Group Interpersonal Psychotherapy for Depression in Rural Uganda

A Randomized Controlled Trial

Paul Bolton, MBBS
Judith Bass, MPH
Richard Neugebauer, PhD, MPH
Helen Verdeli, PhD
Kathleen F. Clougherty, MSW
Priya Wickramaratne, PhD
Liesbeth Speelman, MA
Lincoln Ndogoni, MA
Myrna Weissman, PhD

Context Despite the importance of mental illness in Africa, few controlled intervention trials related to this problem have been published.

Objectives To test the efficacy of group interpersonal psychotherapy in alleviating depression and dysfunction and to evaluate the feasibility of conducting controlled trials in Africa.

Design, Setting, and Participants For this cluster randomized, controlled clinical trial (February-June 2002), 30 villages in the Masaka and Rakai districts of rural Uganda were selected using a random procedure; 15 were then randomly assigned for studying men and 15 for women. In each village, adult men or women believed by themselves and other villagers to have depressionlike illness were interviewed using a locally adapted Hopkins Symptom Checklist and an instrument assessing function. Based on these interviews, lists were created for each village totaling 341 men and women who met *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*) criteria for major depression or subsyndromal depression. Interviewers revisited them in order of decreasing symptom severity until they had 8 to 12 persons per village, totaling 284. Of these, 248 agreed to be in the trial and 9 refused; the remainder died or relocated. A total of 108 men and 116 women completed the study and were reinterviewed.

Intervention Eight of the 15 male villages and 7 of the 15 female villages were randomly assigned to the intervention arm and the remainder to the control arm. The intervention villages received group interpersonal psychotherapy for depression as weekly 90-minute sessions for 16 weeks.

Main Outcome Measures Depression and dysfunction severity scores on scales adapted and validated for local use; proportion of persons meeting *DSM-IV* major depression diagnostic criteria.

Results Mean reduction in depression severity was 17.47 points for intervention groups and 3.55 points for controls ($P<.001$). Mean reduction in dysfunction was 8.08 and 3.76 points, respectively ($P<.001$). After intervention, 6.5% and 54.7% of the intervention and control groups, respectively, met the criteria for major depression ($P<.001$) compared with 86% and 94%, respectively, prior to intervention ($P=.04$). The odds of postintervention depression among controls was 17.31 (95% confidence interval, 7.63-39.27) compared with the odds among intervention groups. Results from intention-to-treat analyses remained statistically significant.

Conclusions Group interpersonal psychotherapy was highly efficacious in reducing depression and dysfunction. A clinical trial proved feasible in the local setting. Both findings should encourage similar trials in similar settings in Africa and beyond.

JAMA. 2003;289:3117-3124 www.jama.com

DEPRESSION IS A LEADING cause of disability in both developed and developing regions of the world, including Africa.[1,2] In 2000, we conducted a community-based survey in an impoverished part of southwest Uganda that has been severely affected by the human immunodeficiency virus (HIV) epidemic. Using *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*) major depression criteria, we found a current depression prevalence rate of 21% (P.B., unpublished data, 2000), consistent with previous research implicating socioeconomic disadvantage and bereavement in depressive symptoms.

World Vision International, a nongovernmental humanitarian organization, was interested in addressing this substantial mental health burden in Uganda. Both antidepressants and psychotherapy have been shown to be efficacious in numerous controlled trials in developed countries, including evidence of equivalence in reducing the symptoms of acute depression.[3] However, use of antidepressants is not feasible in this region because of high cost

Author Affiliations are listed at the end of this article.
Corresponding Author and Reprints: Paul Bolton, MBBS, 159 Tilden Rd, Scituate, MA 02066 (e-mail: pbolton@jhsph.edu).

and limited supply infrastructure. Psychotherapy was therefore the preferred option, although its use raised other issues. While there is substantial evidence for the efficacy of "talking therapies,"[4] these have been developed in industrialized nations in the Western Hemisphere. The extent to which the concepts and therapeutic strategies they use are appropriate among other populations is unknown. In sub-Saharan Africa, conditions are very different from those in which psychotherapy was developed, in ways that could reduce effectiveness. For example, many populations are reluctant to communicate directly about sensitive issues; others live in conditions of extreme chronic deprivation that are rare in developed countries.

The need to test the local effectiveness of psychotherapy raised an additional problem. Such testing has been hampered in Africa by a lack of field methods for cross-cultural adaptation and validation of assessment instruments. The lack of these methods, as well as perceived logistic and ethical difficulties, have led some to believe that clinical trials of psychotherapy are not feasible in Africa. We therefore began by developing a field method that has since been successfully tested in 2 sites—Rwanda and the same villages in Uganda as in the current study.[5,6] In both settings, we created or modified and then validated measures of depressive symptoms and social functioning. These instruments were then used in community-based prevalence surveys. The instruments developed in Uganda form the basis of the current study.

The intervention we studied is a group-based interpersonal psychotherapy (IPT) for depression. Extensive evidence for its efficacy and effectiveness comes from randomized controlled clinical trials in which treatment was time-limited and specified in a procedural manual.[4] "Time-limited" means that treatment is not open-ended but that the number, frequency, and duration of sessions are specified at treatment outset. Selection of this intervention allowed us to more accurately budget the intervention and also made it cost-effective compared with open-ended therapies. The IPT manual (available by e-mail from the authors at mmw3@columbia.edu or kfclougherty@aol.com)[7] was essential for accurate provision of IPT to this population. Prior experience also suggested that the focus of IPT on interpersonal relationships was compatible with Ugandan culture. The full rationale behind the development of IPT, its adaptation for use in Uganda, and the training of local care providers is described elsewhere.[4,7]

This article reports the results of a controlled clinical trial of group IPT. The study was conducted in the same Ugandan villages surveyed in 2000. Screening and baseline assessments were conducted in February 2002. The IPT took place from March through June 2002 (all groups began and finished within a week of each other), and the follow-up assessment was conducted within 2 weeks of IPT completion.

Our purposes were (1) to test the efficacy of group IPT for Uganda (IPT-G-U) in relieving depressive symptoms and improving functioning and (2) to evaluate the feasibility of such studies in sub-Saharan Africa. To our knowledge, this is the first published controlled clinical trial of a psychological intervention in resource-poor sub-Saharan Africa.

METHODS

Site and Population

The study area included all of Rakai province and the contiguous half of Masaka province in southwest Uganda, the area of operations for World Vision International. There are 154 villages in this area, each with several hundred to several thousand adults and separated from other villages by surrounding agricultural fields. Because of prior agreements with local leaders, we studied the same 30 villages assessed in the 2000 prevalence survey. These were originally chosen by weighted random sampling based on government census data. Hence, sample size was based on these agreements rather than on sample size calculations. However, calculations were later performed using a formula that accounts for individual clustering within villages: $m = 2(z_{\forall} + z_{\exists})^2 \Phi^2 (1 + (n - 1)\Delta)/nd^2$ where m is the number of clusters (villages) in each study arm; n is the number of individuals per group; Φ^2 is the variance of the change score for each individual; Δ is the correlation of change scores between any 2 individuals in the same village; and d^2 is the square of the expected difference between baseline and follow-up for controls minus the expected difference for the intervention groups. These calculations suggested that 30 is an appropriate sample size, assuming a 50% difference in symptom severity scores between intervention and control groups after the intervention, using a 2-tailed test with $\alpha = .05$ and 80% power.

Villages formed the unit of randomization, each with 1 IPT or control group. Because sex-segregated groups were more culturally acceptable, villages were randomly sorted into blocks of 15 villages for male participants and 15 villages for female participants. Within each block, each village was randomly allocated to the intervention or control study arm to ensure a sex-balanced design. Study participants in the control villages did not receive IPT; however, participants in both control and intervention villages were free to seek whatever other interventions they wished throughout the study. Prior to randomization, all potential participants were informed that if the intervention proved effective, it would later be offered to controls (currently being implemented by World Vision International).

Assessment of Outcome

The outcome measures were originally created for the 2000 study. They include the depression section of the Hopkins Symptom Checklist,[8] adapted and validated for local use by a combination of ethnographic and quantitative methods,[6] and a sex-specific 9-item questionnaire to assess functional impairment.

This latter measure was also developed and validated locally and is based on tasks that participants in the ethnographic study reported were important to local people. For each task, respondents were asked to state whether they were experiencing "no more" difficulty (scored as 0), "a little more" difficulty (score, 1), "a moderate amount more" difficulty (score, 2), "a lot more" difficulty (score, 3), or were "frequently unable to do the task" (score, 4) compared with most others of their age and sex. Overall severity of dysfunction was calculated for each individual by summing the scores for all tasks. A similar process was used to calculate overall depression severity using the responses to the Hopkins Symptom Checklist (details of the development and validation of the depression and function questionnaires have been published previously[5,6]). Questions on demographics and duration of symptoms and 12 questions on HIV-related knowledge, behavior, and attitude were added to form the final study instrument. The latter were included to investigate the link between depression and HIV, a topic outside the scope of this article.

To assess presence or absence of major depression, we used a *DSM-IV*[9]–based algorithm originally developed by Mollica et al[10] and expanded to include the *DSM-IV* A, C, and E diagnostic criteria. This algorithm includes the number and duration of symptoms and dysfunction. The *DSM-IV* B and D criteria were not included because they require exclusion of medical conditions and drug effects, which was beyond our resources. The diagnostic algorithm is described in detail elsewhere.[11]

Study Eligibility and Screening

Eligible persons were identified using a 3-stage screening process performed by 10 locally based World Vision International staff who had a minimum education status of high school graduate. Nine had been interviewers for the 2000 study; the other had previous survey experience.

Stage 1 screening involved contacting local leaders, healers, or other knowledgeable persons in each village. They were asked for a list of at least 20 men or women (according to the village's sex assignment) aged at least 18 years who were believed by other villagers to have depressionlike illness. In the local Luganda language, there is no single term to describe depression. Instead the interviewers asked for persons with *Yo'kwekyawa* or *Okwekubazida*, 2 locally described depressionlike syndromes identified in the 2000 study (P.B., unpublished data, 2000) These syndromes are frequently comorbid and together include all the *DSM-IV* major depression symptom criteria.

For stage 2 screening, interviewers visited each person on the village list. Informed consent was obtained by reading the consent form to the participant because illiteracy is common in this area. If the respondent agreed to participate, the interviewer signed the form as a witness. They then asked the respondent if they thought that they had *Yo'kwekyawa* and/or *Okwekubazida*. If the person denied having either syndrome, they were not interviewed further. The stage 1 and 2 screenings were performed because the likelihood of depression was approximately 60% when both the individual and others believed they had either of the local syndromes, according to the 2000 study data (P.B., unpublished data, 2000). If the person thought that he/she currently had *Yo'kwekyawa* and/or *Okwekubazida*, the participant was interviewed using the study instrument, which constituted the third stage of screening. Because each local syndrome is only an approximation of depressive illness, self-report and outside reports of their presence were used in screening only and not as outcome measures.

The purpose of this process was to develop lists of approximately 12 eligible persons per village—those who believed themselves and were believed by others to have depressionlike illness and who at interview met the *DSM-IV* criteria according to the algorithm. However, unanticipated problems arose. First, in only 5 villages did at least 12 persons meet the algorithm diagnostic criteria. Second, there was a very wide age range among eligible persons in some villages; this was a problem because experience suggests that IPT-G-U is more effective among participants of similar age. Third, some interviewees stated that they were currently bothered "quite a bit" or "extremely" by thoughts of suicide and, hence, could not be ethically enrolled in a trial of an (as yet) unproven intervention.

The eligibility criteria were therefore revised: (1) diagnostic criteria were expanded to include persons with subthreshold depression, meaning that they fell short of a major depression diagnosis by a single *DSM-IV* symptom criterion; (2) in villages where 1 eligible person was of a very different age than the rest, that person was excluded; and (3) persons who appeared to be currently suicidal were revisited by the interviewer. If the person confirmed a current danger of suicide, he/she was not enrolled but referred to a clinical psychologist (L.N.) for further attention. In villages where more than 12 interviewees met these criteria, we chose the 12 with the highest depression scores.

Randomization and Participation

After creation of the lists of eligible participants, 8 of the 15 male villages and 7 of the 15 female villages were randomly assigned to the intervention arm and the remainder to the control arm. Random assignment was performed by enumerating the villages and using a random-number table to determine study allocation. Each list began with those who met the original diagnostic criteria, followed by those who fell short by a single criterion, in order of decreasing depression score. Interviewers visited each person in the order they appeared on the list. The interviewer reread the consent form, advised the person about the study group to which their village had been allocated, and asked them to confirm their willingness to continue in the study. Interviewers continued down the list until they had at least 8 participants (at which point they did not contact the remainder of the list) or until they reached the end of the list. The target group size was between 8 and 10, based on the clini-

cal judgment of the 2 senior IPT therapists (H.V. and K.F.C.).

Timing of the Follow-up Assessment and Interviewer Blinding

Both the intervention and control groups were reinterviewed with the study instrument within 2 weeks of the intervention groups' completion of IPT. Because interviewers likely knew which villages had received the intervention, all interviewees were taken to the same town and interviewed there. They were instructed not to mention their village during the interview or whether they had received IPT. Each was interviewed by the same interviewer as at baseline to reduce interviewer effects. The study was approved by the Johns Hopkins University Institutional Review Board and by local government authorities in Rakai and Masaka districts.

The IPT-G-U Intervention

Interpersonal psychotherapy was originally developed for individual treatment of major depression. Since then, it has been found effective for a variety of mood and nonmood disorders in numerous clinical trials.[4] A group adaptation of IPT was devised for treatment of binge eating disorder, social phobias, adolescent depression, and posttraumatic stress disorder. Interpersonal psychotherapy makes no assumption about etiology but uses as a critical point for intervention the connection between symptom onset and current interpersonal problems. The therapist begins with a careful diagnostic assessment, then explains the diagnosis and works with the patient to identify the problem areas associated with the onset of current symptoms. Difficulties in 4 interpersonal areas are considered triggers of depressive episodes, and 1 or more of these form the treatment focus: grief (due to death of a loved one), interpersonal disputes (disagreements with important people in one's life), role transitions (changes in life circumstances), and deficits (persistent problems in initiating or sustaining relationships).

In Uganda, each IPT group met for 90 minutes weekly for 16 weeks. Groups were led by a local person of the same sex as the group who had received 2 weeks of intensive training in IPT-G-U by 2 of the authors (H.V. and K.F.C.). During each session, the group leader reviewed each participant's depressive symptoms. The participant was then encouraged to describe the past week's events and to link those events to his/her mood. The group leader then facilitated support and suggestions for change from other group members. Attendance was high; the dropout rate was 7.8%, and 54% of participants attended at least 14 sessions. Details of IPT and its adaptation for Uganda have been described elsewhere.[4,7]

Statistical Methods

Baseline characteristics of intervention recipients and controls were compared using χ^2 tests for categorical data and *t* tests for continuous data. The main indicators of effect derive from a comparison of the mean changes in the depression and function scores from baseline to postintervention assessment in the intervention groups with those in the control groups. Change in depression prevalence in both study arms was also assessed using a dichotomous variable (1=depression present and 0=depression absent according to the algorithm).

Three sets of analyses were performed using different categories of participants. The first analysis included only persons interviewed before and after the intervention; we excluded those on the village lists whom we tried but failed to *re*contact for inclusion in the trial, those whom we found but who refused, and those whom we found and who agreed but were lost to follow-up. The second analysis included all of these persons. Therefore, this analysis consisted of all of those included in the lists of eligible persons who were then sought by interviewers for inclusion in the trial, whether or not they were found. The third analysis included all of those on the lists of eligible participants, whether or not we tried to *re*contact them. Hence, the second and third analyses were on an intention-to-treat basis to allow for possible selection bias by participants or interviewers. For participants whom we failed to recontact and/or reinterview, we used an end-point analysis, a method widely used in clinical trials of major depression disorder. The initial interview was carried forward and imputed as the end-of-treatment score, thereby assuming no change over the course of intervention for these individuals.

Since the unit of analysis is the participant and the study design is a cluster randomized trial, we used a 2-level model to allow for within-village (cluster) correlation and between-village variability.[12] It is also necessary to adjust for potential nonindependence of outcomes of the individuals within each village. For continuous outcome measures such as change in depression scores, we used mixed-effects models to determine the intervention effect. In these models, treatment status was regarded as a fixed effect and the effect of villages was treated as random. Mixed models were also used to determine whether variation among group leaders was associated with the change in depression scores. For dichotomous outcome variables such as prevalence of depression, logistic regression models based on the generalized estimating equations with robust variance estimates[13] were used to adjust for potential correlation of outcomes within a village. Odds ratios and their corresponding 95% confidence intervals (CIs) were computed for categorical outcomes and regression coefficients and corresponding *P* values were computed for continuous outcomes. The level of statistical significance was set at $P<.05$. We used Stata version 7.0 (Stata Corp, College Station, Tex) for the descriptive analyses and SAS version 8.0 (SAS Institute Inc, Cary, NC) for the cluster analyses.

RESULTS

In the 30 study villages, 631 people were identified by themselves and others as having a depressionlike illness and consented to be interviewed. Of

these, 341 met the revised eligibility criteria, of whom 271 were diagnosed as having major depression using the *DSM-IV* algorithm. A total of 6 men and 7 women expressed suicidal ideation at the initial interview. However, on further assessment, none were determined to be actively suicidal and all were included in the study.

As described in the "Methods" section, in each intervention village interviewers visited eligible persons in the order they were listed, inviting them to join the village IPT group. When at least 8 persons had been recruited, those remaining on the list were not approached (in some villages, interviewers reached the end of the list before recruiting 8). In this way, interviewers attempted to contact 139 of the 163 eligible persons on the intervention village lists. Of these, 9 declined, 12 could not be found, and 2 had died since the screening interview, leaving 116 enlisted. Nine of these 116 were lost to follow-up for various reasons during the intervention, leaving 107 (53 men and 54 women) who completed the intervention and were reinterviewed at follow-up. The FIGURE presents a flow diagram of the study participants.

Similarly, we attempted to contact 145 of the 178 eligible persons on the control village lists. Two had died since screening and 11 were unavailable, leaving 132 who agreed to continue in the study (in some control villages, >8 persons were recruited to compensate for recruiting <8 persons in other control villages). Of these 132, 1 died during the intervention and 14 could not be found, leaving 117 controls (55 men and 62 women) who were reinterviewed (Figure).

Baseline Characteristics

Intervention and control groups did not differ significantly by age, symptom duration, years of education, or baseline depression or function scale scores. However, there was a significant difference in the proportions who met the original depression diagnostic criteria, both among those who completed the study and all those on the original lists of eligible participants (TABLE 1). (Tests for differences in baseline characteristics were performed using standard signifi-

Figure. Flow of Participants Through the Trial

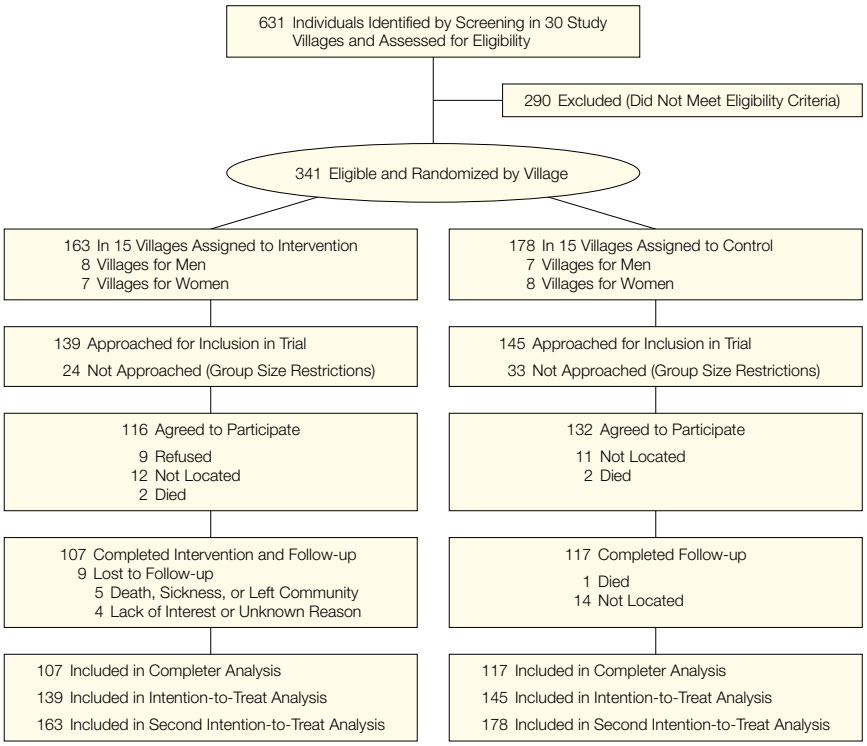

Table 1. Baseline Characteristics of Intervention and Control Groups*

Characteristics	Interviewed Before and After Intervention			All Persons Sought			All Eligible Persons		
	Intervention (n = 107)	Control (n = 117)	*P* Value	Intervention (n = 139)	Control (n = 145)	*P* Value	Intervention (n = 163)	Control (n = 178)	*P* Value
Age, y†	47.6 (15.2)	45.4 (16.6)	.30	46.4 (16.1)	44.1 (16.5)	.22	47.4 (17.0)	45.2 (17.0)	.25
Education, y	4.7 (2.8)	3.9 (3.3)	.07	4.6 (2.8)	4.0 (3.3)	.09	4.6 (2.8)	4.1 (3.4)	.15
Female, No. (%)‡	54 (50)	62 (53)	.71	70 (50)	76 (52)	.73	83 (51)	94 (53)	.73
Symptom duration, y	6.0 (6.4)	5.6 (5.7)	.66	5.6 (5.9)	5.6 (5.5)	.99	5.6 (5.8)	5.8 (5.4)	.80
Depression score§	23.6 (6.5)	24.3 (6.1)	.34	23.1 (6.6)	24.1 (6.1)	.14	23.0 (6.8)	23.6 (6.3)	.39
Function score§	12.3 (6.7)	12.5 (6.6)	.85	11.6 (6.9)	12.2 (6.6)	.51	11.9 (7.0)	11.9 (6.6)	.92
Major depression diagnosed, No. (%)‖	92 (86.0)	110 (94.0)	.04	116 (83.5)	136 (93.8)	.01	131 (80.4)	159 (89.3)	.02

*Data are presented as mean (SD) unless otherwise noted. See "Methods" section of text for explanation of participant categories.
†Mean (SD) ages of groups ranged between 28.1 (9.1) and 66.1 (10.5) years among controls and 26.7 (13.5) and 65.0 (13.7) years among intervention participants.
‡Although 8 male and 7 female villages were allocated to the intervention group, the average size of male groups was smaller, resulting in an equal distribution by sex in this group.
§Depression and function scores provide an approximation of overall relative severity of depression symptoms and dysfunction for the intervention vs control groups.
‖Major depression was diagnosed using the criteria of the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*.

Table 2. Comparison of Changes in Rates of Diagnosable Major Depression Before and After Intervention*

	Interviewed Before and After Intervention			All Persons Sought			All Eligible Persons		
Prevalence	Intervention	Control	*P* Value for Difference	Intervention	Control	*P* Value for Difference	Intervention	Control	*P* Value for Difference
Baseline	92/107 (86.0)	110/117 (94.0)	.04	116/139 (83.5)	136/145 (93.8)	.006	131/163 (80.4)	159/178 (89.3)	.02
Follow-up	7/107 (6.5)	64/117 (54.7)	<.001	31/139 (22.3)	90/145 (62.1)	<.001	46/163 (28.2)	113/178 (63.5)	<.001

*Data are presented as No./total (percentage). See "Methods" section of text for explanation of participant categories.

Table 3. Risk of Major Depression After Intervention Among Control Groups Compared With Intervention Groups*

	Interviewed Before and After Intervention	All Persons Sought	All Eligible Persons
Intervention	17.31 (7.63-39.27)	5.66 (3.43-9.33)	4.42 (3.08-6.34)
Intervention adjusted for covariates†	18.49 (8.00-42.76)	5.72 (3.52-9.30)	4.45 (3.10-6.39)

*Data are presented as odds ratios (95% confidence intervals) and are adjusted for a cluster effect (see "Methods" section of text). Also see "Methods" for explanation of participant categories.
†Data are adjusted for sex, years of education, and age.

cance tests and were not adjusted for cluster effects. However, because we found a positive correlation between clusters, adjusting for cluster effects would tend to reduce variance and cause group differences to be even less significant than the values reported herein.)

Prevalence of Depression

After completion of IPT, the point prevalence of major depression (those who met the *DSM-IV* algorithmic criteria) was significantly higher among the control groups than the intervention groups, regardless of which participants were included in the analysis (TABLE 2). Similarly, the cluster-adjusted odds ratio for major depression among controls compared with the intervention groups was also highly significant and remained so in the intention-to-treat analyses. Adjustment for covariates had very little effect (TABLE 3).

Change in Depression and Function Scores

The decline in depression scores was substantially greater among the intervention groups (TABLE 4). Mean changes were 17.47 and 3.55 for the intervention and control groups, respectively (11.59 and 2.38, respectively, for the entire group of eligible persons). Within the intervention groups, the mean change among men was 14.43 (95% CI, 12.32-16.55) compared with 20.46 (95% CI, 18.09-22.84) among women. There was no significant sex difference among controls.

Improvement in function scores was also greater among the intervention groups across all 3 analyses. When the change in depression score is included in the model with change in function as the outcome, the latter is statistically the same for both the intervention and control arms, indicating a close correlation between these 2 variables.

We examined how change in ability to perform individual tasks varied between the intervention and control groups. TABLE 5 shows all tasks assessed for both sexes and the mean change between preintervention and postintervention interviews. We found a significant difference in the amount of change comparing intervention participants with controls for all female tasks except "consoling the bereaved." In contrast, there were no statistically significant differences in the amount of change for the individual male tasks.

COMMENT

This study demonstrated that IPT-G-U was effective for depressionlike illness, depression symptoms, and associated dysfunction among persons in our study sample. There was a highly significant decline in overall severity of depression symptoms, the proportion of persons with major depression, and dysfunction among those who received IPT-G-U compared with the control groups. Unlike most function assessment instruments, ours was not restricted to health-related problems but refers to total dysfunction due to all causes (including lack of resources or assistance). Moreover, the tasks assessed were those previously chosen by a sample of the population as being particularly important to them.[6] Therefore, we believe that a large improvement in the ability to do these tasks could significantly affect community welfare and development, depending on how many people are affected by depression and experience improvement. Although the effects on depression and function were less when we based the analysis on intention to treat, they remained substantial and highly significant.

The experimental nature of the study permitted conclusions about cause and effect between depression and dysfunction. In our first assessment in Uganda in 2000, we noted a strong association, but that cross-sectional study could not determine whether depression was the cause or result of dysfunction. This is important in developing countries like Uganda, where dysfunction is common and due to multiple other causes, including physical illness. Since IPT-G-U focuses on depression only, and function improved in this study only in concert with depression (Table 4), we conclude that depression is causing dysfunction. This is in accord with longitudinal data from other countries[14] and suggests that targeting depression may be a useful way of improving function among depressed persons. Why the intervention appears to have benefited women in almost all areas of functioning considered separately whereas among men no such benefits for any par-

Table 4. Effect of Intervention on Depression and Function Scale Scores*

	Interviewed Before and After Intervention			All Persons Sought			All Eligible Persons		
	Intervention	Control	*P* Value	Intervention	Control	*P* Value	Intervention	Control	*P* Value
Depression Scale									
Baseline score, mean (SD)	23.64 (6.5)	24.46 (6.1)		23.06 (6.6)	24.19 (6.1)		23.04 (6.8)	23.65 (6.3)	
Follow-up score, mean (SD)	6.10 (6.3)	20.64 (9.0)		9.56 (9.0)	21.11 (8.5)		11.53 (10.0)	21.14 (8.19)	
Adjusted score change, mean (SE)†‡	17.47 (1.1)	3.55 (1.1)	<.001	13.83 (1.0)	2.70 (1.0)	<.001	11.59 (0.8)	2.38 (0.75)	<.001
Difference in adjusted mean score change (95% CI)†	13.91 (10.99 to 16.84)			11.13 (8.28 to 13.98)			9.20 (7.09 to 11.32)		
Function Scale									
Baseline score, mean (SD)	12.29 (6.7)	12.47 (6.6)		11.64 (6.9)	12.17 (6.6)		11.94 (7.0)	11.86 (6.6)	
Follow-up score, mean (SD)	4.27 (4.7)	8.66 (7.5)		5.47 (5.7)	9.09 (7.3)		6.67 (6.7)	9.36 (7.2)	
Adjusted score change, mean (SE)†‡	8.08 (0.6)	3.76 (0.5)	<.001	6.32 (0.5)	2.95 (0.5)	<.001	5.25 (0.5)	2.52 (0.4)	<.001
Difference in adjusted mean score change (95% CI)†	4.32 (2.80 to 5.84)			3.37 (1.96 to 4.79)			2.73 (1.51 to 3.95)		
Adjusted score change controlling for change in depression score, mean (SE)†‡	6.04 (0.6)	5.62 (0.6)	.65	4.53 (0.5)	4.65 (0.5)	.87	3.51 (0.4)	4.12 (0.4)	.31
Difference in adjusted mean score change controlling for change in depression score (95% CI)†	–0.43 (–1.43 to 2.29)			–0.11 (–1.48 to 1.26)			–0.61 (–1.78 to 0.56)		

Abbreviation: CI, confidence interval.
*See "Methods" section of text for explanation of participant categories.
†Data refer to the number of points by which the mean score changed between the preintervention and postintervention periods. Positive numbers refer to a reduction in scores between periods. For intention-to-treat analyses, no change was assumed for participants without postintervention data.
‡Data adjusted using a mixed-model approach, with village clusters as the random effects and adjusted for baseline scores.

ticular functional area were evident requires further study.

Our results illustrate the importance of a control study arm in assessing effectiveness. In this study, the control groups improved significantly over the course of the intervention. Since the severity of mental symptoms and function can vary with time, these changes are likely due to regression to the mean. They may also be related to other actions taken by those in the control groups, such as treatments by traditional healers (although those we interviewed in 2000 thought they were unable to help with these problems). Whatever the cause, a noncontrolled study of any intervention in this population would likely have shown an effect regardless of its true efficacy.

Because we did not assess whether the control groups took any other actions to relieve their depression, the trial comparison is not IPT vs nothing but IPT vs the usual treatment, whatever that may be. (We did not assess whether the intervention groups took any action to relieve their depression other than attending the IPT sessions.) This approach was

Table 5. Comparison of Mean Change in Ability to Perform Individual Tasks Before and After Intervention*

Tasks	Mean (95% Confidence Interval)	
	Intervention	Control
Male Tasks		
Personal hygiene	0.77 (0.43 to 1.11)	0.38 (0.04 to 0.72)
Farming	0.89 (0.39 to 1.38)	0.73 (0.25 to 1.21)
Head the home	1.02 (0.56 to 1.48)	0.53 (0.05 to 1.01)
Manual labor†	1.04 (0.57 to 1.51)	0.51 (0.12 to 0.90)
Plan for the family	1.06 (0.53 to 1.58)	0.69 (0.23 to 1.15)
Participate in community development activities†	0.85 (0.47 to 1.23)	0.36 (–0.07 to 0.80)
Attend meetings†	0.51 (0.19 to 0.83)	0.15 (–0.09 to 0.38)
Participate in burial ceremonies	0.42 (0.13 to 0.70)	0.16 (–0.01 to 0.33)
Socialize	0.30 (0.04 to 0.56)	0.16 (–0.24 to 0.57)
Female Tasks		
Personal hygiene‡	0.74 (0.42 to 1.06)	0.19 (–0.13 to 0.52)
Care for children‡	1.74 (1.34 to 2.14)	1.17 (0.76 to 1.58)
Cook‡	1.17 (0.82 to 1.51)	0.31 (–0.10 to 0.71)
Wash clothes/utensils‡	1.30 (0.88 to 1.72)	0.76 (0.37 to 1.14)
Clean house and surroundings‡	0.93 (0.56 to 1.29)	0.32 (0.01 to 0.64)
Grow food‡	1.28 (0.84 to 1.72)	0.56 (0.21 to 0.92)
Participate in community development activities‡	1.10 (0.68 to 1.52)	0.46 (0.06 to 0.85)
Attend meetings‡	0.69 (0.28 to 1.09)	0.10 (–0.21 to 0.40)
Console and assist the bereaved	0.24 (–0.02 to 0.50)	0.06 (–0.13 to 0.25)

*Data are only for participants interviewed before and after intervention. A positive 1 point is equivalent to a mean change from one category of response to the next lowest response category (eg, from "a lot of difficulty" to "a moderate amount of difficulty").
†$P<.10$ for the mean change in the intervention vs control groups. None of the differences were $P<.05$.
‡The difference between the intervention and control groups was statistically different ($P<.05$) for this task.

appropriate here because our purpose was not to investigate replacing existing approaches. Rather, it was to determine whether IPT-G-U would be a useful addition by World Vision International to the current methods of coping with depression by local people.

When informing participants of their allocation to the intervention or control arms, we reread the consent form and asked them to confirm their continuing participation in one of the study arms. This contravened standard procedures by which allocation is performed only after obtaining consent (or renewing it, as here). This was done because of our prior experience in sub-Saharan Africa, where random allocation is not well understood or accepted. Many previous attempts by one of us (P.B.) to explain that people "might" receive something were misinterpreted, resulting in resentment and withdrawals from the study. To avoid this, we thought it essential to explain at each stage exactly what would be happening to participants, particularly since one study arm would be receiving nothing beyond usual treatment at this time. We expected this approach to produce better data than the standard clinical trial procedure because, in our experience, few people refuse enrollment when they believe that they accurately understand what will happen. If refusals among both control and intervention groups are few, then there is no significant participation bias on the basis of study arm allocation. In this study, only 9 persons in intervention villages and none in control villages refused to participate, and demographic data (Table 1) suggest that the intervention and control groups were not significantly different overall. Once the study began, only 4 persons withdrew from the intervention groups while none of the controls who were recontacted after the intervention refused reinterview.

We are currently uncertain how long the effects of IPT will last. A postintervention assessment of both the intervention and control groups will be conducted 6 months after the intervention ended. We may find that it will be necessary to add a maintenance component to prevent recurrence. This is usually provided on a monthly basis.

To reduce interviewer bias, each participant was interviewed by the same interviewer at baseline and follow-up. This makes it more likely that an interviewer will know whether a person had the intervention, since they may remember some interviewees and where they are from. We are uncertain how often this occurred and, therefore, the extent to which blinding was compromised.

We still do not know the mechanism by which IPT was effective. Our study design could not suggest which elements were essential and which were not. For example, we cannot separate IPT per se from the group dynamics of simply meeting together. Testimonials from participants suggest that the group problem-solving element of IPT was vital, but this is not conclusive. We plan to investigate this in future studies.

In conclusion, we note that this was the first time that the trainers (H.V. and K.F.C.) had been in Africa or trained local people. This was also the first experience of the IPT trainees; many were hesitant at first and took weeks to gain confidence in the method. Under these circumstances, the effects of this intervention impressed us. We might expect even greater impact with more local experience with this approach.

Author Affiliations: Center for International, Emergency, Disaster and Refugee Studies (Dr Bolton) and Department of Mental Health (Ms Bass), Johns Hopkins University Bloomberg School of Public Health, Baltimore, Md; Epidemiology of Developmental Brain Disorders Department and International Program on Refugee Trauma (Dr Neugebauer) and Division of Clinical and Genetic Epidemiology (Drs Verdeli, Wickramaratne, and Weissman and Ms Clougherty), New York State Psychiatric Institute, New York; Gertrude H. Sergievsky Center, Faculty of Medicine (Dr Neugebauer), and Department of Psychiatry (Drs Verdeli, Wickramaratne, and Weissman and Ms Clougherty), College of Physicians and Surgeons, and Department of Epidemiology Mailman School of Public Health (Drs Verdeli, Wickramaratne, and Weissman and Ms Clougherty), Columbia University, New York, NY; and World Vision International, Washington, DC (Ms Speelman and Mr Ndogoni).

Author Contributions: Dr Bolton, as principal investigator, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Bolton, Neugebauer, Verdeli, Ndogoni, Weissman.
Acquisition of data: Bolton, Bass, Clougherty, Speelman, Ndogoni.
Analysis and interpretation of data: Bolton, Bass, Neugebauer, Wickramaratne, Weissman.
Drafting of the manuscript: Bolton, Bass, Neugebauer, Verdeli, Weissman.
Critical revision of the manuscript for important intellectual content: Bolton, Bass, Neugebauer, Clougherty, Wickramaratne, Speelman, Ndogoni.
Statistical expertise: Bass, Neugebauer, Wickramaratne.
Obtained funding: Bolton, Ndogoni.
Administrative, technical, or material support: Bolton, Bass, Clougherty, Speelman, Ndogoni.
Study supervision: Bolton, Verdeli, Speelman, Ndogoni, Weissman.
Funding/Support: Material support for Dr Neugebauer's participation in this project was provided by Ruth and David Levine and by funds from the International Program on Refugee Trauma, New York, NY.
Acknowledgment: The Mellon Foundation funded this project but had no role in data collection, analysis, manuscript preparation, or authorization for publication. Material support was provided by World Vision International in the form of vehicles, drivers, and local staff time. We thank those staff, local leaders, and the communities who welcomed us to work with them.

REFERENCES

1. *Mental Health: New Understanding, New Hope*. Geneva, Switzerland: World Health Organization; 2001.
2. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet*. 1997;349:1436-1442.
3. Ekin I, Shea M, Watkins J, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46:971-982.
4. Weissman M, Markowitz J, Klerman G. *Comprehensive Guide to Interpersonal Psychotherapy*. New York, NY: Basic Books; 2000.
5. Bolton P, Tang AM. An alternative approach to cross-cultural function assessment. *Soc Psychiatry Psychiatr Epidemiol*. 2002;37:537-543.
6. Bolton P. Cross-cultural validity and reliability testing of a standard psychiatric assessment instrument without a gold standard. *J Nerv Ment Dis*. 2001;189:238-242.
7. Verdeli H, Clougherty K, Bolton P, et al. Adapting group interpersonal psychotherapy (IPT-G-U) for a developing country: experience in Uganda. *World Psychiatry*. In press.
8. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci*. 1974;19:1-15.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.
10. 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;282:433-439.
11. Bolton P, Neugebauer R, Ndogoni L. Prevalence of depression in rural Rwanda based on symptom and functional criteria. *J Nerv Ment Dis*. 2002;190:631-637.
12. Omar RZ, Thompson SG. Analysis of a cluster randomized trial with binary outcome data using a multilevel model. *Stat Med*. 2000;19:2675-2688.
13. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. New York, NY: Oxford University Press; 1994.
14. Ormel J, VonKorff M, Ustun TB, Pini S, Korten A, Oldehinkel T. Common mental disorders and disability across cultures: results from the WHO Collaborative Study on Psychological Problems in General Health Care. *JAMA*. 1994;272:1741-1748.