Development of a patient rated scale for mental health global state for use during humanitarian interventions.

ORIGINAL ARTICLE

WILEY

Development of a patient rated scale for mental health global state for use during humanitarian interventions

Augusto E. Llosa[1] | Carmen Martínez-Viciana[2] | Cristina Carreño[2] |
Stella Evangelidou[3] | German Casas[4] | Caroline Marquer[1] | Marie Rose Moro[5,6] |
Bruno Falissard[6] | Rebecca F. Grais[1]

[1]Epicentre, Paris, France

[2]Médecins Sans Frontières, Barcelona, Spain

[3]Department of Psychiatry and Legal Medicine, Faculty of Medicine, Autonomous University of Barcelona, Barcelona, Spain

[4]Facultad de Medicina, Universidad de Los Andes, Bogotá, Colombia

[5]Médecins Sans Frontières, Université de Paris, Paris, France

[6]Epidemiology and Population Health, CESP/INSERM U1018, Paris, France

Correspondence
Caroline Marquer, Epicentre, Paris, France.
Email: caroline.marquer@epicentre.msf.org

Funding information
Médecins Sans Frontières

Abstract

Objective: We present the results of a cross-cultural validation of the Mental Health Global State (MHGS) scale for adults and adolescents (<14 years old).

Methods: We performed two independent studies using mixed methods among 103 patients in Hebron, Occupied Palestinian Territories and 106 in Cauca, Colombia. The MHGS was analyzed psychometrically, sensitivity and specificity, ability to detect clinically meaningful change, compared to the Clinical Global Impression-Severity scale (CGI-S). Principal component analysis was used to reduce the number of questions after data collection.

Results: The scale demonstrated good internal consistency, with a Cronbach alpha score of 0.80 in both settings. Test retest reliability was high, ICC 0.70 (95% CI [0.41–0.85]) in Hebron and 0.87 (95% CI [0.76–0.93]) in Cauca; inter-rater reliability was 0.70 (95% CI [0.42–0.85]) in Hebron and 0.76 (95% CI [0.57–0.88]) in Cauca. Psychometric properties were also good, and the tool demonstrated a sensitivity of 85% in Hebron and 100% in Cauca, with corresponding specificity of 80% and 79%, when compared to CGI-S.

Conclusions: The MHGS showed promising results to assess global mental health thereby providing an additional easy to use tool in humanitarian interventions. Additional work should focus on validation in at least one more context, to adhere to best practices in transcultural validation.

KEYWORDS
conflict, global scale, humanitarian intervention, mental health, psychology

---

1 | INTRODUCTION

Populations exposed to natural and man-made crises are at increased risk of developing mental health problems (Fazel, Reed, Panter-Brick, & Stein, 2012; Jones et al., 2009; Miller & Rasmussen, 2010; Reed, Fazel, Jones, Panter-Brick, & Stein, 2012). Despite growing evidence of the benefit of a variety of mental health and psychosocial support (MHPSS) interventions in these settings, recommendations remain largely based on expert opinion rather than robust measurement (Interagency Standing Committee (IASC), 2006). Relying on expert opinion may be the only possible recourse in situations without empirical evidence for recommendations on patient care, but generation of evidence remains imperative to improve MHPSS interventions. Additional data are needed to evaluate mental health in conjunction with MHPSS interventions, and to promote evidence driven practices (Charter, 2011; Tol et al., 2011).

This evidence generation aims to improve and adapt mental health interventions to different crisis-affected contexts and better address the needs of these populations.

Mental-health measures, usually through scales, and particularly those including the patient's perspective, provide important information to improve and adapt mental care (Bolton & Tang, 2002; Rodin & van Ommeren, 2009; World Health Organisation, 2013, pp. 1–27). These scales, however, are often either disorder specific, modulated around western concepts and populations, lack cross-cultural validation, require scoring by specialists, and are too lengthy and cumbersome to implement in many crisis-affected contexts (Bolton & Tang, 2002; Chiumento, Khan, Rahman, & Frith, 2016; Dozio, Bizouerne, Feldman, & Moro, 2018; Van Ommeren, 2003; Viswanath & Chaturvedi, 2012; Wolpert et al., 2012; World Health Organisation, 2013, pp. 1–27).

In humanitarian settings, especially conflict affected ones, there is sporadic and inconsistent access to patients due to insecurity along with a shortage and inequitable distribution of psychiatrists and psychologists (Charlson et al., 2019; Patel, Chowdhary, Rahman, & Verdeli, 2011). There is a clear need for tools and approaches which are easy to use, can be easily administered by non-specialized personnel and help assess overall changes in mental health. Despite a large number of mental health status questionnaires, either addressing symptoms or function, few serve to measure global mental health state from the patient's perspective, and none have specifically addressed the specificities of complex humanitarian conflict contexts (Kirmayer, Guzder, & Rousseau, 2013; Lohr, 1988; Rodin & van Ommeren, 2009).

To facilitate the evaluation of MHPSS interventions supported by Médecins Sans Frontières (MSF), we aimed to create a quick and versatile mental health global state (MHGS) scale for use in crisis-affected populations. We aimed for an instrument that is adaptable to a variety of ages, mental conditions and contexts, which is easy to administer, and possible to be done quickly when contact may be limited and intermittent. As such, we developed a short, easy to administer questionnaire, which captures the perspective of the patient. Two independent validations were performed, one in Hebron, Occupied Palestinian Territories (oPT) and one in Cauca, Colombia. In Hebron, MSF has been providing care and mental health support since 1991. The context is characterized as a chronic conflict where populations are exposed to conflict related violence as well as limitations of movement, harassment, demolition of homes and restricted access to health care. In Cauca, MSF provided medical care beginning in 1985 and mental health support between 2003 and 2016. At the time of the study, the population was suffering direct consequences of a prolonged armed conflict, including repeated mass forced-displacement, movement restrictions and poor access to healthcare (Gómez-Restrepo et al., 2016; Sanchez-Padilla, Casas, Grais, Hustache, & Moro, 2009). The most treated conditions were common mood and anxiety disorders including trauma and stress related disorders. Patients with severe and chronic mental disorders (e.g., schizophrenia and bipolar disorders) were referred directly to partners for specialized care, not provided within the MSF intervention and none were included in this study. In both contexts, the mental health interventions comprised a maximum of 16 psychological sessions per patient, delivered by a trained psychologist in as regular intervals as possible, to adapt to patients that might be forced to displace, or due to movement restrictions. Here, we present the development and validation studies of a long and short version for the MHGS scale for adults and adolescents (14 and above). The MHGS was psychometrically examined including inter-rater, test-retest reliability, and sensitivity to change analyses compared to the Clinical Global Impression-Severity scale (CGI-S).

2 | METHODS

The tool is interviewer-administered and aims to capture the patient's perspective on their global mental state. The development of the tool included two phases. The first phase included the identification of the main domains of interest and a list questions to provide information on change in patients receiving the mental health intervention. Through focus groups in the community and pilot testing, understanding of the concepts described behind preselected questions were addressed before assessing psychometrics (Lohr et al., 1996). The second phase examined the psychometric properties of the tool and aimed to simplify the tool to a minimum set of questions. These steps are described below.

2.1 | Creation of the list of questions

Instrument domains and questions were selected through a consensus group made up of experts in instrument development and validation and cross-cultural psychology. These included mental health professionals involved in the mental health interventions at the two project sites (France, Spain, Colombia and oPT). The expert group met over the course of several weeks and derived a list of potential questions reflecting key domains from their professional expertise and previous work in these contexts. Once the list of questions was developed through consensus, community focus

groups in both project sites were held. Questions were then listed by domain of interest and their phrasing was considered for simplicity and cross-cultural interpretability (Kortmann, 1987, 1990; Kortmann & Ten Horn, 1988). A final list of questions was developed after follow-up with additional expert consultation.

2.2 | Translation and adaptation of the tool

Selected questions were translated into local languages (Arabic and Spanish) and back-translated by professional translators in each context. Discrepancies were resolved by consensus and questions revised accordingly. The instrument was also piloted in each context among professionals and laypersons from June to November 2013. The instrument was piloted by a dedicated study coordinator and provided to individuals working in various professions in the project as well as community members, none of whom participated in the study. Once developed, professional local partners (members of the Ministries of Health, psychiatrist, and psychologist) were asked to check the relevance of questions. These steps aimed to assess interpretability of questions and to ensure that concepts represented are as intended and similarly understood in at least two cultural-linguistic contexts.

2.3 | Training

Interviewers were the psychologists working in the MSF programs where the validation took place (oPT Cauca). They were specifically trained over 2 days in interviewing, administration of the tool, the informed consent process, and were supervised by the study coordinator (also a psychologist) at each study site.

2.4 | Study population and sampling

Both contexts were selected due to the conflict affected the populations, and also long-standing provision of mental health care in an insecure context. Study participants were recruited among patients directly affected by the conflict and receiving mental health care at MSF supported clinics. After providing written informed consent, adult and adolescent participants, comprising any patient presenting for mental health and psychosocial care at one of the MSF clinics, were interviewed. Recruitment continued until the target sample size was reached (100). No differences were expected between those with appointments in different seasons. The baseline measure was considered as the session provided to the patient when included in the study.

2.5 | Procedures

Testing a minimum of 7–10 patients per question with no less than 100 participants overall is standard practice (Kline, 2013; Terwee et al., 2007). Interviews were timed with the patient sessions. In

cases where retest interviewers were not possible in person, these took place by phone. After obtaining informed consent, participants were asked demographic questions and administered the MHGS, in an area offering privacy, by first introducing the questionnaire and explaining its purpose and how to rate each question of the MHGS with the aid of pictograms. The Disability Assessment Schedule 2.0 (DAS) was used to measure the patient's overall level of functioning (Üstün, 2010). All questions were read to participants, and their response recorded in study specific forms.

The attending psychologist completed routine information for the patient dossier, including the Clinical Global Impression (for overall severity assessment; CGI-S) and the Global Assessment of Functioning (GAF) during the patient's appointment (Hilsenroth et al., 2000; Guy, 1976). These and other study related data points were then transferred from the chart to the study forms by the study coordinator.

Patients in the test-retest assessment subgroup were to be re-interviewed by the study coordinator within 48 h. For the inter-rater reliability assessment subgroup, a cohabiting adult relative of the patient was to be interviewed within 24 h, either by phone or in person depending on security constraints. The subgroup participating to the responsiveness testing were to be interviewed at different scheduled appointments, within five sessions from the study's baseline interview. For the study, the same routine clinic procedures were used. The patients who accepted to participated received one reminder call for their next appointment.

2.6 | Data analyses

Internal consistency of the instrument was assessed by Cronbach alpha (where alpha values 0.7 = acceptable, 0.8 = good, 0.9 = excellent; DeVellis, 2016). Inter-rater reliability compared assessments by the patient themselves with those of a cohabitating adult family member. Test-retest reliability was assessed between scores from the first study related assessment with those of the retest. Both reliability measures were estimated using intra-class correlation co-efficients (ICC) and presented with corresponding 95% confidence intervals and p-values (F test), following interpretation guidelines by Ciccetti (i.e., <0.40 bad; 0.40–0.59 fair; 0.60–0.74 good; 0.75–1.00 excellent; Cicchetti, 1994). Differences in distribution of initial and retest or second-rater scores were compared with paired t-tests or Wilcoxon signed-rank tests, as appropriate, and p-values presented.

Convergent validity was assessed comparing the results to CGI-S, GAF and DAS, through Pearson & Spearman correlation tests, as appropriate depending on distribution of data, and interpreted accordingly (0.1 small; 0.3 medium; 0.5 large; Cohen, 2013). Criterion validity was checked by calculation of sensitivity and specificity against the CGI-S based on the clinician's assessment through structured interview. The Kaiser–Meyer–Olkin (KMO) test was used to determine sampling adequacy. A value 0.80 was set as the criteria for exploring structure (34). Parallel principal component analysis was performed to better understand the relationship between questionnaire items (internal construct), compute factor loadings, to

eliminate redundancies and specify the number of items for a shortened scale (35). The following goodness of fit statistics were considered: chi-squared test (acceptable model fit if $p > 0.05$); root mean square error of approximation (RMSEA, acceptable model fit if <0.06); comparative fit index (CFI, acceptable model fit if >0.96); standardized root mean residual (SRMR, acceptable model fit if <0.08; Hu & Bentler, 1999).

The ability of the instrument to detect clinically important changes over time was assessed by comparing the instrument's measurements at least two different time points of the mental health care provided. Detected differences were compared to those by another instrument (CGI-S). Responsiveness was considered if the Area Under Curve (AUC) was >0.70 ([CVO] American Psychiatric Association, 2005; Forkmann et al., 2011; Hall, 1995; Williams, 1976). The AUC of receiver operating characteristics (ROC) curve is based on a 3-point increase of the CGI-S. The minimal clinically important difference (MCID) for the simplified version of the scale was calculated using linear regression of MHGS against a 1-point change in CGI-S.

Data were analyzed using STATA SE 12.0 (STATA Corporation).

The protocols received approval from ethics review boards of MSF, the University of Cauca, Colombia, and the Ministry of Health, oPT. All participants were already receiving free mental health care in MSF programs, and were further referred to partners for psychiatric and other medical care if needed.

2.7 | Measures

The domains selected were psychological functioning (cognitive ability, interpersonal relation, and ability to perform daily activities), symptoms of psychological suffering (somatizations, negative thoughts or emotions, and difficulties in sleeping and eating) and general perception of psychological well-being or suffering. Two questions on coping mechanisms were also added as this was an area perceived to be of therapeutic importance to MSF. One question on perceived usefulness of the mental health and psychosocial sessions was included but not intended to be part of the assessment scale.

The complete version of MHGS was composed of 13 questions for adults and adolescents. The patient had to quote or point to the respective image for "not at all", "a little," "some," "fair amount," or "a lot" using the following scale (Figure 1). The short version of MHGS included a total of six questions with a minimum score of 1 ("not at all") and maximum of 5 ("a lot") per question (5-point Likert scale), thus total scores for the scale range between 6 and 30 points, with higher scores indicating greater difficulties (Figure 2).

3 | RESULTS

3.1 | Scale structure and item reduction

Following the psychometrics properties result, the scale was proved robust enough to undergo structure analysis (KMO = 0.80)

on Hebron baseline data. The steep drop in factor loadings eigenvalues over 100 replications in parallel principal component analysis (PCA) suggested a single component model. Six questions showed notably higher loading factors (superior to 0.30 in PCA; Table 1). This model includes two questions (Q) on function (Q2) performance of daily activities and (Q3) getting along with others; two questions on specific symptoms of suffering (Q11) troubled by thoughts or worries, and (Q12) feelings of sadness or nervousness; and two questions related to general well-being, addressing sleep problems (Q6) and (Q13) a general question on suffering (Table 1). This analysis allowed for a simplified version including six questions. Confirmatory factor analysis was carried out using the 6-question version of MHGS on the Cauca dataset. Goodness of fit statistics were consistent with the reduced model, chi-squared test = 6.96, $p = 0.64$ (acceptable model fit if $p > 0.05$); RMSEA = 0.00 (acceptable model fit if <0.6); CFI = 1.00 (acceptable model fit if >0.90); excepting SRMR = 0.28 (acceptable model fit if <0.08). Results that follow are for the reduced version of the instrument.

3.2 | MHGS study participants

In Hebron, 103 participants were enrolled between January and May 2014. In Cauca, 106 participants were enrolled between December 2013 and August 2014. The majority of participants were female in both Hebron and Cauca with a median age of 32 in Hebron and 33 in Cauca (Table 2).

In Hebron, most study participants were already in treatment at enrollment (median four sessions). Median of therapeutic session at follow-up was eight ($n = 30$), with a median of five intra-study therapeutic sessions (range 3–9). All participants responded directly to the interviewer (i.e., no proxy responders). Most ($n = 88$, 85%) were in individual therapy. The most common principal diagnosis at enrollment were other anxiety disorders ($n = 32$, 31%), Post-traumatic Stress Disorder (PTSD; $n = 20$, 19%) and depression ($n = 16$, 16%).

In Cauca, most study participants were on their first session at enrollment, with median of one session ($n = 106$) and three sessions at follow-up ($n = 31$). The median number of intra-study therapeutic sessions was two (range 2–3). Most participants ($n = 104$, 98%) responded themselves. Proxy responders were husband ($n = 1$) or mother ($n = 1$), care takers of the patients who were not able to answer directly. Of the 106 participants, 97 (92%) were enrolled in individual therapy. The most frequent diagnoses at enrollment were depression ($n = 29$, 27%), PTSD ($n = 11$, 10%) and prolonged bereavement ($n = 13$, 12%).

3.3 | Psychometrics properties

Internal consistency was considered good, with a Cronbach alpha score of 0.80 for both Hebron and Cauca.

1. In general how have you felt in the past week?

In the past week

2. Did you have problems performing daily activities, such as cooking, bathing, or chores outside the household?

3. Did you have problems getting along with family or friends?

4. Did you have problems paying attention, or understanding or remembering things?

5. Did you have problems with your appetite, such as having too little or too much hunger?

6. Did you have problems with your appetite, such as having too little or too much hunger?

7. Did you have headaches, pain or discomfort in your body?

8. Did you lose control or become hostile or aggressive?

9. Did you try to do something to make your situation better?

10. Did you get comfort and understanding from someone? (except from MSF psychologist or staff)

11. Have you been bothered by unpleasant thoughts worries or ideas?

12. Have you felt sad or nervous

13. Have you been suffering?

| Not at all | A little | Some | Fair amount | A lot |
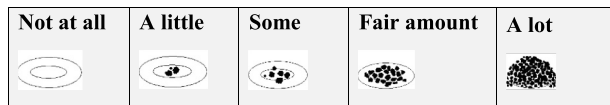
Answer options for first question only were: "quite poorly", "not so well", "ok", "well", "quite well"

FIGURE 1 Adult/adolescent version of Mental Health Global State (MHGS) instrument, questions

In the past week,

1. Did you have problems performing daily activities, such as cooking, bathing, or chores outside the household?

2. Did you have problems getting along with family or friends?

3. Did you have any problems sleeping, such as sleeping too much or not enough?

4. Have you been bothered by unpleasant thoughts worries or ideas?

5. Have you felt sad or nervous?
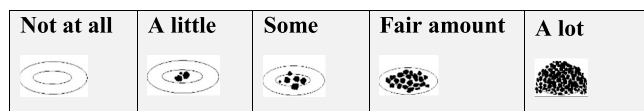
6. Have you been suffering?

| Not at all | A little | Some | Fair amount | A lot |

FIGURE 2 Reduced, adult/adolescent version of Mental Health Global State (MHGS) instrument, questions

3.4 | Test-retest reliability

In Hebron, 35 patients, and in Cauca 42 were re-interviewed with the MHGS scale to assess repeatability. Due to scheduling constraints and traceability of participants between appointments, many patients were re-interviewed after the planned 48 h; in Hebron, 98% of re-tests occurred within 72 h of the baseline interview and in Cauca, 84%. On average scores at retest were lower in both in Hebron (mean difference = 2.5, SD = 5.4, $p < 0.05$), and Cauca (mean difference = 1.9, SD = 3.8, $p < 0.05$). ICC corresponded to 0.70 (95% CI [0.41–0.85], $p < 0.05$) in Hebron and 0.87 (95% CI [0.76–0.93], $p < 0.05$) in Cauca. These results show the instrument tended to consistently indicate lower retest scores for those with lower initial scores and higher retest scores for those with higher initial scores, thus suggesting the instrument had good test-retest reliability.

TABLE 1 Eigenvalues from PCA averaged over 10 replications (left), and Component loading values (right)

Eigen values MHGS, Hebron				Component loading values MHGS, Hebron				
	PCA	PA	Dif	MHGS question	Category	Component 1	Component 2	Component 3
Component 1	**4.11**	1.68	2.43	(1) How felt	General	0.2139	−0.3092	0.2589
Component 2	1.51	1.48	0.03	(2) Daily activities	Function	0.3028	0.1891	0.0648
Component 3	1.2	1.34	−0.14	(3) Getting along with others	Function	0.2974	0.2252	−0.4038
Component 4	0.93	1.23	−0.29	(4) Pay attention and understand	Function	0.2269	0.0703	0.2922
Component 5	0.91	1.13	−0.22	(5) Appetite	General	0.2157	0.0912	0.3589
Component 6	0.82	1.03	−0.22	(6) Sleeping	General	**0.3048**	−0.0464	0.4196
Component 7	0.7	0.95	−0.25	(7) Headaches and pain	Symptom	0.2721	−0.3107	−0.0499
Component 8	0.65	0.87	−0.22	(8) Aggression and loss of control	Symptom	0.2927	0.2368	−0.4421
Component 9	0.61	0.8	−0.19	(9) Action to improve situation	Coping	−0.0085	0.5871	0.3125
Component 10	0.51	0.72	−0.21	(10) Sought support comfort	Coping	0.0025	0.5398	0.0457
Component 11	0.45	0.66	−0.21	**(11) Troubling thoughts and worries**	Symptom	0.3717	0.0023	−0.1942
Component 12	0.35	0.58	−0.23	(12) Feeling sad and nervous	Symptom	**0.4009**	−0.0713	0.1194
				(13) Level of suffering	Symptom	0.353	−0.0933	−0.1606

Abbreviations: MHGS, Mental Health Global State; PA, Parallel Analysis; PCA, Parallel Principal Component Analysis.

TABLE 2 Age and gender of MHGS study participants, Hebron (left), Cauca (right), 2014

Hebron, oPT	Female		Male		Total		Cauca, CO	Female		Male		Total	
Baseline	n	%	n	%	n	%	Baseline	n	%	n	%	n	%
14–24 years	13	18.3	18	56.3	31	30	14–24 years	28	32.2	4	21.1	32	30.2
25–44 years	39	54.9	10	31.3	49	48	25–44 years	42	48.3	7	36.8	49	46.2
45–75 years	19	26.8	4	12.5	23	22	45–80 years	17	19.5	8	42.1	25	23.6
Total	71	100	32	100	103	100	Total	87	100	19	100	106	100

Abbreviation: oPT, Occupied Palestinian Territories.

3.5 | Inter-rater reliability

In Hebron, 36 patients were included in this subgroup, and in Cauca, 42. In Hebron, 34 of 36 (94.4%) secondary raters were interviewed the same day as the patient, while in Cauca all 42 were. On average family members tended to give lower scores than those directly by the patient; in Hebron, mean difference = 1.7, SD = 5.1, $p < 0.05$; in Cauca, mean difference = 0.2, SD = 4.0, $p = 0.4$. The ICC was 0.70 (95% CI [0.42–0.84], $p < 0.05$) in Hebron and 0.77 (95% CI [0.57–0.88], $p < 0.05$) in Cauca. These results show that perceived severity by secondary raters was consistent with that given by patients (i.e., lower scores with lower scores and higher with higher).

3.6 | Convergent validity

Correlations with other instruments at baseline, were mostly medium. In Hebron: these were CGI-S rho = 0.48; GAF rho = −0.21; DAS rho = 0.35. In Cauca these were: CGI-S rho = 0.49; GAF rho = −0.48; DAS rho = 0.59). For greater interpretability, correlations included participants who had measures for all tests at baseline, $n = 71$ in Hebron and $n = 84$ in Cauca. Correlations were significant ($p < 0.05$) for all comparisons except with GAF in Hebron.

Comparison of MHGS to CGI-S was possible among the 32 participants in Hebron and 31 in Cauca who underwent both baseline and follow-up assessments. By this measure in Hebron, 7 (22%) improved by 3 or more points in CGI-S, 11 (34%) each by two and one points, and 3 (9%) did not show improvement by CGI-S. Similarly, in Cauca, 3 (10%) improved by three points in CGI-S, 12 (39%) by two points, 11 (35%) by one point, and five (16%) showed no improvement.

The correlation of change in MHGS and CGI-S was rho = 0.38 ($p < 0.05$) in Hebron and rho = 0.46 ($p < 0.05$) in Cauca for this version of the scale.

When considering the psychologist's perspective expressed through CGI-S as the gold standard measure, MHGS scale correctly classified improvements of 3 or more points among 81% of patients (85% sensitivity and 80% specificity) in Hebron using a cut-off MHGS score change of eight or greater (AUC = 0.85, 95% CI [0.67–1.00]). In
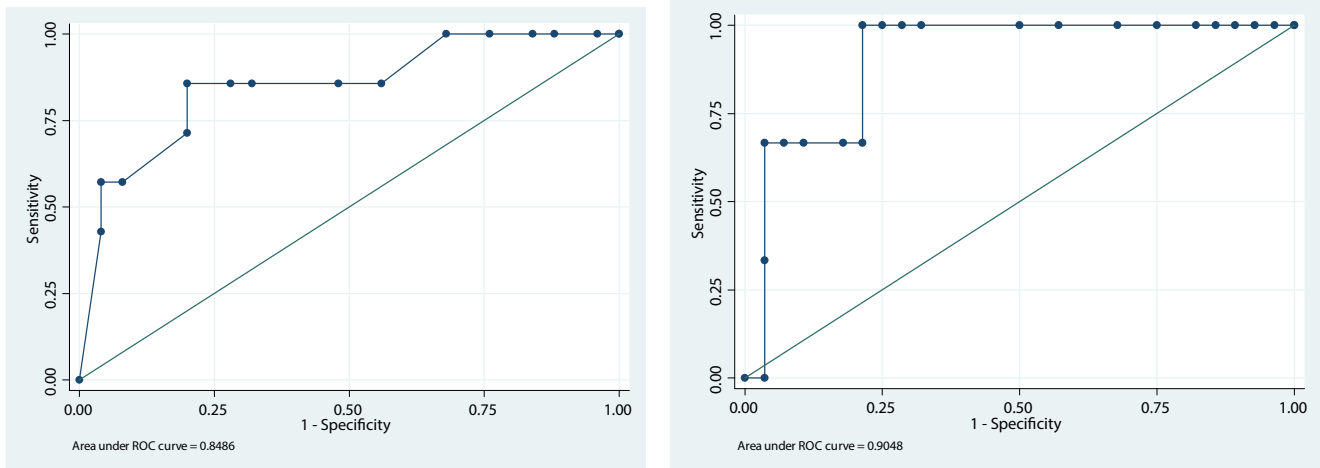
FIGURE 3 Receiver operating characteristic (ROC) curves for score change in six question version of Mental Health Global State (MHGS) for adolescents and adults compared to improvements in Clinical Global Impression (CGI), Hebron (left), and Cauca (right), 2014. Improvement was defined as a score change of three or more points in CGI; $n = 32$ in Hebron, and $n = 31$ in Cauca

Cauca MHGS correctly classified improvements of three or more points among 81% of patients (100% sensitivity and 79% specificity) using a cut-off MHGS score change of 10 or greater (AUC = 0.90, 95% CI [0.76–1.00]). ROC curves for both settings are shown in Figure 3.

There was strong agreement between the scales in more than half of participants assessed, and moderate to strong agreement in at least three quarters of those assessed. There were no serious discrepancies noted in either setting. No patient demonstrated strong improvement by CGI-S and deterioration by MHGS, nor strong improvement in MHGS and worsening by CGI-S.

The minimum clinically important difference (MCID) or clinically relevant cut off for change in MHGS score was four points. This value corresponds to the regression coefficient indicating the number of MHGS point difference associated with a one-point change in CGI-S. For Hebron, this value was 2.3 ($p < 0.05$), and for Cauca 4.2 ($p < 0.05$). Considering the corresponding measurement errors for the scale were 2.9 and 3.3 for Hebron and Cauca, respectively, the most appropriate cut off for improvement or worsening is a change in score of 4 points in either direction.

4 | DISCUSSION

To our knowledge, this is the first development and validation of an outcome scale for adults and adolescents (14 years old and above) created specifically from and for humanitarian interventions. It applies a simple set of questions aimed at determining the global state of the patient at different time-points in care, by applying a few simply worded questions on key domains of interest to mental health workers and populations in these contexts. The MHGS allows for monitoring change in global mental health state over time in several key outcomes of interest relating to function, symptoms and patient suffering. The results suggest that Arabic and Spanish versions are dependable and offer a way to measure patient evolution throughout the care process.

Despite this promising result, it should be noted that we faced security constraints leading to the minimum 100 participants in each setting. Also, correlations with baseline values from other instruments utilized in the study were mixed, albeit it correctly classified over 80% of participants against therapist-noted changes in function and symptomatology. It is important to note that patients' assessment through MHGS asks about difficulties in the previous week, while the psychologist might have taken a more global perspective when rating patient functionality and symptomatology at the same point in time.

A few limitations also need to be considered. First, it should be noted that assessments took place at different points in the therapy. Given that continuity of care is not guaranteed in conflict settings (Sanchez-Padilla et al., 2009), we aimed to have a tool that is practical enough to be applied at any session and would be able to discern change in global state within a few sessions. For this reason, we did not require that the first and last interviews match the first and last therapeutic session, nor a fixed number of sessions. Second, we relied on a live-in relative as proxy responder for two baseline interviews where the patient was unable to answer directly, and proxies were also used to assess inter-rater reliability, compared to the patient's own rating. Despite statistical agreement between the measures (i.e., by ICC), we recognize that proxy responders may not always be a valid alternative to a patient's direct response. For instance, we noted on average significantly lower MHGS scores by the secondary assessor compared to the patients' own assessments. Nonetheless, when a patient is no longer available, an unfortunate reality of conflict settings, a close, live-in relative or caretaker may be the best suited alternative to inform the program about the patient's state. Third, some additional challenges occurred by differences in test versus retest, as the original interview occurred in person and in most cases the re-test occurred by phone, often days after the associated interview. Scores where on average lower than initial ones, which could potentially be due to the phone interview. Additionally, the several

days delay might have increased the chance of the effects of the session or other external factors to play a role. It is to be noted as well, that the MHGS scale requires two measures to detect change, and that a noted change is not, at any one point at individual level, necessarily a result of the intervention; it may reflect changes due to other life events. Fourth, despite formal training of psychologists on the use of CGI-S, it was a recent addition to routine assessments in these projects, and some changes in score were higher than expected. While assessing concordance with MHGS was still possible, the MCID will need to be confirmed in other settings. Strong training in the use of CGI-S, which is recommended to be used alongside MHGS to capture the therapists' perspective, is advisable. Last, the simplified scale was discerned statistically from results of the application of the longer version of the scale. The interspersion of questions from the longer scale could have affected the responses of the questions included in the shorter version. Further testing of the simplified version is therefore planned, though of note, the scale is already in use in several humanitarian contexts and performing well.

At the programmatic level, the tool can help mental health staff to assess impact of their interventions, adjust the objectives and contents of the mental health interventions and care strategy, and provide an overall evaluation of programmatic impact in a brief and comprehensive manner. The use of the MHGS in combination with other evaluation support would hopefully improve the delivery of mental health services during humanitarian intervention. Useful in acute and chronic crises, the tool may be most relevant for conflicts where humanitarian interventions may last for many years and in some cases decades. In these contexts, there is ample opportunity to improve and refine programming. The instrument may be applicable beyond humanitarian interventions, albeit additional validation work would be needed to show it validity outside these contexts.

5 | CONCLUSION

The MHGS showed promising results to help assess MHGS thereby providing an additional easy to use tool for both patient and mental health programmatic follow-up in a humanitarian setting. Additional work should focus on validation in at least one more context, to adhere to best practices in transcultural validation, and fine tune MCID cut offs.

CONFLICT OF INTERESTS
The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS
Conception and design of study: Augusto E. Llosa, Marie Rose Moro, Bruno Falissard, Carmen Martínez-Viciana, Stella Evangelidou, German Casas, Rebecca F. Grais, and Caroline Marquer. Selection of instrument's domain and questions: Marie Rose Moro, Bruno Falissard, Carmen Martínez-Viciana, Stella Evangelidou, German Casas, Augusto E. Llosa, and Rebecca F. Grais. Conducted the research: Augusto E. Llosa and Carmen Martínez-Viciana. All authors read and approved the final manuscript.

ETHICAL APPROVAL
The study received approval from the MSF Ethics Review Board, the Comité de Ética para la Investigación Científica (CEIC), Universidad de Cauca, Colombia, and clearance from the Palestinian Ministry of Health and local authorities. All participants to the research signed the informed consent. Participants younger than 18 years old signed an assent form and their parent/caretaker signed the consent form.

ORCID
Caroline Marquer 🔟 https://orcid.org/0000-0002-9085-715X

REFERENCES
Bolton, P., & Tang, A. M. (2002). An alternative approach to cross-cultural function assessment. *Social Psychiatry and Psychiatric Epidemiology*, *37*(11), 537–543.

Charlson, F., van Ommeren, M., Flaxman, A., Cornett, J., Whiteford, H., & Saxena, S. (2019). New WHO prevalence estimates of mental disorders in conflict settings: A systematic review and meta-analysis. *Lancet*. *394*(10194), 240–248.

Charter, H. (2011). *The sphere project* [Internet]. Response (Vol. 1. p. 402). Retrieved from www.practicalactionpublishing.org/sphere

Chiumento, A., Khan, M. N., Rahman, A., & Frith, L. (2016). Managing ethical challenges to mental health research in post-conflict settings. *Developing World Bioethics*. *16*(1), 15–28.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. *6*(4), 284

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. *Biometrics*. Academic press.

(CVO) American Psychiatric Association. (2005). Global assessment of functioning (GAF) scale. In *DSM-IV-TR manuel diagnostique et statistique des troubles mentaux*.

DeVellis, R. F. (2016). Scale development: Theory and applications. In Sage (Ed.), *Scale development: Theory and applications* (Vol. 26, pp. 109–110). Sage Publications.

Dozio, E., Bizouerne, C., Feldman, M., & Moro, M. (2018). Operational and ethical challenges of applied psychosocial research in humanitarian emergency settings: A case study. *Intervention*. *16*(1), 46

Fazel, M., Reed, R. V., Panter-Brick, C., & Stein, A. (2012). Mental health of displaced and refugee children resettled in high-income countries: Risk and protective factors. *The Lancet*, *379*, 266–282.

Forkmann, T., Scherer, A., Boecker, M., Pawelzik, M., Jostes, R., & Gauggel, S. (2011). The clinical global impression scale and the influence of patient or staff perspective on outcome. *BMC Psychiatry*, *11*(1), 83.

Gómez-Restrepo, C., Aulí, J., Tamayo Martínez, N., Gil, F., Garzón, D., & Casas, G. (2016). Prevalencia y factores asociados a trastornos mentales en la población de niños colombianos, Encuesta Nacional de Salud Mental (ENSM) 2015. *Revista Colombiana de Psiquiatria*. (45), 39–49.

Guy, W. B. R. R. (1976). Clinical global impression (CGI). *ECDEU Assessment Manual for Psychopharmacology*. 217–222.

Hall, R. C. W. (1995). Global assessment of functioning: A modified scale. *Psychosomatics*. 36(3), 267–275.

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R, ... Holdwick, D. J. (2000). Reliability and validity of DSM-IV axis V. *American Journal of Psychiatry*. 157(11), 1858–1863.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*.

Interagency Standing Committee (IASC). (2006). *IASC guidelines on mental health and psychosocial support in emergency settings* [Internet]. Geneva, Switzerland: IASC 2006.

Jones, L., Asare, J. B., El Masri, M., Mohanraj, A., Sherief, H., & van Ommeren, M. (2009). Severe mental disorders in complex emergencies. *Lancet*, 374(9690), 654–661.

Kirmayer, L. J., Guzder, J., & Rousseau, C. (2013). *Cultural consultation: Encountering the other in mental health care*, New York, USA: Springer Science & Business Media.

Kline, P. (2000). *The Handbook of Psychological Testing*. Psychology Press.

Kortmann, F. (1987). Problems in communication in transcultural psychiatry: The self reporting questionnaire in Ethiopia. *Acta Psychiatrica Scandinavica*, 75(6), 563–570.

Kortmann, F. (1990). Psychiatric case finding in Ethiopia: Shortcomings of the self reporting questionnaire. *Culture Medicine and Psychiatry*, 14(3), 381–391.

Kortmann, F., & Ten Horn, S. (1988). Comprehension and motivation in responses to a psychiatric screening instrument. Validity of the SRQ in Ethiopia. *British Journal of Psychiatry*, 153(1), 95–101.

Lohr, K. N. (1988). Outcome measurement: Concepts and questions. *Inquiry*, 25, 37–50.

Lohr, K. N., Aaronson, N. K., Alonso, J., Burnam, M. A., Patrick, D. L., Perrin, E. B., & Roberts, J. S. (1996). Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clinical Therapeutics*, 18(30), 979–984.

Miller, K. E., & Rasmussen, A. (2010). War exposure, daily stressors, and mental health in conflict and post-conflict settings: Bridging the divide between trauma-focused and psychosocial frameworks. *Social Science & Medicine*, 70(1), 7–16.

Patel, V., Chowdhary, N., Rahman, A., & Verdeli, H. (2011). Improving access to psychological treatments: Lessons from developing countries. *Behaviour Research and Therapy*. 49(9), 523–528.

Reed, R. V., Fazel, M., Jones, L., Panter-Brick, C., & Stein, A. (2012). Mental health of displaced and refugee children resettled in low-income and middle-income countries: Risk and protective factors. *The Lancet*, 379, 250–265.

Rodin, D., & van Ommeren, M. (2009). Commentary: Explaining enormous variations in rates of disorder in trauma-focused psychiatric epidemiology after major emergencies. *International Journal of Epidemiology*, 38, 1045–1048.

Sanchez-Padilla, E., Casas, G., Grais, R. F., Hustache, S., & Moro, M.-R. (2009). The Colombian conflict: A description of a mental health program in the department of tolima. *Conflict and Health*. 3(1), 13

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., ... de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42.

Tol, W. A., Barbui, C., Galappatti, A., Silove, D., Betancourt, T. S., Souza, R., ... Van Ommeren, M. (2011). Mental health and psychosocial support in humanitarian settings: Linking practice and research. *The Lancet*, 378, 1581–1591.

Üstün, T. B., Kostanjsek, N., Chatterji, S., & Rehm, J. (2010). *Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0*, Geneva, Switzerland: World Health Organization.

Van Ommeren, M. (2003). Validity issues in transcultural epidemiology. *British Journal of Psychiatry*, 182(5), 376–378.

Viswanath, B., & Chaturvedi, S. K. (2012). Cultural aspects of major mental disorders: A critical review from an Indian perspective. *Indian Journal of Psychological Medicine*, 34(4), 306.

Wolpert, M., Ford, T., Trustam, E., Law, D., Deighton, J., Flannery, H., & Fugard, R. J. (2012). Patient-reported outcomes in child and adolescent mental health services (CAMHS): Use of idiographic and standardized measures. *Journal of Mental Health*, 21(2), 165–173.

World Health Organisation. (2013). Comprehensive mental health action plan 2013–2020. Geneva 66th World Heal Assem [Internet] (pp. 1–27).