translation. The results of the ethnographic study (the language used by the respondents) were consulted to help ensure that words chosen were those known and used by the local population. On this basis we rejected several words because although the translators insisted that these words were used by all Rwandans they did not emerge during the ethnographic study.

The resulting Kinyarwanda version of the instrument was then reviewed by the PSSP staff, and the language further simplified. Finally, the instrument was reviewed by 17 Rwandans hired as interviewers. At this stage it emerged that Rwandans living outside Rwanda between 1959-94 had developed some different word usage to those who remained. These displaced persons included some of the interviewers who pointed out words likely to be misunderstood by new arrivals. All the interviewers and PSSP staff then adjusted the questionnaire to ensure the same meaning regardless of the travel history of the respondent. Once this was completed the instrument was considered ready for pilot testing. To summarize: at each review the emphasis was on using simple language intelligible to all persons and consistent with the language from the ethnographic study.

We also developed and translated a standard consent form, using the same procedure (see Appendix C).

6.5 Instrument Testing

6.5.1 Pilot Study

The purpose of the pilot study was to detect any problems with the interview procedure, the consent form and the instrument (including data entry), and to give the interviewers practice

before the survey began. Each interviewer interviewed a man and woman from areas adjacent to the survey area; a convenience sample of 34 respondents. Interviewers, supervisors and the project directors then met and reviewed the experience. The process went smoothly and few problems emerged. Most significant were the listing of causes of disability as poverty and lack of assistance, and ignoring health problems. Although valid concerns, our focus was to assess disability due to health problems, so changes in the function questions were required. We were reluctant to specify health causes only in the questions because of concerns that this might cause respondents to omit some health causes, due to lack of understanding of what we meant by 'health'. Therefore we asked respondents to assess their function and the causes of any disability while assuming adequate help and no poverty. This worked well in the main survey with most respondents focusing on health issues, including mental problems.

Otherwise no significant changes were required to the questionnaire. Both interviewers and participants found the non-verbal card to be particularly useful.

6.5.2 Validity Study

The validity study was a mini-survey conducted in the study area as a test of criterion validity.[4] Supervisors revisited the knowledgeable persons contacted in the first free list and asked for the names of people who had the local syndrome most similar to Depression. They also asked for the names of persons who do not have this problem. Supervisors assigned respondents to interviewers without revealing the respondent's illness status.

At the end of the interview respondents were asked if they felt they had the local illness. We then threw out all responses where the key informant and the respondent disagreed on whether they had the illness. This was to remove the influence of individual judgement; to ensure that we had true cases and true non-cases. The remaining responses were then compared with Depression diagnoses using the HSCL, to determine whether there was a correlation (see Analysis below).

6.6 Sampling and Survey Procedure

The sampling method used in this survey was a simple random sample of adults in selected Sectors of two Communes. Using simple random sampling is unusual in Africa. It was possible here because of the excellent information available on the distribution of the population within the Sectors (which facilitated a highly accurate delineation of the sampling frame) and an unexpected delay in the survey that provided the additional time needed to create the frame. Given these factors, and the sparse distribution of the population, a simple random survey was feasible. In other situations the more usual 30 cluster method would have been used.

Each house in Rwanda belongs to a *Nyumbakumi* which is the lowest administrative level. *Nyumabkumis* consist of 10-30 houses with the head of one household in charge of the *Nyumbakumi*. From 2-18 or more *Nyumbakumis* form a Cellule and from 3-14 or more Cellules form a Sector. Many Sectors form a Commune: Kinzenze has 14 Sectors and Butamwa has 8. From these sectors we chose 5 from each Commune, based on geographical spread and choosing the larger sectors. Within each chosen sector we visited Cellule and Sector offices to

4 The study also provided a test of the changes recommended after the pilot study, as well as further training for the interviewers.

record the site, name and number of houses of each *Nyumbakumi.* Where this information was not on record, PSSP staff visited the *Nyumbakumis* and obtained this information from the local people.

We entered each house listing into a computer to produce the sampling frame. Since the houses within the *Nyumbakumis* did not have names they were designated by numbers only, for example house #8 in Nyabirondo *Nyumbakumi,* Nyarubande Sector, Butamwa Commune. We then drew a random sample of 600 houses, using SPSS software. Six hundred houses was based on an overall target of 384 interviews with an estimated 20% refusal rate and 20% failure rate to find the house or an eligible respondent within it.[5,6]
Seventeen local interviewers were hired and the PSSP staff acted as supervisors. Qualifications for interviewing were a high school education, ability to read and write, ability to walk long distances and availability for the entire study period. Both interviewers and supervisors were given training in interviewing methods and divided into groups of one supervisor and 2-3 interviewers. The sample was divided according to geography and number of houses, and each section allocated to an interviewing group. Supervisors then allocated *nyumbakumi* to each of the interviewers.

Both interviewers and supervisors received training in how to identify respondents. Interviewers would first find someone living in the *nyumbakumi* and ask them to list the names of all the houses in the *nyumbakumi* in any order. Once the respondent has done this, the interviewer took the nth house named, n being the number of the house chosen for interview (for example #8 in the above example). At this house the person asked anyone living there to name all the adults living there over the age of 18 years, in any order. The interviewer listed these names, then used a random number table to choose the respondent. If the respondent was not there an appointment was made to return at a more convenient time, or the interviewer went and found the person if they were not far. If the respondent refused to be interviewed that interview was marked as a refusal and no-one else from that house was interviewed.

The role of the supervisors was to assist interviewers in finding houses and identifying the correct respondent, to regularly observe and provide feedback to the interviewer on some interviews, to re-interview 10% of respondents as a check of test-retest validity and to detect 'arm chair' interviewing, and to check all completed interviews for completeness and clarity. Supervisors also revisited most refusals, to ensure these were genuine.

Prior to each interview informed consent was obtained. If, during the interview, it was apparent that the person was severely distressed, this was reported to the supervisor who recorded the respondent's contact information. These respondents, and those who are diagnosed as depressed according to the interview, will be re-visited by World Vision staff to be assessed for counseling and referral if possible.

---

5 384 interviews is derived from the formula for sample size for estimating proportions: $n=z^2p(1-p)/(0.5d)^2$. n is the required sample size; p is the best estimate of the proportion prior to the study; d is the width of the confidence interval and z is the level of confidence that the estimate lies within that interval. For this study we do not have an estimate for p (the proportion of the population who are depressed) so, to be conservative, we use p=0.5 since this gives the largest sample size. We wanted to be 95% confident that our measured estimate would be within 5 percentage points of the true value (ie, confidence interval=10 percentage points). Therefore sample size= $1.96^2[0.5*(1-0.5)]/(0.5*0.1)^2 = 384.16$

6 Later, when interviewing at Butamwa was stopped early, 80 more houses were chosen at random in Kinzenze, which boosted the total sample in that Commune to 400.

6.7 Quantitative Analysis

Data from each interview in the validity study and survey was entered into a computer and analyzed using SPSS statistical software. Analysis consisted of tests of validity and reliability, and measurement of Depression and functional disability, and the relationships between them.

6.7.1 Testing Reliability

Reliability refers to the extent to which different measures of the same thing agree with each other. It can refer to measurements taken at the same time, or different times. To be useful an instrument must have good reliability. Reliability must be tested whenever a questionnaire is changed (including translation) or used among a new population. The most important aspect of reliability is internal consistency reliability. This refers to how well questions measuring the same thing on the same occasion agree with each other. For example, two questions that measure different aspects of Depression should agree with each other in that a person should score high or low on both.

Agreement is measured quantitatively by correlations. For questionnaires with many questions, a large number of correlations would be required to check the agreement of every question with every other question, and some summary of these correlations would be needed. Cronbach's alpha is a statistical measure which provides this. It is a single figure which summarizes the average correlation between all pairs of questions in a questionnaire. Cronbach's alphas should be above 0.7 and ideally between 0.8-0.9.[7] The reliability of each question can be assessed by calculating the alpha with and without it. Significant increases in alpha without the question would suggest that the question is not measuring the same thing as the other questions, and should be removed. Studying the effect of each question in this way is called Item Analysis.

Testing reliability over time is also useful. This is called test-retest reliability. The questionnaire is given to the same subject on two different occasions. It is usually done at least a day later, to reduce the effect of memory on the responses, but not too long because what is being measured may actually change (mood, for example). Therefore, the repeat interview is usually done 1-7 days after the first interview. A summary scale is created using all the questions on the same topic (in this case Depression and function) and calculated for both the first and second interview (see validity section below for creation of the function and Depression scales). Test-retest reliability is tested by measuring correlations between these scores. Opinions vary as to what is an acceptable score, although correlations above 0.7 are considered desirable. A problem arises in interpreting low scores. These may be due to a poor instrument, or to different interviewers, or because the concept being measured has changed. Partly for these reasons, test-retest reliability is not considered as important as internal consistency reliability (Streiner et al 1995).

Cronbach's alphas, item analysis and test-retest correlations can be calculated using simple statistical software, and interpreted using the parameters listed above. We measured internal consistency reliability separately for the male function, female function and Depression questions, by generating Cronbach's alphas and item analyses for each. Test-retest reliability

---

[7] Above 0.9 suggests that the questionnaires has too many questions and some could be eliminated (Streiner et al, 1995).

and reliability between different interviewers was tested simultaneously by having supervisors re-interview 10% of survey respondents 1-12 days after the initial interview. We then measured the Pearson correlations between the Depression scores and between the function scores on the initial and repeat interviews.

6.7.2 Testing Validity

While reliability refers to how well questions agree with each other and over time, validity refers to how well they measure what they are supposed to measure. Questionnaires may be reliable and not valid. For example, in our study respondents may have given consistent and reliable responses to a questionnaire assessing Depression even if Depression does not occur in that population, or if Depression occurs but the instrument is measuring something else. Some indicator outside the questionnaire is required to determine if Depression exists and if the instrument is measuring it and not something else.

There are three aspects of validity to be considered when testing a questionnaire:

a) Content validity refers to whether the instrument is considered *by experts* to be appropriate for measuring what it is supposed to measure. Part of content validity is whether experts believe that the questionnaire covers all the important aspects of the concept being studied. In our study we consulted two groups of 'experts.' We consulted experts in psychological measurement who helped us choose the HSCL as appropriate for measuring Depression. We also consulted knowledgeable persons (and confirmed their advice with pile sorts) as to whether other important aspects of Depression-like illness were being missed. In creating the questionnaire on function we took a similar approach; again consulting western experts as to what aspects of function were important to assess - they recommended self-care, and care of the family and community. We then consulted local people using these categories, considering a cross-section of them to be the best experts on the important tasks in their own lives.

b) Construct validity refers to how well items that are supposed to be associated with each other are actually associated. For example, if Depression occurs in our study population and a significant number of people are depressed, we should see associations between questions measuring different Depression symptoms. This is provided the questions measure these symptoms accurately. Therefore, construct validity refers both to the existence of the hypothesized construct (in this case Depression) among the population and to the ability of the questions to measure it. Since all the Depression questions measure aspects of Depression, they should all be associated with each other and Cronbach's alpha can be used as a measure of construct validity in this case.

Factor analysis is another statistical method that assesses construct validity. Factor analysis is a statistical method similar to Cronbach's alpha, except that it examines associations between sub-groups of questions, rather than across all the questions as does alpha. It demonstrates which questions are most highly correlated with each other. This information can be used to understand the different elements of depression that underlie the responses. Similarity between these constructs and hypothesized constructs support construct validity. Factor analysis is useful (but not essential) for this method as it is to be used by NGOs. Factor analysis of the data in this study is included in appendix H.

c) Criterion validity refers to the agreement between the questionnaire and an external measure (criterion) of the same construct known to be accurate. In other words, comparing the

questionnaire with a 'gold standard.' This is not strictly possible by this method, since it is meant to be used where none of the usual gold standards currently exist. Instead, we used an alternative local standard - diagnosis by local people of a local of illness similar to Depression. As noted previously, we used only respondents for which the respondent themselves and a key informant agree as to whether the respondent had this illness, to increase the accuracy of the diagnosis.

Analysis consisted of testing the correlation between diagnoses of Depression and cases/non-cases of the locally-defined illness. We used the Phi correlation coefficient because it measures correlations between dichotomous data. Dichotomous means that the data values can only be yes or no, rather than a number on a scale as, for example, with temperature. High correlation would suggest that Depression and the local illness were similar; no significant correlation would mean that they were unrelated. We also measured the sensitivity and specificity of the local illness for Depression, to provide additional information on the nature of any relationships between the illnesses (See Results section).

Depression is known to have a significant effect on function. Therefore testing for an association between Depression diagnosis using our Depression questionnaire and reduced function using our function questionnaires also provided some criterion validity, both for the Depression questions and the function questions.

6.7.3 Measuring Depression and Functional Disability

From the survey data we calculated the prevalence of Depression using the DSM algorithm. We also calculated the prevalence of significant difficulty with each of the function tasks, separately for men and women and for depressed vs not depressed. We then calculated the proportion of cases of significant difficulty attributable to Depression among the depressed of that sex, and among all persons of that sex. We created a summary scale of functional disability - a single number that summarized the disability reported for all tasks: For each individual we added the scores for all the function questions (but excluding the final question about other tasks since not all respondents answered this. Also, some interviewees did not give responses for some tasks because they were not responsible for those tasks. For example, some women did not have children and so did not care for them. In these cases we assigned a value to this task which was the mean of the responses for the other tasks). We then divided this score by the maximum possible total score for that sex (28 for men and 36 for women). This gave us a score which was a percentage of the maximum possible disability, and which could be compared across sexes. On this scale 0% represents the same impairment as healthy others of the same age and sex and 100% represents the maximum relative impairment. We then used regression techniques to determine the relationship between Depression and functional disability.

We generated a Depression scale that was a simple summation of all the Depression questions. On this scale we determined a cut-off level that was most efficient for detecting significant functional disability and Depression (diagnosed by algorithm). This cut-off is intended to replace the algorithm for future measurements of this population. The reasons for using this approach in future instead of the algorithm are:

- Scales have superior reliability than algorithms
- Generating a scale enabled us to incorporate information on the local symptoms that are found to be useful.

- Using functional disability as the main criterion for choosing the cut-off (see ROC Curve below) enabled us to also incorporate this information into the Depression diagnosis.
- Using scales enables measurement of severity of symptoms as well as diagnosis.

To determine the correct cut-off score we generated a Receiver Operating Characteristic (ROC) Curve with maximum sensitivity and specificity in detecting functional disability as the criterion for selecting the score. This required first setting a cutoff score on the function scale above which significant functional disability was considered to be present. We decided to use > 25% as the cutoff since 25% was the maximum score on either the male or female function scale that a respondent could score and not be significantly impaired on any task.

7. RESULTS

7.1 Study Site

We studied Sectors of the Communes of Butamwa and Kinzenze in the Kigali Rural Prefecture. Each Commune has between 100,000-160,000 people. They were selected because World Vision plans to work in both areas in the near future. Kinzenze was one of the Communes worst affected by the genocide in 1994, because it contained many Tutsi families. Butamwa had few Tutsi and so was less affected.

7.2 Results of the Ethnographic Study

7.2.1 Free Listing

For the first activity - the free list on problems - five PSSP interviewers spoke with 40 knowledgeable persons across both Communes. The key informants were identified by the PSSP staff, based on their experience working in these areas. The results are summarized in Table 1.

Table 1: Results of Free Lists on major problems resulting from the genocide in 1994.*

Problem	Ranking	# of responses that include this problem
Poverty	1	41
Lack of food	2	40
Lack of people (from any cause)	3	25
Suspicion/mistrust with breakdown of neighborly relations	4	23
Too many widows and orphans	5	20
Lack of possessions - land, housing, shelter	5	20
Illness (physical)	7	15
Mental Trauma (*Guhahamuka*)	8	14
Lack of motivation/hope	8	14
Lack of justice	8	14
Too many people in prison causing a shortage and those outside must care for them.	11	13
Lack of schools	12	11

Physical Disability	13	7
grief (*Agahinda*)	13	7
Ignorance	13	7
Government program to resettle villagers	16	5
Drunkenness	16	5
Unwillingness to change	18	4

Persons who mentioned Mental Trauma (*Guhahamuka*) elaborated on it with one or more of the following descriptions:

- losing your mind
- feeling isolated
- feeling like committing suicide
- lack of love
- feeling hopeless
- feeling that your life is not worthwhile
- feeling like you are dead and it would be better if you were
- envying the dead
- having no hope.

Only people in the Kinzenze Commune used the term *Guhahamuka*. Those in Butamwa used the term *Agahinda*. Both PSSP staff and interviewees said that *Guhahamuka* is a new concept in Rwanda, having emerged after 1994, whereas the concept of *Agahinda* is much older. We note that World Vision has conducted an education and community outreach campaign in Kinzenze to teach people about mental trauma whereas no such activities have been conducted in Butamwa, by World Vision or other organizations. It is likely that the mental trauma concept has been introduced from outside the country, and that Kinzenze residents have learned to list all perceived mental and emotional effects of the genocide under the category of *Guhahamuka*. In contrast, Butamwa residents use the more traditional term. If this distinction is indeed artificial then *Guhahamuka* and *Agahinda* should be classed together and would rank #5 on the list of problems.

In a separate free listing activity, interviewers spoke with a convenience sample of 25 men and 25 women on the important tasks they do regularly. The results (after excluding tasks that did not appear to affect others) are shown in the function section of the instrument in Appendix A. 'Transmitting Culture' under the female tasks refers to education of children by their mothers.

7.2.2 Who Do People Consult?

During the free listing exercise, whenever a potential mental or emotional problem was mentioned respondents were asked to state who people consulted for that problem. The results are shown in table 2.

Table 2: Types of Persons Consulted for Mental and Emotional Problems*

Type of Person Consulted	# of Respondents naming that person
Nobody	24
Local Leaders	14

God	6
Friends	4
Don't know	2
World Vision staff	2

* table excludes responses given by only one respondent.

On reviewing these results PSSP staff said that local people also frequently consult traditional healers for these problems. Admitting this is not socially acceptable for persons who profess Christianity, as the vast majority of the population do. This effect may have been exacerbated in our study, since interviewers introduced themselves as working with World Vision, a well known Christian organization. For these reasons we decided to consult a mixed group of traditional healers and local leaders for the next phase of the study.

7.2.3 Key Informant Interviews

Seven traditional healers and local leaders were interviewed twice. They were asked to name and describe the possible problems a person with Depression-like symptoms might have (see Methods). At the end of the interviews respondents were also specifically asked to describe *Akababaro*. This was a term said by the translators to be a local term for Depression, although it was not mentioned by any free list respondents. We wished to check whether it might be a valid local term for Depression-like illness that failed to emerge in the free lists. The results are shown in Figure 1.

Figure 1: Local Depression-like syndromes and their symptoms.

Term: *Guhahamuka* (Mental Trauma)

Symptoms:

failure to sleep
despair, hopelessness
anger
failure to eat
failure to talk
loss of intelligence*
attempting suicide
Confusion about issues/different things*
acting like a crazy person*
Mixed feelings and thoughts in your head at the same time.
Feeling extremely weak
Increased startle reaction
Absentmindedness
Too many thoughts
Worthlessness
feeling as if you would be better to die
Lack of concentration
feel like you have a "cloud" within oneself (a vague painful constant presence internally)*
feeling disconnected from others and the world
falling sick every now and then
to keep dreaming of the events he/she went through
fleeing away from people and hiding*
lack of trust*
feeling like fighting*
Making a lot of noise

being quarrelsome
being rebellious
excessive crying (by women and children)
Talking to anybody who comes your way about your pain*
being violent
occasional chaos in the mind (can be a flashback)
instability of the mind.*
feeling like you are having an epileptic episode (collapse).
acting without thinking*
having a nightmare about being in a fight.
deep sadness that can result in death

Term: *Akababaro**

Symptoms

to be extremely quiet
Not pleased by anything*
Feeling very weak
dying with (not from) sadness.

Term: *Agahinda* (deep sadness or grief)*

Symptoms

isolation
lack of self care
loss of mind
being very talkative*
not working
drunkenness
feeling life is meaningless
feeling shattered
committing suicide
don't feel like talking
excessive alcohol drinking causing crazy behavior
being displeased with your living conditions/situation. Not pleased by anything.*
Inability to withstand whatever happens to you
burying one's cheek in his/her palm
difficulty in interacting with others (poor relationships)
sadness.

*included in the subsequent pile sort exercise.

This activity confirmed the existence of the two major cover terms for Depression-like symptoms. We failed to discover any new cover terms. *Akababaro* appeared to be closely related to *Agahinda*, but with much more limited scope. It did not figure prominently in this activity, or at all in the original free list. It was also apparent from this activity and the first free list that many of the symptoms included under these cover terms can also occur in isolation. Also, *Agahinda* refers to both a grief syndrome and to the symptom of deep sadness that accompanies grief.

Between them *Guhahamuka* and *Agahinda* include all the symptom categories required for diagnoses of Depression as well as PTSD. This supports criterion validity of both syndromes among this population, although perhaps less so for PTSD because of the pre-existing education program centered on this illness. Because of the limited scope of *Akababaro*, and the

wide variety of both PTSD and Depression symptoms in *Guhahamuka* (and its recent origins) we decided that *Agahinda* was the local term closest to Depression yet intelligible to both communities. During the course of the ethnographic study, we learned that *Agahinda* has varying degrees of severity from mild to severe. We therefore chose words for the severe form - *Agahinda gakabije* - as the local syndrome most similar to Depression.

7.2.4 Pile Sorts

Before proceeding with the survey, we explored whether there were additional symptoms characteristic of Depression in this population that should be included in the instrument. We selected symptoms from the key informant interviews that are not part of the PTSD or Depression syndromes, as they are described in the DSM. We then conducted a pile sort of these symptoms with *Agahinda*, *Akababaro* and the symptom 'not being pleased by anything' as 'anchors' for the Depression concept; to confirm whether local people perceive relationships between these Depression symptoms and these local symptoms.

Pile sorts using these factors were conducted on a convenience sample of 40 men and women in both Communes. The results are summarized in table 3.

Table 3: Results of pile sort exercise for local features of Depression-like problems

1	X													
2	5	X												
3	6	9	X											
4	3	12	10	X										
5	6	8	3	10	X									
6	1	5	11	14	7	X								
7	8	5	2	8	12	5	X							
8	10	7	1	6	10	5	8	X						
9	10	9	3	6	19	0	8	15	X					
10	6	6	11	5	11	13	7	4	6	X				
11	7	4	3	4	10	2	11	16	11	6	X			
12	13	11	3	7	8	4	10	13	11	3	8	X		
13	5	8	11	11	13	21	4	5	3	14	4	4	X	
14	1	6	10	10	8	29	1	4	0	15	3	2	18	X
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**

Symptoms are designated by the marginal numbers as follows:

1. Feeling like fighting
2. Fleeing away from people and hiding
3. Feeling like you have a "cloud" within oneself (a vague painful constant presence internally)
4. Instability of the mind
5. Loss of intelligence
6. *Agahinda*

7. Confusion of issues/things
8. To talk to anybody who comes your way about your pain
9. To act like crazy people
10. Not pleased by anything
11. Being very talkative
12. Acting without thinking
13. Lack of trust
14. *Akababaro*

Numbers within the body of the table record the total number of times the two marginal symptoms occurred together in all the pile sorts. For example, symptom # 11(being very talkative) was put in the same pile as symptom # 5 (loss of intelligence) 10 times and in the same pile as 'lack of trust' four times.

Review of the pile sort results shows support for the expected close relationship between *Akababaro*, *Agahinda* and anhedonia. Feeling there is a cloud within oneself, instability of the mind, loss of intelligence, and lack of trust are also possibly related to these problems, and therefore may constitute local symptoms of Depression. In the case of feeling a cloud within, we were unable to arrive at a more acceptable translation of this concept, which made us hesitant to use it. Therefore, questions on the latter three symptoms only were added to the instrument.

7.3 Validity and reliability of survey instruments

7.3.1 Validation Study

We did not identify a single syndrome with all the features of Depression only. Therefore we could not make a direct comparison of Depression diagnosis *vs* a local equivalent. The closest locally described syndrome is *Agahinda gakabije*, which is more accurately translated as severe grief. Since grief and Depression are related in the West, we studied criterion validity by examining the relationship between Depression and *Agahinda gakabije*.

Blinded interviews were conducted on 81 people said by key informants to have *Agahinda gakabije* and 50 said not to have it. These respondents were also asked if they thought they had *Agahinda gakabije*. 38 cases were thrown out because of disagreement between the key informant and the respondent, to ensure that we had true cases and true non-cases.

Using the remaining 93 respondents we then compared the correlation between *Agahinda gakabije* diagnosed by key informant and respondents, and diagnosis of Depression by the instrument. We also assessed the sensitivity and specificity of *Agahinda gakabije* in detecting Depression (see table 4).

Table 4: Phi correlation between diagnoses of Depression and *Agahinda gakabije* and sensitivity and specificity of *Agahinda gakabije* for Depression (N=131)

Diagnosis of *Agahinda gakabije*	Correlation with Diagnosis of Depression	Significance	Sensitivity*	Specificity*
by informant and self	0.375	0.000	94.7%	38.2%

* in detecting Depression.

Correlation is significant but low, suggesting that these problems are somehow related but are not the same. The sensitivity/specificity data suggest that Depression forms a sub-group of those with *Agahinda gakabije*. All but two of thirty-eight respondents diagnosed as depressed also had *Agahinda gakabije*. This is consistent with the findings from other cultures that grief can trigger Depression, but only in a proportion of cases (Craig, 1996).

Tables 5-7 show Cronbach's alpha coefficients and item analyses for the questions on Depression and male and female function. These demonstrate good internal reliability for all three groups of questions. The correlations of the local Depression symptoms added to the instrument are comparable to those of the DSM symptoms. This suggests that they are part of the Depression syndrome as indicated in the pile sorts.

Table 5: Cronbach's alpha coefficient and item analysis for survey questions on Depression- Kinzenze Sectors (N=368).

	Cronbach Alpha if item deleted
B1-ENERGY	.8573
B2-BLAME	.8656
B3-CRYING	.8626
B4-FIDGETING	.8542
B5-APPETITE	.8596
B6-SLEEP	.8579
B7-LACK OF HOPE	.8594
B8-BLUE	.8511
B9-LONELY	.8538
B10-SUICIDE	.8661
B12-WORRY	.8621
B13-LOSS OF INTEREST	.8690
B14-INCREASED EFFORT	.8549
B15-FEELING WORTHLESS	.8560
B16-LOSS OF TRUST*	.8645
B17-LOSS OF INTELLIGENCE*	.8601
B18-INSTABILITY OF MIND*	.8563
B19-LOSS OF SEX INTEREST	.8742
Cronbach's alpha for all questions	**.8672**

* local symptoms added on basis of ethnographic study results.

Table 6: Cronbach's alpha coefficient and item analysis for survey questions on male function- Kinzenze Sectors (N=125).

	Cronbach alpha if item deleted

MA1-WASHING	.7803
MA2-DRESSING	.7783
MA3-ADVISING FAMILY	.8061
MA4-ATTEND COMMUNITY MEETINGS	.7931
MA5-SOCIALIZING	.8090
MA6-LABOR	.7653
MA7-EARNING MONEY	.7992
Cronbach's alpha for all questions	**.8154**

Table 7: Cronbach's alpha coefficient and item analysis for survey questions on female function- Kinzenze Sectors (N=243).

	Cronbach alpha if item deleted
FA1-WASHING	.8091
FA2-DRESSING	.8032
FA3-COOKING	.7917
FA4-WASHING CLOTHES	.7747
FA5-CLEANING HOUSE	.7864
FA6-CARING FOR CHILDREN	.8004
FA7-ATTENDING COMMUNITY MEETINGS	.7939
FA8-SOCIALIZING	.8343
FA9-TRANSMITTING CULTURE TO CHILDREN	.8305
Cronbach's alpha for all questions	**.8223**

7.3.2.1 Test-Retest Reliability

Thirty-seven interviews (10% of total) were repeated by the supervisors. These interviews were chosen at random and occurred 1-12 days after the initial interviews.

A Depression scale was created using all the Depression questions, including the local symptoms. A functional disability scale was also created (see Methods). The Pearson correlation coefficient for test-retest reliability was 0.671 (p=.000) for the Depression scale and 0.574 (p=0.000) for the functional disability scale.

7.4 Survey findings

7.4.1 Number Interviewed

We conducted a community-based survey in Butamwa and Kinzenze Communes. In Butamwa we contacted 90 households and interviewed 72 people. Twelve people refused to be interviewed and 6 could not be located. A true response rate could not be calculated since we ceased interviewing early (see below). However, based on the number of contacts the response rate = 80%.

In Kinzenze we contacted 380 of the 400 houses on the list. We interviewed 368 respondents, 5 people refused and in 7 cases interviewers could not find the house or respondent, or no-one eligible lived in the house (empty or child-headed household). The response rate was 98.7%.

Fewer people were interviewed in Butamwa because interviewing was ceased once it was clear that Depression in that Commune was uncommon. This meant that Butamwa would not be a priority area for any subsequent Depression interventions, and the Butamwa teams were brought to Kinzenze to help complete the survey for that Commune.

7.4.2 Demographics

Table 8: Demographics of Study Sample

	N	Minimum	Maximum	Mean	S.D.
Age					
Kinzenze	368	18	85	35.65	14.62
Butamwa	72	18	78	35.49	13.39
Education					
Kinzenze	368	0	14	3.29	3.03
Butamwa	72	0	11	3.33	3.03

The demography of the two Communes is similar in terms of age and education. However, in Kinzenze 34% of respondents were male whereas in Butamwa only 20.8% were male. The difference may be real - the Butamwa respondents to the first free lists complained of lack of men more than those in Kinzenze. The difference may also be artifactual: in both Communes it was more difficult to locate men at home than women. Butamwa interviewers had less time to follow-up male respondents which may have artificially reduced their number.

7.4.3 Prevalence of Depression

Table 9: Prevalence of Depression and Severe Grief (*Agahinda gakabije*) Across Selected Sectors in Kinzenze Commune.

	Total (N)	# with Depression	% depressed* (95% CI)	# with *Agahinda gakabije*†	% with *Agahinda gakabije* (95% CI)	Odds Ratio: Depression if have *Agahinda gakabije* (95% CI)
Men	125	17	13.6 (7.6-19.6)	52	41.6 (32.9-50.3)	5.75 (1.83-17.88)
Women	243	49	20.2 (15.2-25.2)	102	42.0 (35.8-48.2)	9.46 (4.38-20.41)
Total	368	66	17.9 (14.0-21.8)	154	41.8 (36.8-46.8)	8.1 (4.26-15.44)

* Based on algorithm
† Self described

4/72 or 5.6% +/-5.3% of adults in Butamwa have Depression, by the same criteria. 31.9% +/-4.8% described themselves as having *Agahinda gakabinje*.

Because of concerns about adequate randomization in Butamwa (see above) we did not combine the data from the two Communes in our analyses. Therefore subsequent analyses (and the preceding validity and reliability analysis results) refer to the Kinzenze data only.

7.4.4 Relationship between Grief and Depression

We calculated the sensitivity and specificity of *Agahinda gakabije* as a screening tool for Depression among the survey population (see Table 10). We found that asking people whether they had *Agahinda gakabije* would pick up 80% of Depression cases.

Table 10: Sensitivity and specificity of *Agahinda gakabije* in detecting Depression among survey population

Diagnosis of *Agahinda gakabije*	Sensitivity	Specificity
By informant or self	86.4%	64.1%
By self only	80.3%	66.4%

7.4.5 Function

7.4.5.1 Demographics of functional impairment

Men and women show similar levels of overall functional disability, among both the depressed and non-depressed (See Table 12). However, there are differences in their abilities to perform specific types of tasks, and in how Depression may affect these abilities (Table 11). For men, Depression is associated with significantly elevated dysfunction in all measured tasks except socializing and attending meetings. Since even the not depressed frequently report difficulties in performing labor and earning money, in the presence of Depression two thirds of depressed men report difficulties with these tasks.

Non depressed women do not report difficulties with any of their tasks as often as men do with labor and earning. However, they do frequently report difficulties with cooking, washing clothes and caring for children. Depression is most highly associated with increased difficulty attending meetings, cleaning house, washing clothes and caring for children. Only 'transmitting culture' was not significantly associated with Depression in this sample.

Table 11: Comparison of task-specific functional impairment between Depressed and Non-Depressed.*

	Not Depressed		Depressed		Odd ratio: dysfunction if Depressed (95% CI)†	% Prevalence attributable to Depression among depressed	% Prevalence attributable to Depression among population
	N	% dysfunctional for task	N	% dysfunctional for task			
Male	108		17				

Washing	8	7.4	6	35.3	6.82 (2.08-22.6)	79	33.9
Dressing	7	6.5	5	29.4	6.01 (1.74-21.09)	77.9	32.3
Advising	11	10.2	6	35.3	5.01 (1.61-15.77)	71.1	25
Meeting	8	7.4	2	11.8	1.67 (0-7.75)	37.3	7.5
Socializing	7	6.5	2	11.8	1.92 (0-9.10)	44.9	9.7
Labor	27	25	11	64.7	5.5 (1.91-15.76)	61.4	17.8
Earning	31	28.7	12	70.6	5.96 (2.01-17.60)	59.3	16.6
Female	194		49				
Washing	15	7.7	14	28.6	4.77 (2.14-10.65)	73.1	35.5
Dressing	16	8.2	12	24.5	3.61 (1.60-8.16)	66.5	28.8
Cooking	26	13.4	17	34.7	3.43 (1.69-7.00)	61.4	24.3
Wash clothes	25	12.9	24	49	6.49 (3.24-13.01)	73.7	36.0
Clean house	15	7.7	21	42.9	8.95 (4.17-19.24)	82.1	48.0
Care for children	19	9.8	18	36.7	5.35 (2.55-11.23)	73.3	35.6
Attend meetings	13	6.7	22	44.9	11.34 (5.16-24.92)	85.1	53.5
Socialize	8	4.1	6	12.2	3.24 (1.12-9.46)	66.4	28.8
Transmit culture	9	4.6	3	6.1	1.34 (0.38-4.79)	24.6	6.8

* Respondents were considered dysfunctional for a task if they reported at least moderate difficulty compared with normal people of their age and sex.
† Confidence Interval.

Table 12: Distributions of Functional Impairment Scores for Men and Women, Kinzenze Commune.

	N	Minimum	Maximum	Median (all)	Median (Not depressed)	Median (depressed)
Men	125	0.0	71.43	7.14	7.14	25.00
Women	243	0.0	75.0	8.33	5.56	22.22

7.4.5.2 Functional disability, Depression and other variables

We used regression analysis to study the relationships between these variables. Since most people reported normal function, the distribution of function scores was skewed to the right and linear regression was not appropriate. Therefore, we converted the function score into a dichotomous variable: all scores less than or equal to 25 were considered as not significant functional disability and all scores greater than 25 as significant functional disability. We then

performed a binary logistic regression with the dichotomous functional impairment variable as the dependent variable and age, sex, education, *Agahinda gakabije* and Depression diagnosis by algorithm as the independent variables.

The full logistic regression model is shown in Table 13. Only age and Depression are significantly associated with functional disability. When Depression is removed from the model there is a highly significant relationship between Agahinda and functional disability. This suggests that most reduction in function among those with *Agahinda* (or grief) is among the subset with Depression. Those suffering grief but not depressed do not suffer significant functional impairment, compared with others their age and sex.

Table 13: Logistic Regression of age, sex education and *Agahinda* (self-diagnosed) *vs* significant functional disability.

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Age	.051	.011	.000	1.052
Sex	.287	.341	.400	1.333
Education	-.018	.062	.776	.983
Agahinda	.463	.365	.205	1.59
Depression	1.429	.382	.000	4.18

We repeated this regression analysis for men vs women, low age (less 28 years) vs high age, low education (less than 4 years) vs high education. The cut-offs for age and education are the medians for each variable. The results are in Appendix I. Briefly, the models for men and women were not significantly different. The correlation between Depression and function was greater in those with more education. Depression and function were not associated in the young, but were highly correlated among older respondents.

7.4.5.3 ROC Analysis

We generated an ROC curve of Depression scores *vs* function as a dichotomous (yes/no) variable. The Depression score consists of a simple summation of all the Depression questions (except 'feeling trapped' which was not used in any analyses). We used the same dichotomous function variable that we used in the regression analysis (ie, less than or equal to 25, vs greater than 25). The resulting ROC curve is shown in Appendix G, Figure 2. Table 15 (also Appendix G) shows the plotted points of the same curve. Examining the curve and points shows that the score which maximizes sensitivity and specificity is 30.5. At this score Sensitivity = 68.3% and Specificity = 71.8%. Check against a similar ROC curve of Depression score *vs* Depression diagnosis by algorithm (Appendix G, Figure 3 and table 16) shows that this score has a sensitivity = 98.5% and specificity = 79.1%, suggesting that this cut-off score is also very accurate for diagnosing Depression.

7.4.6 Alcohol Use

This did not prove to be a useful question. The overall mean response was 0.3 drinks/day with 78.5% reporting no alcohol intake. The mean for men only was 0.6 drinks/day with 64%

reporting no alcohol intake. As drinking appears to be widespread it appears that admitting higher levels of intake is not socially acceptable, and the results are of doubtful accuracy.

8. DISCUSSION

8.1 Process

8.1.1 Overview

This study was the first field trial of a method to assess the burden of mental problems across cultures. In designing this method our major considerations were to enable NGOs to:

1. Decide whether standard mental health assessment instruments are appropriate (ie, do the diseases they assess occur locally).
2. Where instruments are appropriate, to adapt them to the local situation.
3. Test the validity and reliability of the adapted instruments.
4. Use these instruments in community based surveys to assess need and baseline figures.

Existing methods to perform the first 3 vital tasks require time and resources beyond the means of most NGOs, and in many cases the necessary gold standards are just not available. *Our approach, was to understand how local people view mental health so that we could enlist their assistance in answering these questions.*

For this first trial we worked in Rwanda because of our interest in transitional populations, and because World Vision has a psychosocial program there. For transitional populations affected by war (and genocide) the two most severe mental health problems are Depression and PTSD, which commonly occur together (Engdahl et al, 1998. Shalev et al, 1998. Peltzer, 1998). Resources prevented us from investigating more than one disorder so we focused on Depression, although we could have investigated PTSD or any mental illness or health issue using this method. We chose Depression because - unlike PTSD - it is also common in countries without experience of war (Weissman et al, 1996). Therefore our results could have relevance beyond the scope of transitional populations.

8.1.2 Details of Rwanda experience

Training and execution took 5 weeks, not including a one-week delay used to create a sampling frame. The ethnographic study and concurrent translation required two weeks, the training and pilot and validation studies required one week, and the survey training and data collection a further two weeks. Six PSSP staff, the PSSP director and 2-3 drivers worked full-time on this project for the duration. Seventeen additional workers were hired for the quantitative interviewing.

Most of project activities proceeded smoothly and the exercise has generated useful data. We believe our experience has shown that it is possible to train local staff and conduct a rapid ethnographic and quantitative study in a very short time with resources currently available to many NGOs and other organizations - the only additional requirements are training and a commitment to understanding local communities. Studies like this one are best done at the beginning of programs, even before the interventions have been decided: Ethnographic methods provide a lot of general information about these communities which can be used to plan a program, as well as an effective way to meet local people and build trust. For example, the first free listing exercise provides information on all the community's problems that can guide

an NGO in setting up all programs, not just programs for mental health. The faculty at JHU have designed other ethnographic toolkits for more general (but still rapid) assessments of transitional populations (Weiss et al, 1999).

Despite reservations by both international and local staff, in the validity study Rwandans did not object to identifying others with severe grief. Nor did those identified object. Using the information from these informants and the respondents themselves to diagnose true cases and true non-cases of grief appeared to work well. However, grief and Depression are not directly comparable, and this reduced the utility of the results. The validity study would be more effective in populations which recognize syndromes closer to Depression.

Some problems did emerge during the field trial, and need to be improved. Preparing the study site by means of consultations with leaders prior to the survey, and maintaining contact with them during the study, was done poorly. This was due to our failure to decide on a study site prior to the arrival of the project co-director, and to competing time demands on the PSSP staff despite the full-time requirements of the project. The decision to use simple random sampling within the Sectors considerably increased the time needed to create a sampling frame, and to collect data. Towards the end of the data gathering interviewers noticed some reactivity to the survey, suggesting that the data gathering was taking too long. Future surveys should use the more rapid 30 cluster sampling procedure and reduce the data gathering phase from 10 to 5 days. Based on our experiences in Rwanda, we developed a time-line to address these issues in future studies (see Appendix D).

The existing instrument also requires some changes, based on our experience. In future studies of Depression using the HSCL we will omit the question on feeling trapped, since this is not part of the Depression criteria and was not used in the survey. It will be replaced by a question on psychomotor retardation, a DSM category missing from the current instrument. The questions on lack of interest, lack of interest in sex, and on alcohol consumption will be retained for the time-being, in case the lack of usefulness was particular to this group. The questions on function performed well, including the response categories and the non-verbal response card. In other words, we recommend that the instrument largely retain its current form (see Appendix A) until further experience is gained.

We trained local staff in instrument preparation, data collection and qualitative data analysis. We did not train them in quantitative data analysis and interpretation, because of time and resource limitations. In future, this training can be provided to selected staff with computer skills. Note that the analytical procedures described in this report are necessarily more complicated than those that will ultimately by used by NGOs using this method. As appropriate with early testing of any new method, we stringently tested the validity of the Depression concept and the questionnaire, using factor analysis and relatively complex forms of reliability testing, regression techniques and ROC curve generation.

Eventually the quantitative analysis section of this method will consist of the following:

- Calculation of correlations between Depression and local illness
- Generating and interpreting Cronbach's alpha scores and item analysis
- Calculation of prevalences
- Creation of simple function and Depression scales
- Calculating cut-off scores on the Depression scale most appropriate for detecting functional disability.

Although some elements of this analysis appear complicated, staff *with computer skills* can be taught to conduct them and interpret the results. Training in these methods should be tested in subsequent field trials of this method.

8.2 Results

Our results suggest that Depression occurs among this population even though it is not recognized locally as a distinct syndrome. The validity and internal reliability of the instrument were also good although the test-retest reliability were adequate only. This combination suggests that moods may vary between tests, or that there is reactivity to the survey which is expressed in the second interview. However, this was not severe enough to invalidate using the instrument.

Depression rates were much lower than grief, and there was a large difference between the two areas: 17.9% of adults depressed in the Kinzenze Sectors and 5.6% in Butamwa. A study of Depression in 10 countries (not including Africa) found prevalences ranging between 0.8-5.8% (Weissman, et al, 1996), which was consistent with results from other parts of Africa (Bhagwanjee et al, 1998). This suggests that the Butamwa data may be close to the background level of Depression in this population, whereas the higher rate in Kinzenze may reflect the more severe genocide experience in that Commune.

We asked respondents to tell us whether they were suffering not just from grief, but from severe grief. The term *Agahinda gakabije* denotes grief so severe that the person is having great difficulty living with it, and we indeed found a very strong relationship with reduced function. However, this relationship was not apparent when depression was accounted for, which suggests that the reduced function is due to the association between depression and grief. Even people with severe grief, in the absence of depression do not suffer significantly reduced function. The more general implication may be that even severe levels of distress in those without mental illness do not greatly affect function. To test this hypothesis these methods should be used in similar research among other populations, and to examine other mental illnesses (such as PTSD). If confirmed, it would support a shift from current approaches which broadly address trauma experience and grief, to focusing on those who have developed mental illness as a result

Of interest is not only the global reduction in function with depression, but which tasks are affected. Among both men and women depression is associated with dysfunction in tasks important to the care of the family, including children. Among those depressed most of the difficulty with these tasks was associated with the depression itself, and with a significant proportion of the difficulty reported by the entire sample. This is remarkable, considering the many other problems and diseases that cause dysfunction in this part of Africa. It also carries implications for social and economic development, since the earning power and well-being of the family are compromised. If Depression is a direct cause, then Depression in an adult will significantly affect the whole family, and treatment will be required before the family can fully take advantage of available social and economic resources to improve their situation.

Since this was a cross-sectional study we could not prove that Depression causes dysfunction. It is possible that primary dysfunction could produce Depression, or that there are mixed effects. Research is required to sort this out. The most efficient way to do this would be a study of effective interventions specifically for depression. Concomitant improvement in depression and

function with depression treatments would then demonstrate the link. This type of research would also demonstrate whether function can be improved by treating depression, and could be used to test new interventions appropriate for areas with limited resources. If Depression is shown to be the cause of dysfunction, to be treatable, and that function improves as a result, then the prevalence of Depression becomes a developmental issue and should be assessed whenever high rates are suspected, such as after a war or disaster. We consider that the measurement method we have described here provides a means of conducting this type of research.

During the survey interviewers noted that those subsequently diagnosed as depressed were obviously distressed, and very keen to talk about their experiences. Interviewers reported that these respondents were grateful for the opportunity to speak about their problems, and many said that this was the first time they had discussed them openly with anyone. PSSP staff also noted that these persons were not picked up by their existing outreach program, despite the program's good coverage. The same was true of persons diagnosed with Depression during the validity study. This may be due to the reclusive behavior of the depressed, as expressed by the respondents and noted by the interviewers. Whatever the cause, it suggests that current programs are not reaching those most in need of assistance. These are the people whose mental distress is the most severe, and whose reduced function is most likely to affect their own well-being and that of the community.

We found that using *Agahinda gakabije* as a screening tool for Depression would detect 80% of Depression cases in the survey population.[8] If WV Rwanda decides to assist person with Depression in future programs, these persons could be found by a two step process:

1. Use current outreach resources to publicize the importance of *Agahinda gakabije* and recommend they present to community outreach workers trained by WV.

2. WV workers use the instrument and Depression scale to identify those depressed.

9. RECOMMENDATIONS

For further assessment

1. World Vision and Johns Hopkins University should repeat this assessment procedure with other populations; to assess need and build up a composite picture of mental illness across Africa, and to continue to refine and simplify the method. Future changes should include assessing exposure to events and their relationship to mental illness, assessing personal and community factors which may mediate reactions to trauma, and including in the ethnographic study investigating community suggestions for addressing these issues. Improvements should also include assessments for other mental problems (our study found good evidence for frequent occurrence of PTSD), and for children. Future field trials should include training in quantitative analysis.

8 The difference between this figure and the validity study (97.4%) likely reflects differences in the severity of grief studied in each group. Cases in the validity study had to be severe enough to be obvious not only to the person affected but also to others, whereas the survey group included cases apparent only to the individual affected.

2. Current programs do not appear to reach those most in need of mental health assistance. World Vision should form a technical advisory group composed of World Vision staff and experts in the field of mental health issues. This group should review these results and consider how WV and other NGOs can best help persons with Depression, given existing NGO resources. Communities should contribute to this process, and particularly to the review of proposals for feasibility, acceptability and implementation.

3. World Vision Rwanda should use the recommendations in 2. to assist those identified in the survey as depressed, and screen for others with Depression. The survey should also be repeated in other areas in which WV Rwanda suspects a need and is planning a mental health intervention, to assess the level of need.

4. Finally, we must emphasize that the free listing and validity study both revealed that poverty and lack of people are more pressing problems for most people than mental and emotional issues. One of the positive features of this assessment method is that it should enable the NGO to put mental health in the context of other problems. WV Rwanda should use information to prioritize all their interventions for Kinzenze and Butamwa Communes, and repeat the free listing in other areas prior to planning interventions.

For new research

1. Likely interventions resulting from the technical advisory group (and other sources) should be tested using standard research protocols. Such protocols could be designed and executed which could investigate three outstanding issues:
    - To find effective treatment for depression, given limited resources and large group of affected persons.
    - The nature of the cause-effect relationship between depression (and other mental illness) and function.
    - To determine if improvement in depression results in improvement in function and how much.

2. To repeat (or expand) these protocols to other mental illnesses associated with significant dysfunction.

10. REFERENCES

Andresen EM. Malmgren JA. Carter WB. Patrick DL. Screening for Depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). American Journal of Preventive Medicine. 10(2)1994 Mar-Apr: pp 77-84.

Bhagwanjee A, Parekh A, Paruk Z, Petersen I, Subedar H. Prevalence of minor psychiatic disorders in an adult African rural community in South Africa. Psychol Med 1998, Sep;28(5): pp 1137-47.

Bleich A, Koslowsky M, Dolev A, Lerer B. Post-traumatic stress disorder and Depression. An analysis of comorbidity. Br J Psychiatry May 1997;170: pp 479-82.

Craig TKJ. Adversity and Depression. International Review of Psychiatry 1996;8:341-53.

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Copyright 1994 by the American Psychiatric Association.

Engdahl B, Dikel TN, Eberly R, Blank A. Comorbidity and course of psychiatric disorders in a community sample of former prisoners of war. American Journal of Psychiatry 1998 Dec;155(12):1740-5.

Health Measurement Scales: A practical guide to their development and use. Streiner DL and Norman GR. Oxford University Press 1995.

Mollica RF, McInnes K, Sarajlic N, Lavelle J, Sarajlic I, Massagli MP. Disability associated with psychiatric comorbidity and health status in Bosnian refugees living in Croatia. *JAMA* 1999 Aug 4;282(5):433-9.

Peltzer K. Traumatic experiencing and post traumatic psychological symptoms in South African University students. Central African Journal of Medicine 1998 Nov;44(11):280-3.

Shalev AY, Freedman S, Peri T, Brandes D, Sahar T, Orr SP, Pitman RK. Prospective study of posttraumatic stress disorder and Depression following trauma. American Journal of Psychiatry 1998 May;155(5):630-7.

Trauma Research Method, First Edition. Eve B. Carlson editor. Sidran publishers, Lutherville Maryland, 1996.

Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H, et al. Cross-National Epidemiology of Major Depression and Bipolar Disorder. JAMA 1996, Volume 276(4), 24/31 July pp 293-299.

Weiss W, Bolton P, Shankar A. 'Rapid Assessment Procedures: A guide to understanding the perceived needs of refugees and displaced populations.' Johns Hopkins University, December 1999.

APPENDIX A: ASSESSMENT INSTRUMENT (ENGLISH VERSION)

Respondent ID # __ __ __

SURVEY OF POPULATIONS OF KINZENZE AND BUTAMWA CommuneS, RWANDA.

Nov-Dec 1999.

Preliminary Instructions.

Hello. My name is _____. I work for World Vision and we are conducting a study of this community to help us to better serve the people here. Would you have 30 minutes to answer some questions?

If the respondent answers no ask them if you can come back another time and interview them. If the respondent asks questions answer them, but do not provide information that could affect their answers to the questionnaire. If the respondent answers yes, then say the following:

Before we start I want to read a form to you which explains more about what we want to do.

READ CONSENT FORM HERE.

Insist that the interview be conducted in private. If this is questioned by anyone, explain that this is an important part of our confidentiality procedure, and that we have found that some people give different answers when there are other people present.

Address:

_____

Respondent ID __ __ __ __

Age: _____

Sex: M F

Number of years of School and higher education: _____

Number of years repeated_____

Interviewer: ___ ___
Date of Interview: ___ ___ / ___ ___ / ___ ___

Part A:

There are two versions of this section - one for men and one for women. If the respondent is a man, use only the version for men. If the respondent is a woman, use only the version for women.

I am going to read a list of tasks and duties. These are tasks and duties that other people around here told us were important for people to be able to do. For each task I am going to ask you how much more difficulty you are having doing it THAN A NORMAL PERSON OF YOUR AGE AND SEX. By normal, I mean a person your age and sex who has no problems. You should tell me whether you are having no more difficulty, a little more, a moderate amount more, or a lot more or you often cannot do that task.

To make it easier to remember I have a card here with pictures. Each picture represents a different amount of difficulty.

Show the respondent the card illustrating levels of difficulty. Point to each picture as you describe it.

The first picture shows someone who has no difficulty compared with normal people of their age and sex. The second picture shows someone who has a little more difficulty. The third picture shows someone who is having a moderate amount more difficulty. The fourth picture shows someone who is having a lot more difficulty. The last picture shows someone who has so much more difficulty that they often cannot do the task. For each task or duty, I will ask you to point to the picture which shows how much difficulty you are having in doing that task, compared to a normal person of your age and sex.

Now say each task, and after each one say

1. Are you having no more difficulty than normal people your age and sex, a little more, a moderate amount more, a lot more, or you often cannot do this task?

pointing to each picture as you say it.

Record the response by marking the appropriate box next to the symptom in the table below.
If the respondent indicates no more difficulty or a little more difficulty in doing a task, go to the next task and ask the same question. If the respondent indicates at least a moderate amount more difficulty, record the response and then ask the following question before going on to the next task:

2. What causes this difficulty?

In the table below record the response in the 'cause of difficulty' column beside the symptom.

Respondent ID # __ __ __

FOR MALES:

Task or function	Degree of Difficulty completing task or function None Little Moderate A lot Often Cannot					Cause of difficulty
A1 Wash	0	1	2	3	4	
A2 Dress	0	1	2	3	4	1
A3 Advise the family	0	1	2	3	4	
A4 Attend meetings	0	1	2	3	4	1
A5 Socialize	0	1	2	3	4	
A6 Manual labor	0	1	2	3	4	
A7 earning money	0	1	2	3	4	
A8 Other	0	1	2	3	4	

FOR FEMALES:

Task or function	Degree of Difficulty completing task or function None Little Moderate A lot Often Cannot					Cause of difficulty
A1 Wash	0	1	2	3	4	
A2 Dress	0	1	2	3	4	
A3 Cook	0	1	2	3	4	
A4 Wash clothes	0	1	2	3	4	
A5 Clean house	0	1	2	3	4	
A6 Care for Children	0	1	2	3	4	
A7 Attend meetings	0	1	2	3	4	
A8 Socialize	0	1	2	3	4	
A9 Transmit Culture	0	1	2	3	4	
A10 Other	0	1	2	3	4	

Part B-Depression in Adults (the Hopkins Symptom Checklist)

I am going to read you a list of problems that people sometimes have. For each one I am going to ask you how much you have experienced each one IN THE LAST WEEK, including today.

Say each symptom, and after each one ask how much it has bothered the respondent. Repeat the categories after each symptom and let the respondent choose one. Record the response by ticking the appropriate box next to the symptom.

Depression Symptoms	Not at all	A little	Quite a bit	Extremely
B1. Feeling low in energy, slowed down	1	2	3	4
B2. Blaming yourself for things	1	2	3	4
B3. Crying easily	1	2	3	4
B4. Feeling fidgety	1	2	3	4
B5. Poor appetite	1	2	3	4
B6. Difficulty falling asleep or staying asleep	1	2	3	4
B7. Feeling hopeless about the future	1	2	3	4
B8. Feeling blue	1	2	3	4
B9. Feeling lonely	1	2	3	4

B10. Thought of ending your life	1	2	3	4
B11. Feeling of being trapped or caught	1	2	3	4
B12. Worrying too much about things	1	2	3	4
B13. Feeling no interest in things	1	2	3	4
B14. Feeling everything is an effort	1	2	3	4
B15. Feeling of worthlessness	1	2	3	4
B16. Lack of trust (in others)	1	2	3	4
B17. Loss of intelligence	1	2	3	4
B18. Instability of mind	1	2	3	4
B19. Loss of sexual interest or pleasure	1	2	3	4

Part C- Self-assessment

	Yes	No
C1. **Do you think you have severe grief/sadness?**	1	2
C2. **Has anyone said that you have severe grief/sadness?**	1	2

If respondent answers yes to C1, then ask C3:

C3. **How long have you had this** *Agahinda***?**
____ days____ months____ years.

C4. **How much alcohol do you drink per day?**
____ 750ml bottles of beer

Final instructions

State that this is the end of the questionnaire. Review the questionnaire while still with the respondent, checking that all questions have been answered and the answers are clear. If not, review the question with the respondent and insert the missing information. Once you are satisfied that all questions have been answered and the responses are clear, ask the respondent if it would be OK for you, or your supervisor, to return if they have further questions. If this is OK, ask if there are any restrictions on when you might return and record these in the space below. Finally, ask the respondent not to discuss the interview with anyone until after December 15, by which time the study should be over. This is so that their comments do not affect the answers of others who are interviewed.

Restrictions on Re-visiting
_____
_____
_____

End the interview

APPENDIX B: SHORT REVIEW OF ETHNOGRAPHIC METHODS USED

Ethnographic (or qualitative) methods are structured conversations. They are methods by which interviewers gather information on a particular topic without leading the respondent. By a variety of methods the interviewer stimulates the respondent to talk on a particular topic without revealing their own opinions or world views. Much of the technique involves eliciting relevant responses using as few concepts or language as possible, thereby minimizing the influence of the interviewer's world view on the response. In the absence of leads the respondent provides their own views which are recorded as accurately as possible. The interviewer then uses ethnographic analytic methods to make sense of the data, avoiding interpretations based on their own viewpoint or understanding. This requires an appreciation of how much the interviewer's belief system affects their conversation and understanding.

Ethnographic methods are useful whenever workers are dealing with cultures or situations different from their own. They are particularly useful in the design of survey instruments. Prior use of ethnographic methods helps ensure appropriate design of questionnaires. Most questionnaires are necessarily lists of leading questions designed to acquire very specific information. Without prior ethnographic data these questionnaires may be inappropriate and responses may better reflect the designer's world view than that of the respondent. For example, a simple questions like 'what are the major health problems here?' may yield different information than that intended. If the respondent considers mental health to be a spiritual issue rather than a health issue, problems in this area will be omitted. Prior ethnographic research can provide insight into these differences and produce more appropriately-phrased questions. Different ethnographic methods exist for different purposes. Following is a short description of the ethnographic methods we used, and their rationale.

Free Listing

Here an interviewer asks a primary question designed to elicit a list. For example, in the Rwanda study the primary ethnographic question was:

What are the main problems that affect people in this Commune as a result of the genocide in 1994?

As with other ethnographic methods, the fewer concepts we introduce while still keeping the respondent on the topic the better. This question contains two major concepts; the genocide and problems.

Having put the primary question to the respondent, the interviewer lists all the responses, recording the exact terms used as much as possible. The interviewer uses non-leading probes (for example, "is there anything else), until the respondent can no

longer think of any more items. When this happens the interviewer then returns to each item and asks the secondary question. This usually consists of additional information about the item. In the Rwanda study we asked for a short description of each item, again recording the exact language used by the respondent as much as possible.

Free listing is an efficient way of getting an overview of a topic. Researchers can then choose the most relevant items in the list for further study.

Key informant interviewing proceeds much like a dialogue between informant and interviewer. Questions are even usually even more open-ended and the interviewer makes an active effort at building rapport with the informant. The interviewer can use an interview guide (a general outline of the topics to be discussed), but does not need to follow it exactly. The interviewer usually explores relevant topics as the informant brings them up during the interview. In addition, the interviewer usually interviews the same informant several times to discuss certain issues in-depth, and to further develop rapport. As with other ethnographic methods, the interviewer must keep the interviewee talking around the topic of interest without imposing his own belief system.

When key informant interviews are the first (or only) ethnographic method used, interviewers often open the first interview with very general questions which may not even be about the topic (for example, 'Tell me about your day'). When the informant mentions something related to the topic of interest, the interviewer then asks him/her to talk more about that topic, referring to it using the local term used by the interviewee. As the interview (and interviews) progress, the interviewer can use his/her expanding knowledge of the local belief system to focus on the topic of interest. In the Rwanda study we used free listing first, and used the information from the listing to begin with more specific questions. We used symptoms mentioned in the free list to create a scenario which was then used to stimulate responses. We were careful in the scenario to use only symptoms and language that respondents used in the free lists, to avoid injecting our own ideas or views.

Pile Sorts

Pile Sorts require a set of cards, each representing a single item. For example we were interested in studying symptoms, so each card represented a different symptom. Usually cards represent the item visually (for the illiterate) and in writing. Respondents are given the cards and asked to sort them into groups (or piles) based on any criteria they like. The only restrictions are that the criteria refer to the item and not the card (for example, respondents should not group cards based on how well the drawings are executed) and the cards are not sorted into a single pile or each card by itself. After the respondent has sorted the cards, the interviewer asks for each pile why those cards were put together. The interviewer then records the piles and the reasons given.

Analysis consists of recording how many times, for all respondents collectively, each possible pair of cards were occurred. Items which are put together most frequently are considered related and the reasons reviewed to decide why this might be so.

APPENDIX C: VERBAL CONSENT FORM

The Johns Hopkins University
School of Hygiene and Public Health
Committee on Human Research

Verbal Consent Form for Research Study.

Submission Date: _____

Instructions for the Interviewer:
The following is to be read to the subject prior to the interview. If the subject then agrees to participate, you must sign on the line marked 'Witness to Consent Procedures', at the end of this form. Also mark the date on the appropriate line.

Study of health effects due to displacement by war.

Purpose of the Study

You are being asked to be part of a research study. We want to find out how the events of 1994 in Rwanda has affected the health of the people and how many people have been affected. This research is being done by Johns Hopkins University and World Vision.

Procedures

To obtain this information we are talking with some people in the community. From a list of all the houses in the community we selected your house by chance. You were also selected by chance from the people living in this house. If you agree to help us, you will be asked some questions. These questions are about your health. We may also want to return and talk with you again.

Risks and Discomforts

Each interview will take about 30 minutes. It is possible that some questions may upset you. You may refuse to answer these questions, or any questions, if you wish. You may stop the interview at any time.

Benefits

This information will help World Vision to provide better programs to improve the health of the people in this area. However, there may be no direct benefit to you if you do not have health problems.

Confidentiality

During the interview I will write down the information you tell me. This is the information we will use for our study. The record of this information will not have any information which can be used to identify you. I will also record your name and address, but this will be stored separately from the record, and will be placed in a special cabinet in the project director's office. This cabinet will be locked and only he will have the key. Only he and the researcher from Johns Hopkins University will be able to see this information. Every effort will be made to protect the confidentiality of this information as far as is legally possible.

Voluntariness

It is your decision whether or not to be in this study. You can stop being in this study at any time. This will not affect any assistance you get from World Vision or any other organization.

Whom to Contact

If you have any questions you can ask Dr. Paul Bolton of the Johns Hopkins University. He is in charge of the study. You can also contact Lincoln Ndogoni. Both can be contacted through the World Vision Office in Kigali, telephone 85481. In the future if you have any questions about the study, you should ask Lincoln Ndogoni. He and the other researchers will tell you if they learn anything new that they think will affect you.

Do you agree to participate in this study?

_____
Witness to Consent Procedures (to be signed by interviewer after subject has verbally consented).

_____
Signature of Investigator

_____
Date

APPENDIX D: SUGGESTED TIME-LINE FOR SUBSEQUENT STUDIES

This preliminary time-line is based on our experiences with our field trial in Rwanda. The time allocations given here assume 8 data gatherers working in two areas with two vehicles. Time taken for data gathering will be shorter with more workers and/or vehicles.

Prior to study

Select community to be studied.

Select NGO workers to work on project and give them background information.

Conduct preliminary discussions with community leaders and representatives and obtain permission to conduct study.

Prepare logistics

Identify community members knowledgeable about community problems.

Map community as a guide for the study and for sampling.

Day 1

Meet with NGO team and provide project overview and address questions.

Director meets local leaders.

Conduct 'town hall' meeting with community if possible.

NGO workers appoint to meet with knowledgeable persons the following afternoon.

Train NGO workers in research principles. Provide general overview of qualitative and quantitative methods.

Day 2

Train NGO workers in free listing method

Conduct 40 free lists with knowledgeable community members of the major problems affecting the community, and who people consult for mental or emotional problems.

Day 3

Complete first free list if necessary

Conduct 40 more free lists on a convenience sample of the most important tasks people must do regularly.

Make appointments to interview in 2 days persons who are consulted for mental and emotional problems.

Day 4

Complete free listing if necessary.

Make appointments for interviews the following day, if not arranged the previous day

Analyze free list data and identify target mental and emotional problems, and important tasks.

Commence training in key informant interviews.

Day 5

Complete training in key informant interviews.
Proceed to field and interview persons consulted about mental and emotional problems, about the mental problems identified in the free listing exercise and related problems.
Briefly write up interview

Day 5

Review results of previous day's interviews and provide feedback for next interview.
Conduct second interview with same informant about same topic.
Briefly write up interview.

Day 6

Collectively analyze results of key informant interviews.
Provide training in Pile sorts.
Use existing data to create pile sorts cards and practice.

Day 7

Conduct at least 40 pile sorts on a convenience sample in the community.

Day 8

Analyze pile sort data and review other ethnographic data
Adapt instrument data, including creation of functional assessment questionnaire.

Days 9 and 10

Group translation of instrument by translators (drawn from local population if possible).
Back-translation by separate individual
All translators meet with project director to resolve differences between translation and back-translation, with reference to language of respondents in the ethnographic study.

Day 11

Review translated instrument with NGO staff and incorporate suggestions.
Train NGO staff who worked in the ethnographic study as survey supervisors
Interview and hire interviewers for survey.

Day 12

Didactic training in interviewing for supervisors and interviewers.
Supervisors and interviewers review translated instrument and incorporate suggestions.
Practice interviewing using instrument.

Day 13

Continue interview practice and provide feedback
Finalize logistics for field work and transport
Conduct pilot study in the community- 2 interviews each

Day 14

Complete pilot study if necessary. Meet and review pilot study experience.
Review pilot questionnaires and test data entry procedures.
Provide feedback on basis of review of pilot interviews
Adapt instrument in light of pilot study experiences.

Day 15

Commence validation study: Supervisors identify 50 persons said to have and 50 said to not have a local mental illness equivalent to the illness being studied (Depression).
Conduct blinded interviews of these persons..

Day 16

Continue validation study interviews

Day 17

Complete validation study interviews.
Meet and review experiences.
Make last minute changes to instrument in the light of the validation study.
Analyze validation study data.

Day 18

Didactic training on sampling method, including selecting houses and respondents.
Allocated areas of responsibility to supervisors for division among their team.
Finalize logistics for survey, including transport.
Commence interviewing

Days 19-23

Continue interviewing.
Supervisors and interviewers, and supervisors and director meet every day to monitor interview quality and discuss any problems.

Day 24

Cease interviewing
Commence data analysis.

Post-survey

Complete data analysis and report.
Share results with community and incorporate into programming.

APPENDIX E: DSM-IV CRITERIA FOR MAJOR DEPRESSIVE EPISODE

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.

1) depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful)

2) markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)

3) significant weight loss when not dieting or weight gain (e.g., a change of more_than 5% of body weight in a month), or decrease or increase in appetite nearly every day.

4) insomnia or hypersomnia nearly every day

5) psychomotor agitation or retardation early every day (observable by others, not merely subjective feelings of restlessness or being slowed down)

6) fatigue or loss of energy nearly every day

7) feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)

8) diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)

9) recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

B. The symptoms do not meet criteria for a Mixed Episode

C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism).

E. The symptoms are not better accounted for by Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.

APPENDIX F: DSM-BASED ALGORITHM FOR DIAGNOSING Depression WITH THE HSCL.

The original 25 question version of the Hopkins Symptom Checklist (HSCL) measures both anxiety and Depression. We used only the questions assessing Depression. We chose the HSCL because of its history of successful use across many situations and cultures. Since it was created prior to the DSM Depression criteria, it is not entirely consistent with the DSM 'A' criteria for Depression. To improve consistency with DSM we dropped one question from the analysis ('feeling trapped') and added a question on psychomotor agitation. To diagnose Depression with the HSCL we used an algorithm based on the DSM criteria. This algorithm was developed and tested by the Harvard Program in Refugee Trauma, and is shown in Table 14.

Table 14: Algorithm for Matching HSCL Depression questions to DSM Criteria for Major Depression.

DSM-IV Criteria	Comment	Adapted HSCL Question
A1. Depressed mood	Must have at least one of these (and only one counts) or one of A2.	B3 crying easily B7 feeling hopeless B8 feeling blue B9 feeling lonely
A2. Diminished interest or pleasure	Must have at least one of these (and only one counts) or one of A1.	B13 loss of interest B19 loss of sexual pleasure or interest
A3. Significant weight loss or change in appetite		B5 poor appetite
A4. Insomnia or hypersomnia		B6 difficulty sleeping
A5. Psychomotor agitation		B4. feeling fidgety
A6. Fatigue or loss of energy	Count only one of these	B1 feeling low in energy B14 feeling everything is an effort
A7. Feeling worthless or guilty	Count only one of these	B2 blaming yourself for things B15 feeling worthless
A8. Diminished ability to think or concentrate		B12 worrying too much
A9. Recurrent thoughts of death		B10 thought of suicide
B. Do not meet criteria for a mixed episode		
C. Clinically significant distress or functional impairment		C1 presence of severe grief significant functional disability

D. Symptoms not due to substance abuse or medical condition		
E. Symptoms not due to bereavement, last more than 2 months, or involve marked functional impairment, feeling worthless, suicidal or psychotic symptoms or psychomotor retardation.		C2 duration of severe grief significant functional disability B15 feeling worthless B10 thoughts of suicide

Under DSM criteria a Depression diagnosis requires the presence of 5 or more of the A criteria during the same 2 week period and represent a change from previous functioning. At least one of the symptoms must be A1 or A2. Criteria for a mixed episode (B) were not assessed, nor were substance abuse (apart from alcohol) or medical conditions which might cause these symptoms (D). Significant functional disability was assessed using a dichotomous 'function' variable (see Analysis in the Methods section for how this variable was created).

APPENDIX G: ROC CURVES FOR Depression SCORES *Vs* DICHOTOMOUS FUNCTIONAL DISABILITY AND Depression DIAGNOSIS

Figure 2: ROC Curve for Depression Score *vs* Dichotomous Functional Disability Variable.

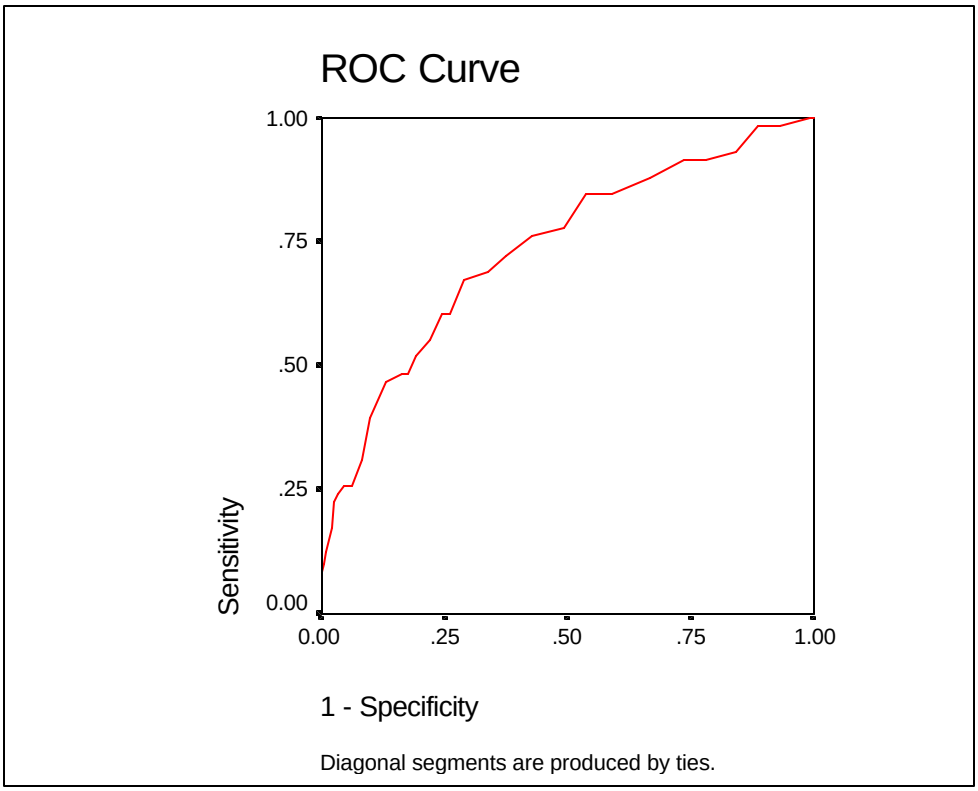


Table 15: Coordinate Points on ROC Curve of Depression Score *vs* Dichotomous Functional Disability Variable.

Positive if Greater Than or Equal To	Sensitivity	1 - Specificity
13.0000	1.000	1.000
15.5000	1.000	.997
17.5000	1.000	.994
18.5000	.983	.932
19.5000	.983	.886
20.5000	.933	.841
21.5000	.917	.776
22.5000	.917	.734
23.5000	.883	.662
24.5000	.850	.588
25.5000	.850	.536
26.5000	.783	.487
27.5000	.767	.422
28.5000	.733	.367
29.5000	.700	.331
30.5000	.683	.282
31.5000	.617	.253
32.5000	.617	.237
33.5000	.550	.214
34.5000	.517	.188
35.5000	.483	.169
36.5000	.483	.159
37.5000	.467	.130
38.5000	.400	.97
39.5000	.317	.081
40.5000	.267	.058
41.5000	.267	.042
42.5000	.250	.029
43.5000	.233	.023
44.5000	.183	.016
45.5000	.117	.006
46.5000	.100	.003
47.5000	.083	.000
50.0000	.067	.000
53.0000	.050	.000
54.5000	.033	.000
56.5000	.017	.000
59.0000	.000	.000

Figure 3: ROC Curve of Depression Score *vs* Depression Diagnosis by Algorithm

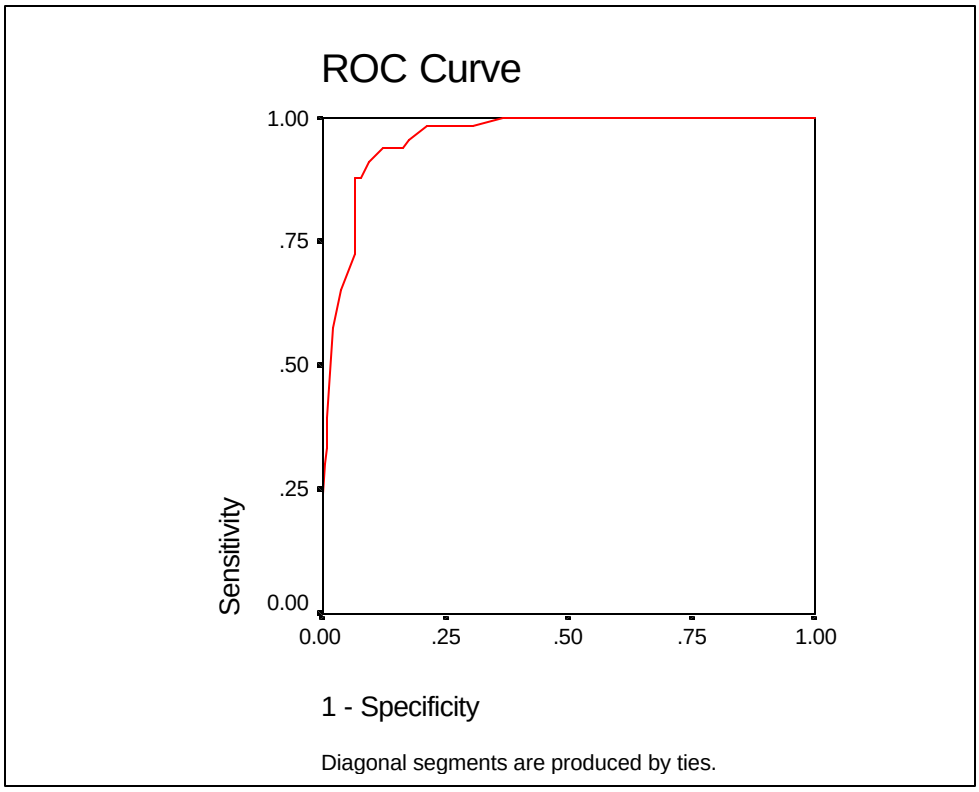


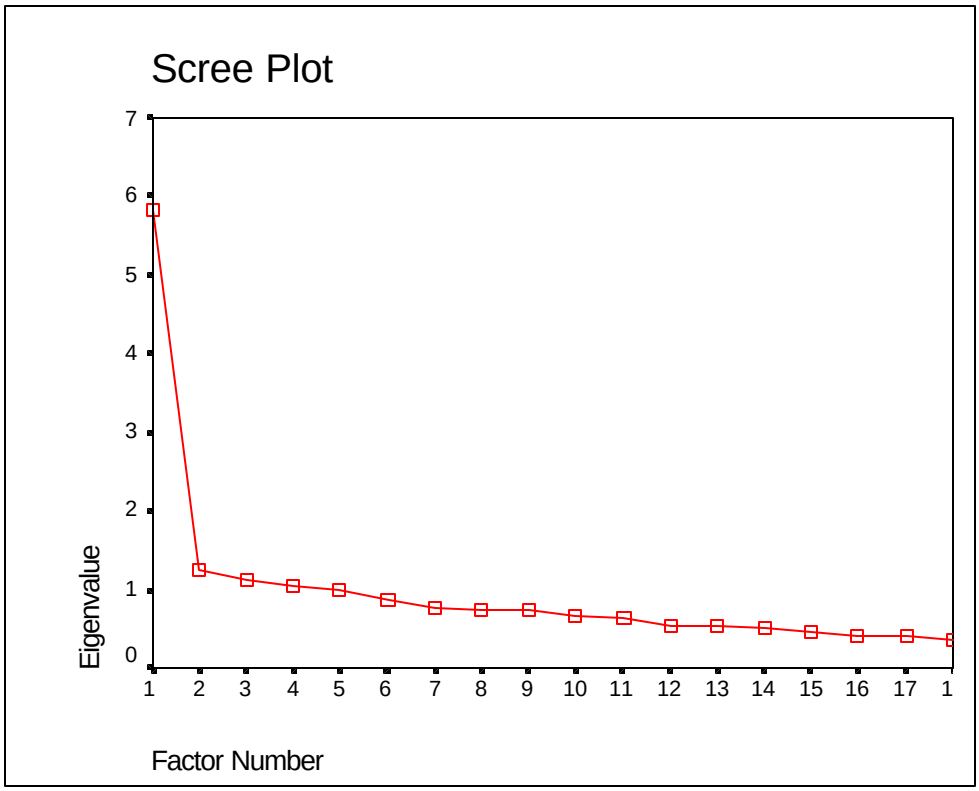
Table 16: Coordinate Points of ROC Curve of Depression Score *vs* Depression Diagnosis by Algorithm.

Positive if Greater Than or Equal To	Sensitivity	1 - Specificity
13.0000	1.000	1.000
15.5000	1.000	.997
17.5000	1.000	.993
18.5000	1.000	.927
19.5000	1.000	.881
20.5000	1.000	.825
21.5000	1.000	.755
22.5000	1.000	.712
23.5000	1.000	.632
24.5000	1.000	.550
25.5000	1.000	.497
26.5000	1.000	.434
27.5000	1.000	.364
28.5000	.985	.305
29.5000	.985	.262
30.5000	.985	.209
31.5000	.955	.172
32.5000	.939	.159
33.5000	.939	.123
34.5000	.909	.096
35.5000	.879	.076
36.5000	.879	.066
37.5000	.727	.066
38.5000	.652	.036
39.5000	.576	.020
40.5000	.455	.013
41.5000	.394	.010
42.5000	.333	.007
43.5000	.303	.003
44.5000	.242	.000
45.5000	.136	.000
46.5000	.106	.000
47.5000	.076	.000
50.0000	.061	.000
53.0000	.045	.000
54.5000	.030	.000
56.5000	.015	.000
59.0000	.000	.000

APPENDIX H: TESTING CONSTRUCT VALIDITY USING FACTOR ANALYSIS OF SURVEY DATA.

We studied the construct validity of the Depression questions using factor analysis with principal components extraction and varimax rotation. Scree plots were used to determine the number of factors to extract. The scree plot is shown in Figure 4.

Figure 4: Scree plot of factor analysis of Depression questions with principal factor extraction.



Although 18 factors were found only one factor is above the elbow of the plot, suggesting that a single factor underlies much of the variation in the Depression questions. In fact, this factor accounted for 32% of the total variance of the 18 Depression questions, compared with 7% for the second most significant factor. Having determined from the plot how many factors to retain, we then reran the analysis, specifying a single factor. The results of this analysis are shown in table 17.

Table 17: Principal component matrix of Depression Questions

Depression Question	Correlation between question and factor
B1Lack of Energy	0.615
B2Blaming self too much	0.392
B3Crying easily	0.496
B4Feeling fidgety	0.690
B5Loss of appetite	0.567
B6Sleeping poorly	0.609
B7Loss of hope	0.581
B8Feeling blue	0.759
B9Feeling lonely	0.714
B10Thoughts of suicide	0.393
B12worrying too much	0.523
B13no interest in things	0.270
B14Tasks require more effort	0.674
B15Feeling worthless	0.666
B16Lack of trust in others	0.444
B17Loss of intelligence	0.585
B18Mental instability	0.653
B19Loss of interest in sex	0.296

Given the standard cut-off of 0.3 to decide whether a question loads on the factor, all questions load adequately except for lack of interest in general and lack of interest in sex. Suicide loads, although not very highly, which may reflect lack of variation in the responses to this question.
The very simple structure (a single factor) resulting from this analysis suggests that there is a single concept being measured by the Depression questions. This supports the hypothesis that Depression exists in this population, and the construct validity of the questionnaire.

A second factor analysis was done, using all the factors with eigenvalues greater than one. This was done to see if lack of interest and lack of interest in sex loaded on any other factors, even though these secondary factors had eigenvalues close to one and so accounted for little variance. Both questions failed to load on any of these factors. The item analysis also suggested that these questions may not be important in diagnosing Depression - the cronbach's alpha increased when these questions were

removed (in fact diagnosis of DSM Depression depends on the interest item in only one respondent). However, given the content validity of the interest item as suggested in the ethnographic study and by the DSM criteria, that the cronbach's alpha scores were only slightly reduced by removing these questions, and the possibility that the poor loading of these factors might be due to peculiarities of the sample, we decided to retain them until further studies using this instrument provide additional evidence.

APPENDIX I: STRATIFIED LOGISTIC REGRESSION ANALYSES

This appendix contains logistic regression analyses of selected variables against a dichotomous functional disability variable. This variable is described in the Results - functional disability scores less than 26 were classed as not significantly disabled, and scores of greater than 25 were classed as significantly functionally disabled. In those analyses in which sex is a variable, the Beta coefficient and Odds Ratio refer to being male compared with females.

Table 18: Males

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Age	.0640	.019	.001	1.066
Education	.051	.103	.618	1.053
Agahinda	.993	.607	.102	2.70
Depression	1.480	.677	.029	4.39

Table 19: Females

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Age	.046	.014	.001	1.047
Education	-.056	.080	.482	.946
Agahinda	.180	.469	.701	1.20
Depression	1.448	.471	.002	4.26

Table 20: Respondents with 3 years of education or less

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Age	.048	.012	.000	1.049
Sex	-.133	.453	.769	.875
Agahinda	.670	.446	.133	1.96
Depression	.980	.484	.043	2.67

Table 21: Respondents with more than 3 years of education

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Age	.057	.020	.005	1.059
Sex	.869	.554	.116	2.39
Agahinda	.066	.641	.918	1.07
Depression	2.156	.644	.001	8.64

Table 22: Respondents aged 27 years or less

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Sex	-.304	.730	.678	.738
Education	.002	.129	.988	1.002
Agahinda	.838	.751	.265	2.31
Depression	.967	.794	.224	2.63

Table 23: Respondents older than 27 years

Variable	Beta Coefficient	Std. Error	Sig.	Odds Ratio
Sex	.727	.375	.053	2.07
Education	-.121	.066	.066	.886
Agahinda	.419	.407	.303	1.52
Depression	1.672	.422	.000	5.32