Abstract:	150
Text:	3,998
Tables:	4

Feasibility, acceptability, and clinical utility of the Cultural Formulation Interview: Mixed-methods results from the DSM-5 international field trial

Roberto Lewis-Fernández, MD, MTS

Neil Krishan Aggarwal, MD, MBA, MA

Peter C. Lam, MPH

Hanga Galfalvy, PhD

Mitchell G. Weiss, MD, PhD

Laurence J. Kirmayer, MD

Vasudeo Paralikar, MD, PhD

Smita N. Deshpande, MD, DPM

Esperanza Díaz, MD

Andel V. Nicasio, MSEd

Marit Boiler, MPH

Renato D. Alarcón, MD, MPH

Hans Rohlof, MD

Simon Groen, MA

Rob C.J. van Dijk, MSc

Sushrut Jadhav, MBBS, MD, MRCPsych, PhD

Sanjeev Sarmukaddam, PhD

David Ndetei, MBChB, DPM, MRCPsych, FRCPsych, MD, DSc

Monica Z. Scalco, MD, PhD

Kavoos Bassiri, MS, LMFT, LPCC, CGP

Sergio Aguilar-Gaxiola, MD, PhD

Hendry Ton, MD, MS

Joseph Westermeyer, MD, PhD

Johann M. Vega-Dienstmaier, MD

Corresponding author: Roberto Lewis-Fernández, M.D., Professor, Department of Psychiatry, Columbia University and New York State Psychiatric Institute, 1051 Riverside Drive, Unit 69, New York, NY 10032; Tel.: 1-646-774-8102; Fax: 1-646-774-8105; rlewis@nyspi.columbia.edu.

Financial support: This research was supported by the American Psychiatric Association, the New York State Office of Mental Health (NKA, PL, HG, AVN, MB) and institutional funds from the New York State Psychiatric Institute (RLF). The Pune site received support from the KEMH Research Centre and the Kenya site was partially funded by the Africa Mental Health Foundation. The funders did not have any input into the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Potential conflicts of interest: Three co-authors (RLF, NKA, LJK) receive royalties from an edited book on the Cultural Formulation Interview, published by American Psychiatric Publishing, Inc. None of the other authors have any conflicts to report.

Data Access, Responsibility, and Analysis: The Corresponding Author hereby acknowledges his full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis, and confirms that all authors have approved the manuscript.

Acknowledgments: The authors are grateful for the assistance of Oanh Meyer, Victoria Mutiso, Lincoln Khasakhala, Anne Mbwayo, NN Mishra, Triptish Bhatia, Antonio Lozano, Luis Fiestas, Adelguisa Mormontoy, Martín Arévalo, Spencer Case, Seung-Hee Hong, Samantha Díaz, Ravi DeSilva, Venkat Bhat, Kwame McKenzie, Loren Olsen, Ladson Hinton, Devon E. Hinton, Sophie Bäärnhielm, James Boehnlein, Cécile Rousseau, Jaswant Guzder, Darrel A. Regier, David J. Kupfer, William Narrow, Diana Clarke, Jennifer Shupinka, and Francis Lu.

ABSTRACT

Background: There is a need for clinical tools to identify cultural issues in diagnostic assessment.

Aims: Assess feasibility, acceptability, and clinical utility of the DSM-5 Cultural Formulation Interview (CFI) in routine clinical practice.

Method: Mixed-methods evaluation of field trial data from six countries. The CFI was administered to diagnostically diverse psychiatric outpatients during a diagnostic interview. Post-evaluation session, patients and clinicians completed debriefing qualitative interviews and Likert-scale questionnaires. Durations of CFI administration and full diagnostic session were monitored.

Results: Mixed-methods data from 318 patients and 75 clinicians found the CFI feasible, acceptable, and useful. Clinician feasibility ratings were significantly lower than patient ratings and other clinician-assessed outcomes. After administering one CFI, however, clinician feasibility ratings improved significantly and subsequent interviews required less time.

Conclusions: The CFI was included in DSM-5 as a feasible, acceptable, and useful cultural assessment tool.

Declaration of interest: Three co-authors receive royalties from the CFI Handbook.

Key words: Cultural assessment, diagnostic interview, implementation science, transcultural psychiatry

Unexamined cultural differences in how patients and clinicians frame illness and care may distort diagnosis and assessments of severity, impose communication barriers, compromise engagement, adherence, and response, and unnecessarily prolong patients' suffering.[1,2] Patient-clinician differences in age, gender, sexual orientation, socioeconomic status, race/ethnicity, religion, language, and/or national origin can contribute to cultural differences in all clinical interactions.[3,4] The DSM-IV Outline for Cultural Formulation (OCF) is a conceptual framework that helps clinicians identify the impact of culture on illness and care during a clinical evaluation.[5,6] The OCF is widely used in clinical training and cultural competence initiatives.[7-10] However, its implementation in routine care has proved challenging:[11] clinicians had to improvise questions to collect the information, received limited guidance on which patients would benefit most, and faced uncertainty about whether to implement the OCF as a separate assessment or embed it in a standard clinical evaluation.[12-14] The lack of a structured instrument also impeded research on cultural assessment and inclusion of cultural information in clinical trials.[15,16]

In response, the American Psychiatric Association's DSM-5 Cross-Cultural Issues Subgroup (DCCIS) developed the Cultural Formulation Interview (CFI)[17] to operationalize the OCF for routine use in the clinical assessment of any patient, based on a literature review and consensus-building discussions with designers of OCF-based interviews.[18] The CFI instruments comprise an initial assessment interview (core CFI), an informant interview for collateral information, and 12 supplementary modules that expand on these basic assessments. The core CFI consists of an introduction, open-ended questions for patients, and instructions to clinicians for each question. Acknowledging the need for global relevance and recognizing international work on the OCF, sites in six countries participated in the field trial.

This report presents findings from the international field trial that tested the 14-item pilot version of the core CFI (Appendix 1) in three service domains based on patient and clinician feedback. Together with other field trial data not reported here, this process resulted in the final 16-item version in DSM-5.[19] We assessed several factors related to successful implementation of clinical innovations in service settings,[20] including patient and clinician perceptions of the CFI's feasibility ("Can it be done in clinical settings?"), acceptability ("Do patients and clinicians like it?"), and potential clinical utility ("Is it

helpful?"). We also considered whether closed- and open-ended assessments yielded similar results, and whether outcomes showed a practice effect, improving with experience. Our study is the first to examine these service domains for a tool to enhance cultural competence in multiple international settings.

Methods

Study design and settings

The CFI field trial was designed by the DSM-5 DCCIS via regular teleconferences.[19,21] The study was conducted from November 2011 to September 2012; the New York site coordinated logistics for all sites. The study design purposively included samples of diverse patients, clinician disciplines, and types of outpatient services, because a goal of the DSM-5 trials was to test the feasibility, acceptability, and utility of proposed diagnoses and assessments under varied clinical conditions to determine inclusion in DSM-5.[22,23] Each site aimed to enrol at least 30 patients from affiliated psychiatric outpatient clinics in Canada (1 site), India (2), Kenya (1), the Netherlands (1), Peru (1), and the USA (5). Sites were chosen based on involvement of a principal investigator (PI) in the DCCIS and aimed to include diverse cultural populations and types of outpatient services (general community, immigrant/refugee, and ethnic-focus clinics).

An opportunity sample of new and existing patients at each site was enrolled using a standard recruitment script. Clinicians who had no prior contact with their study patient conducted the interviews ("study clinicians"). Clinicians did not interview their own patients because prior knowledge and a pre-existing relationship would confound study aims focusing on an initial assessment. Current patients were referred by treating clinicians to local study clinicians. Each study clinician was expected to interview 3-6 patients during the trial to assess practice effects. Each patient participated only once. Patients and clinicians could also invite companions (e.g., relatives) to participate in the interview and subsequent assessments.[24]

All study clinicians participated in a two-hour CFI training session at their site consisting of (1) reviewing the core CFI's written guidelines; (2) a 24-minute video demonstration; (3) interactive

behavioural simulations with coaching and feedback from local PIs; and (4) a question-and-answer period.

The study clinician administered the CFI followed by a routine diagnostic assessment. Topics of the CFI comprise four cultural domains: (i) definition of the problem; (ii) perceptions of cause, context, and support; (iii) factors affecting self-coping and past help-seeking; and (iv) factors affecting current help-seeking.

All sessions were audio-taped with patient consent. The study was approved by each site's Institutional Review/Ethics Board and followed local informed consent regulations. All patients completed their locally-approved consent process.

Participants

Eligible patients were age 16 or older and fluent in the language of the local clinicians. We required the language match to avoid using interpreters who might introduce cultural information not obtained through the CFI. Patients were excluded if they were acutely suicidal or homicidal, intoxicated or in substance withdrawal, or if their condition seriously limited the assessment (e.g., dementia). Eligible study clinicians had a clinical degree permitting them to see patients, consistent with each country's requirements.

Assessments

Pre-interview, patients and clinicians completed demographic surveys. Clinicians also indicated their professional training and cultural competence experiences. Local PIs identified demographic factors recognized by their governments as indicators of social differences, avoiding a USA-based characterization.[19,25] After every session, study clinicians provided patients' DSM-IV diagnoses and patients and clinicians completed follow-up questionnaires and semi-structured qualitative interviews. All assessments were translated into the local languages at each site and reviewed by a bilingual committee of mental health professionals for consensus.[26]

Quantitative: Participants completed two brief questionnaires: the Debriefing Instrument for Patients (DIP) and the Debriefing Instrument for Clinicians (DIC), which comprise self-administered, Likert-scale

items assessing feasibility, acceptability, and clinical utility (Appendix 2) coded as "Strongly disagree," "Disagree," "Agree," and "Strongly Agree." As with other DSM-5 trials,[22] these instruments were created for use in the CFI field trial. Items were selected for measurement by the DCCIS with reference to three domains (feasibility, acceptability, clinical utility) likely to affect the implementation of assessments like the CFI.[20,22] The same content was included in each instrument, with wording adapted for each stakeholder group. As a measure of feasibility independent of self-report, we assessed the duration of the CFI and the total diagnostic interview (including the CFI), based on session audio files.

Qualitative: Separate semi-structured qualitative interviews (8-9 questions, previously reported[19:512]) with patients and clinicians conducted by research assistants at each site provided more detailed accounts of the impact of the CFI on the initial evaluation. These interviews assessed participants' perceptions of the most and least helpful aspects of the CFI, its impact on interview quality and outcomes, and its role in clinical practice, including diagnosis and treatment planning. Each site provided written English summaries of the interviews to the coordinating site.

Analysis

Quantitative: SAS version 9.4 (Cary, NC) was used for all analyses.

Descriptive information: Patient and clinician characteristics were compared cross-nationally using ANOVA for continuous variables and Chi-square (or Fisher's exact test) for categorical variables; the Kruskal-Wallis test was used for ordinal or continuous variables with skewed distributions.

DIC/DIP: Negative DIC/DIP responses were coded as -2 (strongly disagree) or -1 (disagree) and positive responses as +1 (agree) or +2 (strongly agree).[24,27] Missing responses were imputed using the mean of the non-missing items within the assessment domain for the individual. Mean proportion of missing responses was 4.5% (SD=1.4%) for the DIP (range 2.8%-7.6% for a single item) and 2.2% (SD=1.0%) for the DIC (range 0.9%-4.1%).

Cronbach's alpha was used to assess the internal consistency of the three DIC/DIP domains. For domains with alphas <.70, inter-item correlation matrices, item correlation with total, and changes to alpha by item were examined to detect problematic items; these items were excluded from subsequent

analyses.

Mean DIC/DIP scores for feasibility, acceptability, and utility were compared within patient and clinician cohorts, cross-nationally and overall. We also compared the overall patient and clinician mean scores for each assessment domain; remaining items in domains with excluded items were also compared individually. To account for site-specific effects, clinicians seeing several patients, and the inclusion of new and existing patients to the clinic, we used generalized linear mixed-effects models (PROC GLIMMIX in SAS), with random intercepts for site and clinician and a fixed effect for new patient status. Tukey-Kramer post-hoc tests that adjust for multiple comparisons were used to identify significant patient-clinician differences.[28]

Duration: Durations of the CFI and the full diagnostic interview (including the CFI) were compared separately cross-nationally using PROC GLIMMIX to adjust for new patient status and clinician effects. The proportion of total interview time devoted to the CFI was also calculated.

Practice effect: To determine whether clinicians' accumulated experience with the CFI affected their perceptions of the outcomes, we analysed changes in DIC scores over subsequent CFI interviews; we also analysed interview duration and the proportion of time devoted to the CFI in the full interview for each clinician. A mixed-effects model adjusted for clinician and site effects (but not patient newness, since patients were always new to study clinicians). Separate mixed-effects models and Tukey-Kramer post-hoc tests contrasted DIC assessment domains between and within each administration, respectively.

Qualitative

Qualitative analyses were conducted by a three-person multidisciplinary team (public health, sociology, and psychiatry) using deductive content analysis and working independently of the quantitative analysis team. Deductive content analysis codes qualitative data using pre-established categories based on theoretical frameworks.[29,30] Each debriefing interview was coded for feasibility, acceptability, and utility according to a codebook (developed by NKA): feasibility and acceptability were defined as per Proctor et al[20] and their definition for appropriateness was used to define utility, consistent with the terminology of the DSM-5 trials.[25] Coder training consisted of two 1-hour sessions. Each coder labelled each interview

phrase with one unique code for feasibility, acceptability, or utility to minimize bias.[31] Inter-rater reliability of 80% was achieved using a random 10% selection of transcripts. Iterative revision of the codebook was conducted over 5 weeks by reviewing concordance among codes and concepts, developing new sub-codes, memoing, specifying code definitions with parameters (appropriate and inappropriate use), and reviewing data examples until new information produced no change to coding categories. All debriefing interviews were uploaded into NVivo (QSR International 2012) and randomly assigned for coding. NVivo reports were generated for codes, exploring patterns, and drafting analytical memos by theme. Qualitative codes were counted by individual respondent and by number of mentions per text to analyse data by session and for the total sample.

Results

Patient characteristics

The field trial enrolled 321 patients; three were under 16 and were excluded, leaving 318 for analysis, of whom 189 were new and 129 existing patients. They had a mean age of 41.4 and 10.6 years of education; half were female (Table 1 and Appendix 3). Most countries had an even distribution of employed, unemployed, and out-of-the-labour-force respondents (e.g., retired), except for the USA where nearly half were disabled. Marital status differed by country. Proportion of foreign-born individuals ranged widely, from 0% in Peru to 97% in Canada. Patients' primary language varied by site. Significant cross-national differences were observed for all socio-demographic variables (gender: $p<0.05$; all others: $p<0.001$). Clinically, 70% of patients received one DSM-IV Axis I diagnosis, 20% received two, 7% three or more, and 2% none (Table 1); this proportion varied significantly across countries ($p<0.001$). Depressive disorders were diagnosed most frequently, followed by anxiety disorders.

INSERT TABLE 1

Clinician characteristics:

Seventy-five clinicians were enrolled, with an average age of 38.4; over 50% were female, except in the Netherlands and Peru (Table 2). Nearly 50% were psychiatrists or psychiatric trainees, 28% psychologists, and 15% social workers. Countries differed substantially on several indices. Kenyan clinicians had a mean of 3 years of practice, had seldom/never treated patients of different cultures, and all had <10 hours of cultural training. By contrast Dutch clinicians had 15.6 years of practice, 91% had daily cross-cultural contacts, and half had >50 hours of cultural training. The proportion of foreign-born clinicians ranged from 0% in India and Peru to 57% in Canada. All variables differed significantly across countries, except for age and gender.

INSERT TABLE 2

Self-report outcome ratings:

Cronbach's alphas for the DIC were high: 0.78 (feasibility), 0.80 (acceptability), and 0.89 (utility). DIP internal consistency was high for utility (0.82) but minimal for feasibility (0.18) and acceptability (0.17). Item-based analyses identified one problematic item under feasibility ("Took more time to share my perspective then I wanted") and acceptability ("Were too personal"); both items were negatively worded. Removing these items[27] increased Cronbach's alpha for feasibility (0.45) and acceptability (0.48) (see Appendix 2), these domains each now containing two items. Prior research on cross-cultural variation with negatively-worded survey items supports this approach.[32]

Patient and clinician ratings of feasibility, acceptability, and clinical utility were positive, but varied significantly cross-nationally (Appendix 4). Once adjusted for site effects, mean overall results for all three outcomes (Table 3) were positive among patients—scoring 1.26-1.33 on a scale from -2 to +2—but evaluations were less positive among clinicians, with scores of 0.93-0.98 on utility and acceptability and 0.75 on feasibility. Overall feasibility was significantly lower than the other indices among clinicians, and significantly lower than patients' feasibility rating. Clinicians also rated acceptability and utility lower than patients, but not significantly. By contrast, patient scores across assessment domains were nearly

identical.

After excluding the two problematic DIP items, comparison of remaining single-item ratings of feasibility (easy to understand, $t(10)=5.27$, $p<.001$; improved flow, $t(10)=2.32$, $p=.043$) and acceptability (encourage clinician use, $t(10)=2.17$, $p=.055$; felt at ease, $t(10)=21.3$, $p=.059$) across patient and clinician assessments revealed the same pattern as the analysis of means. DIC single-item results (Appendix 2) identified clinician concerns about CFI comprehensibility and interview flow (feasibility) and about CFI impact on clarification of diagnosis, cultural background, severity, and patient-clinician differences (utility). DIP single-item results did not indicate specific concerns, although identification of barriers to care (utility) scored somewhat lower than other items.

INSERT TABLE 3

Duration:

Average CFI duration ranged from 18.8 minutes in the Netherlands to 29.2 in Kenya ($p<0.001$) and total interview duration ranged from 37.6 minutes in Kenya to 88.2 in the Netherlands ($p<0.001$). Average overall CFI duration was 23.4 minutes, within a 54.1-minute intake. Cross-nationally, the proportion of the interview devoted to the CFI varied significantly (Appendix 4).

Practice effects

Clinician (DIC) feasibility ratings improved significantly with practice, from an average of 0.59 at first use to 0.96 at the sixth or subsequent administration (Table 4). Acceptability and utility scores, by contrast, were stable and positive over time. Feasibility differed significantly from acceptability and utility ratings only for the first administration. Mean CFI duration decreased significantly, by over 4 minutes, consistent with clinician's reports of increasing confidence in feasibility. This effect on CFI duration was evident by clinicians' second CFI administration, and remained stable at 22-23 minutes thereafter. Mean total diagnostic interview duration also decreased significantly but gradually, by over 12 minutes from first to last administration. CFI proportion of the total interview time increased slightly with

practice.

INSERT TABLE 4

Qualitative interviews

Qualitative coding of the post-CFI open-ended debriefing interviews identified a pattern similar to the closed-ended quantitative DIC/DIP analysis (Appendix 5). Clinicians had a more negative perception of CFI feasibility than patients: 107 of 318 clinician interviews included negative feasibility comments about the CFI as a tool, and 39 negative feasibility comments concerning prospects for clinical implementation, compared to only 26 and 7 negative comments, respectively, among 318 patients. By contrast, patients made 81 positive feasibility comments about the CFI and 14 positive feasibility comments about its implementation prospects, while clinicians only made 30 and 9 positive comments, respectively. Clinicians' concerns focussed on feasibility; acceptability and utility elicited more positive views. By contrast, patients' comments were largely positive across all assessment domains. These patterns were identical whether views were coded by participant or by total number of utterances.

Clinicians were concerned about the CFI's feasibility as a tool, faulting its organization ("jumbled") and its placement early in the clinical interview. They also worried about implementation-related issues, such as time burden and whether the format was overly structured. Patients were more positive about feasibility, praising the CFI structure ("from basic questions to more complex…in the sense of how you feel") and clinicians' non-"pressured" administration. However, some patients found "all the details" confusing; they also worried the CFI might be too time-consuming for busy clinicians.

Regarding acceptability, clinicians praised the CFI's ability to generate empathy but found some questions difficult to administer (e.g., on the clinician-patient relationship). Patients liked the flow and person-centeredness of the CFI questions ("I felt like I was talking to someone I knew"), though some became upset by the life content elicited.

The views on CFI utility were the most positive. Generally, both groups of participants found the

CFI useful with respect to diagnosis, treatment planning, and understanding the patient's situation, including the role of culture in mental illness (e.g., "will help me get better treatment;" "will help me understand the patient's problem extensively on the basis of cultural, religious things").

Discussion

The DSM-5 Cultural Formulation Interview field trial was the first international study to examine clinician and patient perceptions of the feasibility, acceptability, and clinical utility of a cultural assessment interview designed for use in routine clinical practice in diverse cross-national settings. The international trial recruited 321 patients and 75 clinicians over 11 sites in 6 countries. Mixed-methods analyses showed that both patients and clinicians found the CFI to be feasible, acceptable, and clinically useful, and these findings supported its inclusion in DSM-5. The diversity of the samples and sites – and the fact that both closed-ended and open-ended assessments yielded similar results when analysed blind to one another – enhance the clarity, robustness, and generalizability of our findings.

The strategy for our quantitative analysis was developed at one of the study sites in India and used here with minor modifications. [27] Site-specific analyses of the field trial data have also found positive perceptions of implementation-related outcomes.[19,24,27] In the full sample, patients assessed the CFI more positively than clinicians, and the difference was significant for feasibility. Clinicians were more concerned about feasibility than about acceptability or utility. The qualitative data, based on post-CFI open-ended interviews, likewise showed greater clinician concern about feasibility, compared to patient views and other clinician-rated outcomes.

To be successfully implemented, a new assessment should address the concerns of all stakeholders;[33] our design enabled us to examine views of both clinicians and patients. Differing views of feasibility among stakeholders probably reflect practical concerns and limited time of busy clinicians,[34] relevant for effective allocation of health system resources that must balance clinical values and practical constraints.[35] Although stakeholders' perceived acceptability and utility of an assessment or intervention may conceivably differ,[20,36] we found no significant differences in our field trial.

Our mixed-methods design identified barriers to implementation of the CFI field trial version. DIC single-item analysis and qualitative data largely converged. They also confirm a previously-published subanalysis of New York-site qualitative data, which had identified lack of differentiation of the CFI from routine clinical assessments, question clarity and ordering, and the time required for the interview as main concerns.[19] The consistency of these concerns in our cross-national analysis is striking, given the cultural and clinical diversity among study participants.

Many of these issues were addressed in the revised version of the CFI published in DSM-5. Based on the field trial results, the revision clarified confusing wording, improved the flow of questions, and distinguished the intent of the CFI from other aspects of clinical management. Four questions were condensed into two, and one question on cultural identity and three on the views of the patient's social network were added. Future research should examine the impact of implementing the CFI on clinical practice and outcomes, and in cultural competence training.

The practice effect identified from self-report and interview-duration data has important implications for questions about feasibility. Findings suggest that two hours of training followed by experience administering a few interviews may be sufficient to address clinicians' concerns about feasible use of the instrument, even in a diverse sample of provider disciplines and of cultural competence experience across sites.[25] Consideration of the practice effect may facilitate uptake of the CFI, mindful that implementing any new tool may initially evoke resistance,[27] which may lessen over time if its relative advantage becomes clear in routine practice.[37] Indeed, by the second CFI administration, clinician feasibility scores increased substantially and no longer differed significantly from clinician acceptability and utility scores. Duration of the CFI interview, an objective indicator of feasibility, showed a similar practice effect, decreasing by 4 minutes by the second administration and remaining stable thereafter.

Duration of the full diagnostic interview also decreased significantly albeit more gradually. By the last administration, the duration of the full intake assessment, including 22 minutes for the CFI, was 50 minutes. This is comparable to the time required for an initial assessment in many mental health settings. In the USA, for example, average duration of community-based psychiatric visits (initial and follow-up

combined) was 32–38 minutes in 1989–2006;[38-40] intakes are often 45-50 minutes. Our study found substantial international variation in intake duration. Some of this variation may derive not only from resource constraints—few clinicians for many patients—but also from clinic characteristics. The sites with the longest intakes (Canada and the Netherlands) included specialized programs for immigrants and refugees, whereas most other sites operated in general community clinics. Sites also differed significantly in the proportion of total interview time devoted to the CFI, yet all were able to integrate the CFI into routine intake procedures. The proportion of the interview devoted to the CFI increased slightly with experience, suggesting clinicians continued to find it useful and that the information it yielded was relevant to other aspects of the diagnostic interview, inasmuch as less time was required for the overall interview as a practice effect.

This study has several limitations. Participating clinics were recruited purposively and may pay higher-than-average attention to cultural issues; clinicians who were most interested may have done more interviews, potentially confounding the positive practice effect. However, clinicians' interest did not prevent them from stating their concerns candidly in the qualitative interviews. Second, we developed our own self-report measures of service outcomes because at the time of the field trial there were no psychometrically-validated quantitative measures of implementation-related outcomes.[41] The DIP feasibility and acceptability domains of assessment had psychometric limitations. One-time use of these assessments is consistent with the DSM-5 field trial goal of testing proposed diagnostic criteria (or tools such as the CFI) for inclusion or revision in the final manual.[22] The congruence of the qualitative and quantitative results as a benefit of the mixed-methods design supports the robustness of the DIP data. Third, the study interview consisted of the CFI session followed by the routine diagnostic assessment. All clinicians were asked to inform patients when they transitioned from the CFI to the routine assessment. It is possible that some patients did not distinguish the CFI component of their evaluation from the routine diagnostic component when responding to questions in their debriefing interviews.

Despite these limitations, the DSM-5 international field trial results support the feasibility, acceptability, and clinical utility of the Cultural Formulation Interview. The positive valuation by patients

and clinicians suggests that it is worth investing about 20 minutes of an initial evaluation on a cultural assessment that holds promise for enhancing clinical communication, diagnostic accuracy, effective treatment planning, patient satisfaction, engagement, and clinical response.[19,21] The promise of such benefits argues for further study of CFI implementation effects on clinical and service outcomes (e.g., cost and sustainability).[20] As a practical matter, the field trial suggests an attractive learning curve, with clear benefits after two hours of training and a single interview. A 2014 *Lancet* Commission on culture and health advocated use of the CFI in all medical subspecialties, not just psychiatry,[42] highlighting its broad relevance. Although further studies of implementation outcomes are needed, our findings indicate good prospects for meeting these acknowledged needs.

References

1. Adeponle AB, Thombs BD, Groleau D, Jarvic E, Kirmayer LJ. Using the cultural formulation to resolve uncertainty in diagnoses of psychosis among ethnoculturally diverse patients. *Psychiatric Serv* 2012;**63**: 147-53.
2. Bhui K, Bhugra D. Explanatory models for mental distress: Implications for clinical practice and research. *Br J Psychiatry* 2002; **181**:6-7.
3. Ayonrinde O. Importance of cultural sensitivity in therapeutic transactions: Considerations for healthcare providers. *Dis Manage Health Outcomes* 2003;**11**: 233-48.
4. Schouten B, Meeuwesen L. Cultural differences in medical communication: A review of the literature. *Patient Educ Couns* 2006;**64**: 21-34.
5. Lewis-Fernández R. Cultural formulation of psychiatric diagnosis. *Cult Med Psychiatry* 1996;**20**: 133-44.
6. Mezzich J. Cultural formulation and comprehensive diagnosis: Clinical and research perspectives. *Psychiatric Clin North Am* 1995;**18**: 649-87.
7. Aggarwal NK, Rohrbaugh RM. Teaching cultural competency through an experiential seminar on anthropology and psychiatry. *Acad Psychiatry* 2011;**35**: 331-2.
8. Scarpinati Rosso M, Bäärnhielm S. The cultural formulation: a model to combine nosology and patients' life context in psychiatric diagnostic practice. *Transcult Psychiatry* 2009;**46**: 406-28.
9. Hansen H, Dugan TM, Becker AE, Lewis-Fernández R, Lu FG, Oquendo MA, et al. Educating psychiatry residents about cultural aspects of care: A qualitative study of approaches used by U.S. expert faculty. *Acad Psychiatry* 2013;**37**: 412-6.
10. Rohlof H, Knipscheer J, Kleber R. Use of the cultural formulation with refugees. *Transcult Psychiatry* 2009;**46**: 487-505.
11. Groen S, Richters A, Laban K, Deville W. Implementation of the Cultural Formulation through a newly developed Brief Cultural Interview: Pilot data from the Netherlands. *Transcult Psychiatry* in press.
12. Lewis-Fernández R. The cultural formulation. *Transcult Psychiatry* 2009;**46**: 379-82.
13. Caballero-Martínez L. DSM-IV-TR Cultural formulation of psychiatric cases: Two proposals for clinicians. *Transcult Psychiatry* 2009;**46**: 506-23.
14. Aggarwal NK. Adapting the cultural formulation for clinical assessments in forensic psychiatry. *J Am Acad Psychiatry Law* 2012;**40**: 113-8.
15. Alarcón RD. Culture, cultural factors and psychiatric diagnosis: Review and projections. *World Psychiatry* 2009;**8**: 131-9.
16. Weiss M. Cultural epidemiology: An introduction and overview. *Anthropol Med* 2001;**8** :5-29.
17. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 4th ed*. Washington, DC: American Psychiatric Association, 1994.
18. Lewis-Fernández R, Aggarwal N, Baarnhielm S, Rohlof H, Kirmayer LJ, Weiss MG, et al. Culture and psychiatric evaluation: Operationalizing cultural formulation for DSM-5. *Psychiatry: Interpers Biol Processes* 2014;**77**: 130-54.
19. Aggarwal NK, Nicasio AV, DeSilva R, Boiler M, Lewis-Fernández R. Barriers to implementing the DSM-5 cultural formulation interview: A qualitative study. *Cult Med Psychiatry* 2013;**37**: 505-33.
20. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Admin Policy Men Health* 2011;**38**: 65-76.
21. Lewis-Fernández R, Aggarwal N, Hinton L, Hinton D, Kirmayer L (eds.). *The DSM-5 Handbook on the Cultural Formulation Interview.* Washington, DC: American Psychiatric Publishing, 2016.
22. Clarke DE, Narrow WE, Regier DA, Kuramoto SJ, Kupfer DJ, Kuhl EA, et al. DSM-5 field trials in the United States and Canada, Part I: Study design, sampling strategy, implementation, and analytic approaches. *Am J Psychiatry* 2013;**170**: 43-58.

23. Regier D, Narrow W, Clarke D, Kraemer HC, Kuramoto SJ, Kuhl EA, et al. DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of the selected categorical diagnoses. *Am J Psychiatry* 2013;**170**: 59-70.
24. Hinton L, Aggarwal N, Losif A-M, Weiss M, Paralikar V, Deshpande S, et al. Perspectives of family members participating in cultural assessment of psychiatric disorders: Findings from the DSM-5 international field trial. *Int Rev Psychiatry* 2015;**27**: 3-10.
25. Aggarwal NK, Lam P, Castillo EG, Weiss MG, Díaz E, Alarcón RD, et al. How do clinicians prefer cultural competence training? Findings from the DSM-5 Cultural Formulation Interview field trial. *Acad Psychiatry* 2016;**40**: 584-91.
26. Bravo M, Canino GJ, Rubio-Stipec M, Woodbury-Farina M. A cross-cultural adaptation of a psychiatric epidemiologic instrument: The diagnostic interview schedule's adaptation in Puerto Rico. *Cult Med Psychiatry* 1991;**15**: 1-18.
27. Paralikar V, Sarmukaddam S, Patil K, Nulkar A, Weiss M. Clinical value of the cultural formulation interview in Pune, India. *Indian J Psychiatry* 2015;**57**: 59-67.
28. Kramer CY. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* 1956;**12**: 307-10.
29. Krippendorff K. *Content Analysis: An Introduction to its Methodology.* Thousand Oaks, Sage Publications, 2013.
30. Elo S, Kyngas H. The qualitative content analysis process. *J Advanced Nursing* 2008;**62**: 107-15.
31. Barbour RS. Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *BMJ* 2001;**322**: 1115-7.
32. Wu C. An examination of the wording effect in the Rosenberg Self-Esteem Scale among culturally Chinese people. *J Soc Psych* 2008;**148**: 535.
33. Proctor E, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: An emerging science with conceptual, methodological, and training callenges. *Admin Policy Men Health* 2009;**36**: 24-34.
34. Aggarwal N. Clinical implementation of the Cultural Formulation Interview: Planning and assessment. In *The DSM-5 Handbook on the Cultural Formulation Interview* (eds R Lewis-Fernández, NK Aggarwal, L Hinton, DE Hinton, LJ Kirmayer). 191-207. Washington, DC: American Psychiatric Publishing, 2016.
35. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health, scarcity, inequity, and inefficiency. *Lancet* 2007;**370**: 878-89.
36. Fischer EP, Shumway M, Owen RR. Priorities of consumers, providers, and family members in the treatment of schizophrenia. *Psychiatric Serv* 2002;**53**: 724-9.
37. Greenhalgh T, Robert G, MacFarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;**82**: 581-629.
38. Mechanic D, McAlpine D, Rosenthal M. Are patients' office visits with physicians getting shorter? *New England J Med* 2001;**344**: 198-204.
39. Olfson M, Cherry D, Lewis-Fernández R. Racial differences in visit duration of outpatient psychiatric visits. *Arch Gen Psychiatry* 2009;**66**: 214-21.
40. Olfson M, Marcus S, Pincus H. Trends in office-based psychiatric practice. *Am J Psychiatry* 1999;**156**: 451-7.
41. Bird VJ, Boutillier CL, Learny M, Williams J, Bradstreet S, Slade M. Evaluating the feasibility of complex interventions in mental health services: standardised measure and reporting guidelines. *Br J Psychiatry* 2014;**204**: 316-21.
42. Napier A, Ancarno C, Butler B, Calabrese J, Chater A, Chatterjee H, et al. Culture and health. *Lancet* 2014;**384**: 1607-39.

Affiliations and Addresses of Co-authors

Neil Krishan Aggarwal, MD, MBA, MA
Department of Psychiatry, Columbia University and New York State Center of Excellence for Cultural Competence at the New York State Psychiatric Institute, New York, NY, USA
NYSPI, Unit 11
1051 Riverside Drive, New York, NY 10032 USA

Renato D. Alarcón, MD, MPH
Department of Psychiatry, Mayo Clinic College of Medicine, Rochester, MN, USA and Universidad Peruana Cayetano Heredia, Lima, Peru
1 Lakeside Dr. Apt. 1602, Oakland, CA, 94612, USA

Sergio Aguilar-Gaxiola, MD, PhD
Department of Internal Medicine and Center for Reducing Health Disparities, University of California, Davis, Sacramento, CA, USA
2921 Stockton Blvd., Suite 1408, Sacramento, CA 95817 USA

Kavoos Bassiri, MS. LMFT, LPCC, CGP
Richmond Area Multi-Services, Inc. and Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA
RAMS, Inc.
639 14th Avenue, San Francisco, CA 94118, USA

Marit Boiler, MPH
New York State Center of Excellence for Cultural Competence at the New York State Psychiatric Institute, New York, NY, USA
NYSPI, Unit 11
1051 Riverside Drive, New York, NY 10032 USA

Smita N. Deshpande, MD, DPM
Department of Psychiatry, Post-Graduate Institute of Medical Education and Research, Dr. Ram Manohar Lohia Hospital, New Delhi, India
Dr. Ram Manohar Lohia Hospital, President's Estate, New Delhi, Delhi 110001, India

Esperanza Díaz, MD
Department of Psychiatry, Yale University, New Haven, CT, USA
34 Park Street, New Haven, CT 06519 USA

Hanga Galfalvy, PhD
Departments of Psychiatry and Biostatistics, Columbia University and Division of Biostatistics, New York State Psychiatric Institute, New York, NY, USA
NYSPI, Unit 48
1051 Riverside Drive, New York, NY 10032 USA

Simon Groen, MA
Department of Anthropology, University of Amsterdam, Amsterdam, The Netherlands and De Evenaar | Centre for Transcultural Psychiatry, GGZ Drenthe Mental Health Care, Assen The Netherlands
PO BOX 30007, NL-9400 RA Assen, The Netherlands

Sushrut Jadhav, MBBS, MD, MRCPsych, PhD

Division of Psychiatry, University College London, London, United Kingdom
UCL Research Department of Mental Health Sciences
67-73 Rid, 2nd floor, Charles Bell House, London, United Kingdom W1W 7EJ

Laurence Kirmayer, MD
Division of Social & Transcultural Psychiatry, McGill University and Culture & Institute of Community & Family Psychiatry, Jewish General Hospital, Montreal, Quebec, Canada
4333 Cote Ste Catherine Rd., Montreal, Quebec H3T 1E4, Canada

Peter C. Lam, MPH
New York State Center of Excellence for Cultural Competence at the New York State Psychiatric Institute, New York, NY, USA
NYSPI, Unit 11
1051 Riverside Drive, New York, NY 10032 USA

Roberto Lewis-Fernández, MD, MTS
Department of Psychiatry, Columbia University and New York State Center of Excellence for Cultural Competence, and Hispanic Treatment Program, at the New York State Psychiatric Institute, New York, NY, USA
NYSPI, Unit 69
1051 Riverside Drive, New York, NY 10032 USA

David Ndetei, MBChB, DPM, MRCPsych, FRCPsych, MD, DSc
Department of Psychiatry, University of Nairobi and Africa Mental Health Foundation, Nairobi, Kenya AMHF
PO Box 4823-00100, Nairobi, Kenya

Andel V. Nicasio, MSEd
New York State Center of Excellence for Cultural Competence at the New York State Psychiatric Institute, New York, NY, USA
NYSPI, Unit 11
1051 Riverside Drive, New York, NY 10032 USA

Vasudeo Paralikar, MBBS, MD, PhD
Psychiatry Unit, KEM Hospital, Pune, India
King Edward Memorial Hospital
489, Rasta Peth, Sardar Moodliar Road, Pune, India 411011

Hans Rohlof, MD
Centrum '45, Oegstgeest, The Netherlands
Rijnzichtweg 35, 2342 AX Oegstgeest, The Netherlands

Sanjeev Sarmukaddam, PhD
Department of Biostatistics, Maharashtra Institute of Mental Health, Sassoon Govt. Hospital Campus, Pune, India

Monica Z. Scalco, MD, PhD
Department of Psychiatry, University of Toronto and Toronto Western Hospital, Toronto, Ontario, Canada
New East Wing, 9th Floor, Room 461
399 Bathurst Street, Toronto, Ontario, Canada M5T2S8

Hendry Ton, MD, MS
Department of Psychiatry and Behavioral Sciences and Center for Reducing Health Disparities, University of California, Davis and Asian Pacific Community Counseling-Transcultural Wellness Center, Sacramento, CA, USA
2230 Stockton Blvd, Sacramento, CA 95817 USA

Rob CJ van Dijk, MSc
Parnassia Psychiatric Institute, The Hague, The Netherlands
Dadelplein 1 2552 DH The Hague, The Netherlands

Johann M. Vega-Dienstmaier, MD
Department of Psychiatry, Universidad Peruana Cayetano Heredia and Hospital Nacional Cayetano Heredia, Lima, Peru
Av. José Pardo 1142-701, Miraflores, Lima 18, Peru

Mitchell G. Weiss, MD, PhD
Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute and University of Basel, Basel Switzerland
Socinstrasse 57, CH-4002 Basel, Switzerland

Joseph Westermeyer, MD, PhD
Department of Psychiatry, University of Minnesota and Minneapolis VA Medical Center, Minneapolis, MN, USA
1935 Summit Avenue, St. Paul, MN 55105 USA

Table 1. Patient sample characteristics (N=318)

Patients	Canada (n=33)		India (n=101)		Kenya (n=29)		Netherlands (n=30)		Peru (n=34)		USA (n=91)		Total (n=318)		Test Statistic (df)	p-value
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)			
Age	51.12 (15.85)		35.42 (12.85)[b]		31.97 (10.77)		41.87 (15.33)		36.50 (10.47)[b]		49.25 (13.62)[b]		41.44 (14.95)		$F(5,306) = 17.56$	**<0.001*****
Years of Education	7.53 (4.94)[a]		11.37 (4.21)		9.83 (3.37)		12.03 (4.97)		12.56 (2.83)		9.94 (4.78)[c]		10.64 (4.52)		$F(5,101) = 6.92$	**<0.001*****
	N	%	N	%	N	%	N	%	N	%	N	%	N	%		
Female	21	63.64	42	41.58	14	48.28	9	30.00	20	58.82	50	54.95	156	49.06	Fisher's Exact Test	**0.011***
Foreign-born	32	96.97	1	0.99	1	3.45	17	56.67	0	0	61	67.03	112	35.22	$\chi^2(5) = 187.75$	**<0.001*****
New to CFI clinic	33	100	101	100	0	0	20	66.67	34	100	1	1.10	189	59.43	$\chi^2(5) = 286.25$	**<0.001*****
Number of Axis I diagnoses																
0	2	6.06	3	2.97	0	0	1	3.33	0	0	1	1.10	7	2.20	$\chi^2(5) = 55.97$	**<0.001*****
1	25	75.76	91	90.10	26	89.66	12	40.00	20	58.82	49	53.85	233	70.13		
2	5	15.15	6	5.94	3	10.34	9	30.00	10	29.41	32	35.16	65	20.44		
3 or more	1	3.03	1	0.99	0	0	8	26.67	4	11.76	9	9.89	23	7.23		

a. Data unavailable for 1 participant; b. for 2 participants; c. for 10 participants

*p<0.05; **p<0.01, ***p<0.001.

Table 2. Clinician sample characteristics (N=75)

Clinicians	Canada (n=7)		India (n=21)		Kenya (n=5)		Netherlands (n=11)		Peru (n=5)		USA (n=26)		Total (N=75)		Test Statistic (df)	p-value
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)			
Age	37.57 (7.76)		34.67 (7.48)		33.40 (4.39)		43.64 (11.46)		39.60 (8.26)		40.08 (10.26)[a]		38.35 (9.12)		$\chi^2(5) = 10.39$	0.065
Years providing mental health care	10.14 (5.24)		7.48 (7.09)		3.00 (1.22)		15.55 (12.64)		6.60 (4.34)		10.16 (8.61)[a]		9.47 (8.58)		$\chi^2(5) = 13.12$	**0.022***
	N	%	N	%	N	%	N	%	N	%	N	%	N	%		
Female	6	85.71	12	57.14	3	60.00	5	45.45	1	20.00	14	53.85	41	54.67	Fisher's Exact Test	0.377
Professional Discipline																
Psychiatrist/ Psychiatry Trainee	2	28.57	15	71.43	5	100	2	18.18	5	100	8	30.77	37	49.33	Fisher's Exact Test	**<0.001*** **
Psychologist	1	14.29	2	9.52	0	0	6	54.55	0	0	12	46.15	21	28.00		
Social Worker	1	14.29	4	19.05	0	0	3	27.27	0	0	3	11.54	11	14.67		
Other Mental Health Clinician[1]	3	42.86	0	0	0	0	0	0	0	0	3	11.54	6	8.00		
Foreign-born	4	57.14	0	0	1	20.00	2	18.18	0	0	11	42.31	18	24.00	Fisher's Exact Test	**<0.001*** **
Frequency of contact with patients of different cultures																
Daily	4	57.14	12	57.14	0	0	10	90.91	1	20.00	19	73.08	46	61.33	Fisher's Exact Test	**<0.001*** **
Weekly or Monthly	0	0	9	42.86	0	0	1	9.09	4	80.00	4	15.38	18	24.00		
Seldom or Never	3	42.86	0	0	5	100	0	0	0	0	3	11.54	11	14.67		
Hours of cultural training																
< 10 hours	2	28.57	6	28.57	5	100	3	50.00[b]	0	0	3	11.54	19	27.14	$\chi^2(5) = 14.05$	**0.015***
10 – 50 hours	1	14.29	11	52.38	0	0	0	0[b]	3	60.00	11	42.31	26	37.14		
>50 hours	4	57.14	4	19.05	0	0	3	50.00[b]	2	40.00	12	46.15	25	34.71		

a. Data unavailable for 1 participant; b. for 5 participants.

1. Other clinicians: Licensed Marriage and Family Therapist (n=1 person), Social Work intern (n=1), Rehabilitation Counselor (n=1), Psychology trainee (n=2), Unspecified clinician (n=1).

*p<0.05; **p<0.01, ***p<0.001.

Table 3. Comparing feasibility, acceptability, and clinical utility of the CFI from Likert-scale debriefing questionnaires, by clinicians and patients (N=315)

	Domain				
	Feasibility M (SD)	Acceptability M (SD)	Clinical Utility M (SD)	Test Statistic (df)	p-value
Patients	1.33 (0.57)	1.27 (0.71)	1.26 (0.53)	F(2,833) = 1.41	0.246
Clinicians	0.75 (0.90) [a,b]	0.98 (0.75) [a]	0.93 (0.70) [b]	F(2,864) = 13.37	**<0.001*****
Test Statistic (df)	t(10) = 3.53	t(10) = 1.65	t(10) = 2.14		
p-value	**0.005****	0.131	0.058		

Mixed-effect models compared domain score differences within and between groups, controlling for clinicians seeing multiple patients, multiple clinicians within a site, and whether the patient seen was new to the clinic.

Data unavailable for the following parameters: Patient acceptability (n=16), Patient feasibility (n=13), Patient utility (n=5), and Clinician acceptability (n=3).

a,b. Values with paired superscripts in the same row differ significantly ($p < 0.05$) after adjusting for multiple comparisons, Tukey-Kramer test.

*p<0.05; **p<0.01, ***p<0.001.

Table 4. Practice effects on feasibility, acceptability, clinical utility, interview duration, and proportion of total interview devoted to the CFI, in successive clinician interviews using the CFI (N=316)

Number of CFI administrations	First (n=74) Mean (SD)	Second (n=68) Mean (SD)	Third (n=67) Mean (SD)	Fourth (n=42) Mean (SD)	Fifth (n=26) Mean (SD)	Sixth[1] (n=39) Mean (SD)	Beta (95% CI)	p-value
Feasibility	0.59 (1.02) †‡	0.81 (0.95)	0.72 (0.92)	0.84 (0.66)	0.72 (0.94)	0.96 (0.67)[a]	0.053 (0.003, 0.103)	**0.039***
Acceptability	1.01 (0.72) †	0.98 (0.78)	0.97 (0.76)[a]	0.98 (0.79)[b]	0.87 (0.74)	0.98 (0.70)[a]	-0.011 (-0.051, 0.029)	0.591
Clinical Utility	0.96 (0.65) ‡	0.92 (0.82)	0.84 (0.66)	0.91 (0.74)	0.98 (0.66)	1.06 (0.66)[a]	-0.013 (-0.046, 0.021)	0.458
Duration of CFI in minutes	26.44 (10.40)[c]	22.23 (9.64)[e]	22.87 (9.38)[b]	22.16 (8.77)[a]	23.42 (9.57)[a]	22.28 (8.39)	-1.017 (-1.616, -0.418)	**0.001****
Duration of full diagnostic interview in minutes	62.70 (27.41)[d]	54.26 (25.95)[f]	53.67 (23.58)[g]	48.21 (21.49)[g]	47.92 (22.55)	50.43 (28.61)[b]	-1.609 (-2.708, -0.510)	**0.004****
CFI proportion of total diagnostic interview	47.49% (21.95)[d]	47.62% (22.47)[f]	48.91% (22.72)[g]	51.67% (21.62)[g]	54.07% (17.69)[a]	51.94% (18.61)[b]	0.046% (-0.753, 0.845)	0.910

Mixed-effect model comparisons control for clinicians seeing multiple patients and multiple clinicians within a site.

1. Combines the sixth administration or greater into one group. 6^{th}=18 individuals, 7^{th}=9, 8^{th}=5, 9^{th}=4, 10^{th}=3.

a. Data unavailable for 1 participant; b. for 2; c. for 6; d. for 10; e. for 4; f. for 5; g. for 3.

†‡. Values with paired superscripts in the first-administration column differ significantly ($p< 0.05$) after adjusting for multiple comparisons, Tukey-Kramer test. No other values differed significantly within administrations.

*p<0.05; **p<0.01, ***p<0.001.

Appendix 1

The Field Trial Version of the Core Cultural Formulation Interview

1. What problems or concerns bring you to the clinic?
2. What troubles you most about your problem?
3. People often understand their problems in their own way, which may be similar or different from how doctors explain the problem. How would you describe your problem to someone else?
    3a. Sometimes people use particular words or phrases to talk about their problems. Is there a specific term or expression that describes your problem?
    3b. What is it?
4. Why do you think this is happening to you? What do you think are the particular causes of your problem?
5. What, if anything, makes your problem worse, or makes it harder to cope with?
    5a. What have your family, friends, and other people in your life done that may have made your problem worse?
6. What, if anything, makes your problem better, or helps you cope with it more easily?
    6a. What have your family, friends, and other people in your life done that may have made your problem better?
7. Is there anything about your background, for example your culture, race, ethnicity, religion or geographical origin that is causing problems for you in your current life situation? In what way?
8. On the other hand, is there anything about your background that helps you to cope with your current life situation? In what way?
9. Sometimes people consider various ways of making themselves feel better. What have you done on your own to cope with your problem?
10. Often, people also look for help from other individuals, groups, or institutions to help them feel better. In the past, what kind of treatment or help from other sources have you sought for your problem?
    10a. What type of help or treatment was most useful? Why?/How?
    10b. What type of help or treatment was not useful? Why?/How?
11. Has anything prevented you from getting the help you need—for example, cost or lack of insurance coverage, getting time off work or family responsibilities, concern about stigma or discrimination, or lack of services that understand your language or culture? What got in the way?
12. Now let's talk about the help you would be getting here. Is there anything about my own background that might make it difficult for me to understand or help you with your problem? 12a. In what way?/Why not?
13. How can I and others at our clinic be most helpful for you?
14. What kind of help would you like from us now, as specialists in mental health?

Appendix 2

Reliability of the Debriefing Instrument for Patients and the Debriefing Instrument for Clinicians

The Debriefing Instrument for Patients (DIP) and the Debriefing Instrument for Clinicians (DIC) are each composed of three domains that assess respondents' perceptions of the feasibility, acceptability, and clinical utility of the Cultural Formulation Interview. We estimated the reliability (both raw and standardized coefficient α) of each domain prior to calculating mean DIP and DIC scores. While the DIC items appeared reasonably reliable as written, the DIP items showed greater variation; psychometric evaluation led to the exclusion of two items, one in the feasibility domain and one in the acceptability domain. This document describes the procedures we followed to reach this conclusion. The psychometric analyses of the DIP clinical utility domain and all three DIC domains are not presented in this document, since no differences between raw and standardized α's were observed; all of the domains had adequate reliability, item correlations with total, inter-item correlations, and changes to α by item; and no items in these domains were changed or removed. The final DIP and DIC item-based results and α coefficients are presented in Table 1 below.

Table 1. Final domain items, means, and reliability estimates for Debriefing Instrument – Patients and Clinicians[1]

Patient interviews			Clinician interviews		
Domains and items	Mean	SD	Domains and items	Mean	SD
Feasibility (n = 302) **α =0.45** (Raw = 0.45)			**Feasibility** (n = 312) **α = 0.78** (Raw = 0.77)		
09. Were easy to understand	1.37	0.72	12. Were easy to administer	0.97	0.97
11. Improved the flow of the interview	1.30	0.71	13. Were easily understood by the patient	0.56	1.14
			14. Contributed positively to the flow of my clinical interview	0.77	1.11
Acceptability (n = 299) **α = 0.48** (Raw = 0.49)			**Acceptability** (n=297) **α = 0.80** (Raw=0.79)		
13. Should be asked by every clinician.	1.14	0.99	15. Helped make the patient feel more at ease during the interview	0.91	1.01
14. Helped me feel more at ease with the interview	1.40	0.75	16. Can be incorporated by mental health clinicians into routine clinical interviews	1.06	0.87
			17. Facilitated a good assessment of cultural factors relevant to clinical care	0.95	1.01
			18. I would recommend for use by other mental health clinicians	1.08	0.84
Clinical Utility (n = 275)			**Clinical Utility** (n = 290)		
α = 0.82 (Raw = 0.82)			**α = 0.89** (Raw=0.89)		
01. Helped me explain my main concerns	1.48	0.56	01. Helped me understand the patient's cultural background	0.74	1.11
02. Helped me communicate important aspects of my background, such as religious faith and/or culture	1.20	0.83	02. Clarified the patient's ideas about the cause of the problem	0.98	1.01
03. Helped me understand how my background and current situation affect my problem	1.20	0.85	03. Clarified my understanding of the patient's symptoms and problems	0.95	1.03
04. Helped me explain what kinds of help I would like	1.36	0.69	04. Gave me confidence in the diagnosis	0.58	1.24

05. Gave me confidence that the clinician understood my situation	1.50	0.68	05. Facilitated treatment planning	0.98	1.06
06. Helped me identify things that could get in the way of my treatment	1.05	0.97	06. Helped me identify issues that could interfere with treatment adherence	1.11	0.95
07. Encouraged me to share important information that might not have been mentioned otherwise	1.21	0.93	07. Helped me identify additional aspects or dimensions of the patient's clinical problems	1.04	1.01
08. Were useful overall	1.44	0.62	08. Helped me assess the severity of the patient's clinical problems	0.77	1.11
			09. Facilitated my rapport with the patient	1.16	1.02
			10. Clarified how my perspective on the patient's presentation was similar or different to the patient's	0.79	1.10
			11. Were useful overall	1.12	0.75

[1.] Standardized α's are reported in bold, with raw α's in parentheses

All reliability calculations were conducted with SAS Version 9.4 (Cary, NC) using the CORR procedure with the ALPHA and NOMISS options. Only patients and clinicians who answered every item within a domain were included for the reliability analyses. Sample sizes increased between calculations done with the full DIP domains and the reduced DIP domains because the former excluded patients who did not provide answers for items that were ultimately removed (DIP item 10 and 12). Alpha estimates and item means were recalculated after these items were identified and removed. Subsequent domain scores are based on these reduced scales.

DIP DOMAINS ON FEASIBILITY AND ACCEPTABILITY

The Feasibility and Acceptability domain each included one negatively-worded item intended to be scored in reverse (items 10 and 12; Table 2); these two domains had the lowest standardized α's (0.18 and 0.17, respectively) after applying the reverse-scoring scheme. There was also a large difference between the raw and standardized α's of these two domains (0.07 and 0.18; 0.07 and 0.17, respectively), suggesting that the variance of at least one item within each domain was appreciably different from that of the other items (DeVellis, 2012). The Clinical Utility domain produced acceptable α values with no difference between the raw and standardized versions (both α's=0.82) and is included fully in Table 1. Such low α's for the Feasibility and Acceptability domains warranted a search for problematic items.

Table 2. Debriefing Instrument for Patients (DIP), original domain composition using all items: reliability estimates and item means.

Patient interviews			
Domains and items	Domain α	Mean	SD
Feasibility (n=298)	Raw: 0.07 Standardized: 0.18		
09. Were easy to understand		1.36	0.72
10. Took more time to share my perspective than I wanted. [r.]		-0.20	1.33
11. Improved the flow of the interview		1.30	0.72
Acceptability (n=295)	Raw: 0.07 Standardized: 0.17		
12. Were too personal. [r.]		0.28	1.23
13. Should be asked by every clinician.		1.14	1.00
14. Helped me feel more at ease with the interview		1.41	0.74

[r.] Reverse-scored

***Feasibility*:** Table 3A lists the correlation of each item of the Feasibility domain with the total correlation of the

remaining items. DIP 10 clearly stands out as unusual, producing a weak negative correlation with the other items despite reverse-coding the negatively-worded item.

Table 3. DIP individual item correlation with total, by domain

A.

DIP Feasibility	Correlation with Total	
	Raw	Standardized
DIP 09	0.197065	0.262147
DIP 10 [r.]	-0.053345	-0.053513
DIP 11	0.013819	0.102627

[r.] Reverse-scored

B.

DIP Acceptability	Correlation with Total	
	Raw	Standardized
DIP 12 [r.]	-0.088108	-0.075912
DIP 13	0.044750	0.130489
DIP 14	0.215033	0.246186

[r.] Reverse-scored

Table 4A examines the inter-item correlations more closely, displaying the correlation matrix among the three items in the domain. Item 09 appears to be moderately correlated with item 11, but weakly correlated with item 10. Item 11 also appears to be poorly correlated with item 10, as well as negatively correlated.

Table 4. DIP inter-item correlation matrix, by domain

A.

DIP Feasibility	DIP 09	DIP 10 [r.]	DIP 11
DIP 09	1.000	0.05439	0.28935
DIP 10 [r.]		1.000	-0.1403
DIP 11			1.000

[r.] Reverse-scored

B.

DIP Acceptability	DIP 12 [r.]	DIP 13	DIP 14
DIP 12 [r.]	1.000	-0.13121	-0.00803
DIP 13		1.000	0.31649
DIP 14			1.000

Table 5 presents changes to α values if an item were dropped. The α of the Feasibility domain (Table 5A) increases substantially when item 10 is removed (raw and standardized α=0.44), providing further evidence that

item 10 may not fit with the other two items. If any other item were to be removed, the α value decreases further (dropping item 11) or results in a negative value (item 09). In the latter case, the hypothetical domain composed of items 10 and 11 produces a negative α, suggesting that the items may not be measuring the same construct. The fact that the raw and standardized α's are essentially identical after dropping item 10 also suggests that the variance between item 09 and 11 is similar.

Table 5. Original DIP coefficient α after deleted item

A.

DIP Feasibility		
Deleted Item	Raw α	Standardized α
DIP 09	-0.27	-0.33
DIP 10 [r.]	0.45	0.45
DIP 11	0.09	0.10

B.

DIP Acceptability		
Deleted Item	Raw α	Standardized α
DIP 12 [r.]	0.46	0.48
DIP 13	0.01	0.02
DIP 14	-0.29	-0.30

For these reasons we excluded item 10 from DIP-Feasibility scoring. Including the patients who were excluded from the full DIP-Feasibility reliability calculation due to missing item 10, the final standardized α estimate was 0.45.

***Acceptability*:** In the Acceptability domain the reverse-scored item was also problematic. Item 12 correlated poorly and negatively with the total remaining items (Tables 3B and 4B). Alpha also improved substantially once the item was removed (Table 5B). Removing any other item resulted in an α nearly at 0 (dropping Item 13) or a negative α (dropping Item 14). As in the Feasibility domain, it appears that the reverse-scored item may not be measuring the intended construct. Dropping this item also produced very similar raw and standardized α estimates.

Therefore, we excluded item 12 from DIP-Acceptability scoring. Including the patients who were excluded from the original reliability calculation for missing item 12, the final standardized α for DIP-Acceptability was 0.48.

REFERENCE

DeVellis RF. Scale development: Theories and applications. 3rd ed. Thousand Oaks, CA: Sage Publications, Inc.; 2012.

Appendix 3. Additional patient sample characteristics of the CFI international field trial

Patients	Canada (n=33)		India (n=101)		Kenya (n=29)		Netherlands (n=30)		Peru (n=30)		USA (n=91)		Total (n=318)		Test Statistic	p-value
	N	%	N	%	N	%	N	%	N	%	N	%	N	%		
Employment Status															Fisher's Exact Test	**<0.001*** **
Employed (full- or part-time for pay)	12	36.36	45	44.55	7	24.14	7	23.33	12	35.29	16	17.58	99	31.13		
Unemployed	11	33.33	18	17.82	13	44.83	13	43.33	11	32.35	15	16.48	81	25.47		
Out of labour force	10	30.30	38	37.62	9	31.03	9	30.00	11	32.35	57	62.64	134	42.14		
Other	0	0	0	0	0	0	1	3.33	0	0	3	3.30	4	1.26		
Marital Status															Fisher's Exact Test	**<0.001*** **
Never married	2	6.06	34	34.00[a]	11	37.93	15	50.00	22	64.71	39	43.33[a]	123	38.92		
Married/living with spouse	26	78.79	60	60.00[a]	9	31.03	13	43.33	9	26.47	17	18.89[a]	134	42.41		
Separated/Divorced	3	9.09	3	3.00[a]	7	24.14	2	6.67	3	8.82	27	30.00[a]	45	14.24		
Widowed	2	6.06	1	1.00[a]	2	6.90	0	0	0	0	5	5.56[a]	10	3.16		
Other	0	0	2	2.00[a]	0	0	0	0	0	0	2	2.22[a]	4	1.27		
Primary Language																
African languages[b]	1	3.03	0	0	29	100	5	16.67	0	0	0	0	35	11.01		
Chinese languages[c]	0	0	0	0	0	0	0	0	0	0	10	10.99	10	3.14		
Dutch	0	0	0	0	0	0	14	46.67	0	0	0	0	14	4.40		
English	1	3.03	3	2.97	0	0	2	6.67	0	0	28	30.77	34	10.69		
Indian languages[d]	0	0	98	97.03	0	0	0	0	0	0	0	0	98	30.82		
Portuguese	30	90.91	0	0	0	0	0	0	0	0	0	0	30	9.43		
Spanish	0	0	0	0	0	0	0	0	34	100	50	54.72	84	26.42		
Other[e]	1	3.03	0	0	0	0	9	30.00	0	0	3	3.30	14	4.09		
Regional Race/Ethnicity-Related Characteristics[f]																
Foreign birth	32	96.97					17	56.67								

State of birth		
Andhra Pradesh	1	0.99
Assam	1	0.99
Bihar	11	10.89
Gujarat	2	1.98
Haryana	3	2.97
Himachal Pradesh	1	0.99
Madhya Pradesh	3	2.97
Maharashtra	32	31.68
National Capital/ Territory of Delhi	29	28.71
Not born in India	1	0.99
Punjab	1	0.99
Rajasthan	5	4.95
Tamil Nadu	1	0.99
Uttar Pradesh	8	7.92
Uttarakhand	2	1.98

Tribe		
Arab	2	6.90
Kalenjin	1	3.45
Kamba	5	17.24
Kikuyu	13	44.83
Kisii	1	3.45
Luhya	3	10.34
Luo	1	3.45
Somali	2	6.90
Taita	1	3.45

Race		
Mixed, primarily indigenous	14	41.18
Mixed, primarily black	2	5.88
Mixed, primarily white	16	47.06
Mixed, primarily Asian	1	2.94
White non-Hispanic	1	2.94

Race/Ethnicity		
Hispanic	54	60.00[a]

Non-Hispanic white											13	14.44[a]		
Non-Hispanic black											5	5.56[a]		
Non-Hispanic American Indian											1	1.11[a]		
Non-Hispanic East Asian											14	15.56[a]		
Non-Hispanic South Asian											2	2.20[a]		
Mixed/Other											1	1.11[a]		
Number of Patients with at Least One Diagnosis in Disorder Cluster[g]														
Anxiety Disorders	6	18.18	21	20.79	0	0	17	56.67	11	32.35	28	30.77	83	26.10
Bipolar Disorders	1	3.03	8	7.92	7	24.14	0	0	2	5.88	14	15.38	32	10.06
Depressive Disorders	23	69.70	33	32.67	3	10.34	22	73.33	19	55.88	46	50.55	146	45.91
Psychotic Disorders	1	3.03	11	10.89	15	51.72	3	10.00	2	5.88	25	27.47	57	17.92
Substance Disorders	3	9.09	9	8.91	4	13.79	2	6.67	4	11.76	14	15.38	36	11.32
Other Disorders	3	9.09	20	19.80	3	10.34	6	20.00	9	26.47	10	10.99	51	16.04

a. Data unavailable for 1 participant.
b. Fular, Kirundi, Kiswahili, Moroccan, Moroccan Arabic, Rwandese, and Wolof
c. Cantonese and Mandarin
d. Gujarati, Hindi, Marathi, Punjabi, Tamil, Telgu, and Urdu
e. Arabic, Armenian, Bosnian, Dari, French, Hmong, Indonesian, Ingushetian, Kurdish, and Turkish
f. As there is no standard for reporting race and ethnicity in international trials, we instead report salient demographic factors as identified by local sites and recognized by governments. Therefore, not all factors will be relevant to every country.
g. Diagnoses made after conducting CFI and a diagnostic interview. Patients can have multiple diagnoses so percentages will sum to over 100%.

*p<0.05; **p<0.01, ***p<0.001.

Appendix 4. Cross-national comparison of feasibility, acceptability, and clinical utility of the CFI

	Canada (n=33)	India (n=101)	Kenya (n=30)	Netherlands (n=30)	Peru (34)	USA (n=91)	Overall (n=315)	Statistic F(5)	p-value
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)		
Patients									
Feasibility	1.50 (0.51)	1.32 (0.47)[a]	1.75 (0.42)[a,b,c,d]	1.02 (0.66)[c,e]	1.21 (0.70)[d]	1.30 (0.58)[b,e]	1.33 (0.57)	4.32	**0.001****
Acceptability	1.23 (0.77)[a]	1.21 (0.67)[b]	2.00 (0)[a,b,c,d,e]	0.76 (0.83)[c,f]	1.23 (0.67)[d]	1.295 (0.63)[e,f]	1.27 (0.71)	9.34	**<0.001*****
Clinical Utility	1.49 (0.43)	1.17 (0.49)	1.56 (0.34)[a]	0.99 (0.56)[a]	1.29 (0.42)	1.27 (0.60)	1.26 (0.53)	2.60	**0.026***
Clinicians									
Feasibility	0.24 (0.83)	0.88 (0.91)	1.36 (0.49)	0.52 (0.94)	0.62 (0.76)	0.75 (0.92)	0.75 (0.90)	2.30	**0.046***
Acceptability	0.44 (0.85)	0.96 (0.80)	1.47 (0.50)	0.66 (0.69)	1.07 (0.52)	1.09 (0.65)	0.98 (0.75)	2.13	0.063
Clinical Utility	0.27 (0.81)[a,b,c]	1.02 (0.65)[a]	1.45 (0.43)[b]	0.58 (0.59)	1.28 (0.43)[c]	0.90 (0.67)	0.93 (0.70)	4.26	**0.001****
Duration (min)									
CFI	26.89 (7.73)	23.05 (10.89)	29.21 (3.05)[a,b]	18.82 (8.78)[a]	19.68 (8.59)[b]	23.69 (9.24)	23.41 (9.57)	3.12	**0.01***
Total Interview	84.39 (26.45)[a,b,c,d]	43.43 (17.07)[b,e]	37.57 (3.80)[c,f]	88.18 (29.20)[e,f,g,h]	37.83 (13.52)[d,g]	54.04 (16.68)[a,h]	54.12 (25.61)	22.46	**<0.001*****
CFI proportion of total interview	35.20% (15.53)[a]	53.79% (18.49)[b]	77.37% (2.86)[a,c,d]	23.61% (12.00)[b,c,e,f]	54.29% (14.55)[e]	47.67% (20.42)[d,f]	49.52% (21.41)	9.43	**0.001*****

Mixed-effect model comparisons control for clinicians seeing multiple patients and whether the patient was new to the clinic. No adjustment for site was included given collinearity between site and country.

Exact N's vary for each row, due to missing data. Data available upon request.

*p<0.05; **p<0.01, ***p<0.001

a,b,c,d,e,f,g,h. Values with paired superscripts in the same row differ significantly (p< 0.05) after adjusting for multiple comparisons, Tukey-Kramer test.

Appendix 5. Qualitative data on reasons for feasibility, acceptability, and clinical utility, differentiated by patients and clinicians

Code	Operationalized Sub-code	Magnitude	Patients (N=318)		Representative Quotes	Clinicians (N=318)		Representative Quotes
			n=Total by Distinct Patients	N=Total by Text Coded		n=Total by Distinct Clinicians	N=Total by Text Coded	
Feasibility: *Any discussion of how the CFI can be used in service settings.*	Issues related to the CFI as a tool	Positive response	81	110	"I thought it was really good. You go from basic questions to more complex. Complex in the sense of how you feel."	30	38	"Having a set of questions clubbed together focuses or brings my attention to the cultural, background aspects of patients. It brings my attention back to these factors, which is definitely good for me in terms of a reminder."
		Neutral/ indifferent response	10	13	"I did not understand it initially but then I think I got it."	11	14	"I think it's the kind of thing I would like everybody to be trained on. When they're evaluating someone for the first time, they can use these questions towards the end of the interview to ask anything that was not elicited. I don't think I could see it being used at the beginning of the interview."
		Negative response	26	36	"It was troubling because there were a lot of questions. I don't understand. All the details confuse me."	107	274	"Compared to other diagnostic interviews I've done, I feel things got jumbled."
	Issues related to implementing the CFI within a clinical setting	Positive response	14	17	"The difference is the patience of the doctor. I didn't notice a pressure in him. I didn't feel forced."	9	9	"It can definitely be used at the intake process."
		Neutral/ indifferent response	3	3	"It didn't affect anything. I was not comfortable, but there were no major problems."	6	6	"The CFI is mostly relevant for non-language concordant services."
		Negative response	7	9	"I don't know if doctors can spend this much time with patients."	39	47	"Human resources are limited and the extra time required for this will increase patient waiting time which is already strained."
Acceptability -*Any discussion of how the CFI elicits emotions among patients and clinicians.*	Positive response		187	350	"It was a good flow. It was calm, easy, relaxed. I wasn't stressed. I wasn't nervous. I felt like I was talking to someone I knew, like a friend or something like that."	39	52	"It allows me to empathize more with the patient."
	Neutral/indifferent response		19	20	"It didn't change my thoughts or feelings."	4	4	"I feel equally comfortable [addressing cultural aspects of patient presentations] as I was prior to using CFI."

	Negative response		19	23	"It reminded me of how sad I was and how much I was suffering and worried; it made me think about my future."	10	11	"I was not at all comfortable. Even though I explained the questions to him [patient], he didn't get. So I kept wondering what more I could do."
Clinical Utility *-Any discussion of the CFI's perceived fit, relevance, or compatibility to address a specific clinical problem.*	- Diagnosis	Positive response	6	7	"I think it will be better because they will understand what my real mental illness is."	32	39	"It will help in certain patients who are a diagnostic query. For example, for this patient, initially it seemed like psychosis, but it wasn't so upon talking to the patient at length."
		Neutral/ Indifferent response	1	1	"I cannot say. Maybe it helps me. It might help the caretaker to draw conclusions."	15	15	"It doesn't seem to modify the diagnosis."
		Negative response	0	0	n/a	5	5	"It significantly lengthens interview time without getting a clear diagnostic picture."
	- Treatment	Positive response	137	202	"The CFI will help me get better treatment because it will help the team better assess me and my problem."	63	80	"What brings them to the clinic is always very important and helps get a general idea, because that sets up what they consider the problem to be and how you can engage that problem with treatment."
		Neutral/ Indifferent response	15	15	"I don't think that there will be any effects."	8	9	"I don't think it will have any impact on treatment."
		Negative response	1	1	"The questions couldn't help me improve myself."	1	1	"Cultural facts might not help in deciding the pharmacological management."
	- Role of culture in mental illness	Positive response	21	27	"The doctors need to understand that patients from this country are this way and they are that way from this other country. They are all not the same."	61	83	"This will help me understand the patient's problem extensively on the basis of cultural/ religious things."
		Neutral/ Indifferent response	2	2	"That question about our cultural background was not about rites, rituals, and religion. It didn't apply to us, so those didn't seem so helpful."	15	15	"I'm finding that there are some people, like this particular patient who I saw today, who jump on it and say, 'This is how culture has affected me,' and other people who look at me like, 'Are you crazy? I don't even know what you're talking about'."
		Negative response	11	11	"I don't think that asking people about their religious backgrounds is right."	23	24	"Different cultures have different beliefs, and to incorporate these might be difficult."
	- General information gathering (not specific to diagnosis or treatment)	Positive response	154	300	"I would say in this interview that today we touched on a lot of things that I would have taken many different sessions to discuss with my talk therapist or psychiatrist. So there's a lot more personal	106	159	"I learned how he being able to talk about his symptoms opened up his family to talking about the symptoms."

					information in a shorter amount of time.”			
		Neutral/ Indifferent response	24	26	“It seems like kind of the same; the types of interviews I’ve given in the past are very similar.”	10	11	“The CFI provides some psychosocial information, but that is already assessed in a good clinical history taking.”
		Negative response	6	6	“The only thing is that sometimes one needs to investigate a little about the problem, about your mental or emotional problem. And those are the things that I don't like to talk about very frequently. I don't like to talk much about the problem."	17	21	“I think they need to streamline [the question about] the groups that have been helpful or not helpful. I think there’s almost too much emphasis on outside groups as opposed to what the individual has experienced.”

n/a - not applicable