Strategies to improve the quality and usefulness of mental health trials in humanitarian settings

Kenneth E Miller, Andrew Rasmussen, Mark J D Jordans

Lancet Psychiatry 2023; 10: 974–80

Published Online October 22, 2023 https://doi.org/10.1016/S2215-0366(23)00273-0

This online publication has been corrected. The corrected version first appeared at thelancet.com/psychiatry on October 27, 2023

Department of Educational and Counselling Psychology, and Special Education, Faculty of Education, University of British Columbia, Vancouver, BC, Canada (K Miller PhD); Department of Psychology, Fordham University, New York City, NY, USA (Prof A Rasmussen PhD); Health Service and Population Research, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK (Prof M J D Jordans PhD); Amsterdam Institute of Social Science Research, University of Amsterdam, Amsterdam, Netherlands (Prof M J D Jordans)

Correspondence to: Dr Kenneth E Miller, Department of Educational and Counselling Psychology, and Special Education, Faculty of Education, University of British Columbia, Vancouver, BC V6T1Z4, Canada
kenneth.miller@ubc.ca

See Online for appendix

A striking rise in the number of people affected by humanitarian crises has led to an increase in mental health and psychosocial support interventions to reduce the psychological effects of such crises. In a parallel trend, researchers have brought increased methodological rigour to their evaluation of these interventions. However, several methodological issues still constrain the quality and real-world relevance of the existing evidence base. We examine five core challenges in randomised controlled trials of mental health and psychosocial support interventions with conflict-affected and disaster-affected populations. These challenges are: translating intervention effects into metrics of real-world significance; giving adequate consideration to the selection and monitoring of control conditions; following rigorous processes to ensure outcome measures are culturally appropriate and psychometrically sound; ensuring and monitoring implementation variables, including fidelity, exposure, participant engagement, and the competence of implementation staff; and assessing mechanisms of change.

Introduction

Over the past decade, there has been a remarkable increase in the development and evaluation of mental health and psychosocial support interventions for survivors of humanitarian crises.[1–3] This increase reflects an escalation in the number of people affected by armed conflicts[4] and natural disasters,[5] and a consensus that humanitarian crises are causally linked to a host of negative mental health outcomes.[6]

At the same time, researchers have brought greater methodological rigour to the evaluation of mental health and psychosocial support interventions in humanitarian settings. These improvements include adequately powering randomised controlled trials to examine primary outcomes, pre-registering study protocols, making datasets open access, and using various study designs to address the complexity of such settings.

Despite methodological progress, the strength of evidence from trials of mental health and psychosocial support interventions in humanitarian contexts remains limited by insufficient attention to several issues.[1,3] These issues, which have received extensive consideration in other areas of psychological research, include the real-world meaningfulness of statistically significant intervention effects; the effect of different control conditions on the magnitude of between-group differences; confusion regarding essential steps to ensure that outcome measures are psychometrically sound in culturally diverse populations; the relationship of implementation factors, such as fidelity, exposure, and engagement, to trial outcomes; and the identification of mechanisms of change within interventions. Because these topics have not been adequately addressed, systematic reviews and meta-analyses cannot draw strong conclusions about the effectiveness, clinical relevance, or mechanisms of change of mental health and psychosocial support interventions in humanitarian settings.

The aim of this Personal View is to examine these methodological topics, drawing on evidence from other areas of intervention research to suggest ways in which trials of mental health and psychosocial support interventions might enhance both the strength and real-world relevance of the evidence they provide. We recognise a shift within humanitarian research towards greater consideration of these topics, and draw on exemplars to illustrate innovative approaches. We also recognise, from our first-hand experience, that conducting mental health trials in humanitarian settings is uniquely challenging due to conditions of persistent instability and adversity. Our aim is to support and strengthen the efforts of researchers conducting randomised controlled trials in such settings, not to criticise those undertaking this challenging work. We also acknowledge the ongoing debate regarding the primacy of the randomised controlled trial as a mechanism for generating evidence of effectiveness.[7] Our aim is to strengthen randomised controlled trials in humanitarian settings, not to advocate for their superiority relative to other methodologies.

We include recommendations for addressing the issues we discuss (panel) and provide additional examples of trials that have effectively managed these issues (appendix pp 1–3).

Assessing the real-world relevance of study outcomes

There is an emerging consensus that significance testing and p values, by themselves, provide little information about the clinical significance and real-world meaningfulness of intervention effects.[8] Reporting effect sizes with corresponding confidence intervals is an improvement, and is now common practice in reports of mental health and psychosocial support trials. However, there is little clarity concerning effect sizes in terms of their real-world significance.[8] Many trials of mental health and psychosocial support interventions use Cohen's original benchmarks (0·2, 0·5, and 0·8 for small, medium, and large, respectively).[9] However, Cohen acknowledged that these benchmarks were not empirically derived.[9] Funder

Panel: **Recommendations for addressing five key methodological issues in mental health and psychosocial support trials in humanitarian settings**

Assessing the real-world relevance of study outcomes

- Reports of trials can be strengthened by going beyond the designation of intervention effects as small, medium, or large on the basis of generic benchmarks, and providing additional information about how meaningful trial findings are in actual practice.
- Researchers should consider various metrics that allow them to contextualise their findings, so that effect sizes can be assessed relative to benchmarks of real-world significance. Examples of such metrics include the Reliable Change Index, the number needed to treat, and improvements in functioning and quality of life, bearing in mind the need to designate a priori how much change or improvement will be considered clinically meaningful. Other simple, yet clinically useful, metrics include bar charts and scatterplots that illustrate the heterogeneity of treatment effects, and the percentages of participants whose symptoms improved, did not change, or worsened.
- When possible, trials should aim to assess the maintenance of intervention effects using long-term follow-ups, ideally 6 months to 1 year, or longer. We recognise the complexity of conducting long-term assessments in conditions of instability and ongoing adversity, and also appreciate that such conditions might contribute to a gradual attenuation of intervention effects. The latter underscores the importance of examining ways of maintaining effects, for example through the provision of booster sessions.

Thorough and transparent consideration of control conditions

- The rationale for selecting particular control conditions should be an element of all trial protocols and reports.
- The monitoring and reporting of help-seeking behaviours in no-treatment, waitlist control, and treatment-as-usual conditions should be routine.
- Researchers should consider the potential effect of control conditions on trial outcomes when reporting their findings. Systematic reviews and meta-analyses should likewise consider the potential effect of control conditions on trial outcomes, particularly given the preponderance of waitlist control and no-treatment control conditions, both of which have been shown to be associated with larger intervention effects relative to active control conditions.

Measurement

- Reports of trials should include reliability and validity statistics for study samples, and not rely solely on psychometrics on the basis of research on other populations. We recognise that validating assessment tools can present logistical as well as funding challenges and might not always be feasible. Although we believe that there is a growing openness among donors to support the validation of measures as part of their funding for trials (appendix p 2), in situations where validation is not possible, measures should be selected that are valid in culturally similar populations.
- Descriptions of cultural-conceptual translation, as well as linguistic translation of outcome measures, should be considered in trial reports. Measures should be assessed through cognitive interviewing and pilot testing, with adaptation as needed.

Implementation

- In addition to fidelity and attendance, reports of randomised controlled trials could be strengthened by including data on participant engagement with intervention activities that are theoretically linked to key outcomes.
- Reports of trials should clearly describe the criteria for selecting implementation staff, and how their competence was monitored during the trial.
- To facilitate the eventual scale up of effective interventions into everyday practice, we encourage researchers to adapt trial designs to answer questions related to implementation science.

Examining mechanisms of change

- Research designs should examine mechanisms of change within established and novel interventions, potentially by developing new tools to assess the degree to which intervention-specific components mediate outcomes.
- Research on the effectiveness of interventions that use task-shifting should examine the competence of non-specialist providers as a contributor to intervention outcomes.

and Ozer[10] took Cohen's precautions a step further by stating that "the terms small, medium, and large are meaningless in the absence of a frame of reference. They immediately require an answer to at least one of two questions: (a) small, medium, or large compared with what? And (b) small, medium, or large for what purpose?"

In their review of the long-term benefits of psychological interventions with refugees and asylum-seekers, Turrini and colleagues found substantial treatment effects on post-traumatic stress disorder and depression.[2] However, symptom reduction was not matched by improvements in post-treatment functioning or quality of life, underscoring the importance of context when interpreting the meaning of effect sizes. Grice and colleagues' innovative work on person-centred effects offers numerous examples of studies that yielded large effects, in which a large majority of subjects, up to 75%, did not show the hypothesised change.[11]

Numerous metrics of real-world meaningfulness exist, such as the Reliable Change Index[12] and the number needed to treat (ie, the number of individuals who would need to be treated to see meaningful improvement in one person).[13] The definition of meaningful improvement needs to be established a priori. Approaches to operationalising improvement include falling below a culturally valid clinical cutoff or no longer meeting diagnostic criteria. Improved functioning and quality of life are other readily assessed metrics that have clear real-world meaning, although how much improvement equates to meaningful change would again need to be defined a priori. A clearer sense of the helpfulness of an intervention might be illustrated with bar charts or scatterplots that illustrate the heterogeneity of effects among participants. At a minimum, reports should include the percentages among those receiving the intervention whose symptoms improved, did not change, and worsened, and by how much.

Thorough and transparent consideration of control conditions

In the past 15 years, progress has been made in understanding the effect of different control conditions on the size of effects in randomised controlled trials of psychological interventions.[14,15] Gold and colleagues offer the unsettling conclusion that "the selection of the control condition can have a greater effect on treatment outcomes than the intervention itself".[15] No-treatment controls and waitlist controls are associated with substantially larger intervention effects in clinical than active comparators.[15] There is also some evidence that waitlist control participants anticipate their eventual access to the intervention and thus discontinue help-seeking or natural coping behaviours during the study period (the so-called nocebo effect), increasing between-group differences on study outcomes.[16] Treatment-as-usual is something of an unknown, because participants randomly assigned to treatment-as-usual groups can have access to highly variable resources. In humanitarian settings, various factors can influence the help-seeking and coping resources to which people have access (eg, proximity to clinics and non-governmental organisations, availability of kinship networks and traditional healers, and opportunities for income generation). Unfortunately, the help-seeking activities engaged in by patients assigned to treatment-as-usual, and the resources to which they have access, are seldom monitored, making it unclear precisely which treatment-as-usual circumstances are being compared with an intervention.[15,17]

There is no best control condition for randomised controlled trials; rather, the control condition should reflect numerous factors, such as the stage of intervention development and evaluation, and the availability of empirically supported alternatives, among others.[15] However, published reports of trials of mental health and psychosocial support interventions seldom offer a rationale for their selection of a particular control condition. Moreover, although considerable attention is paid to what happens within the intervention group, typically little is reported about what sort of monitoring was done to document the experience of participants in the control group. For example, what help-seeking resources did participants in the control group use (or stop using) during the trial? The effect of control conditions on study outcomes is rarely considered in discussions of trial findings. Inattention to the influence of control conditions is evident in systematic reviews of mental health and psychosocial support interventions in humanitarian settings.[2,18,19] Although a majority of the studies in these reviews have used a waitlist control group or a no-treatment comparator group, the potential influence of control conditions on overestimating intervention effects is seldom considered.

Measurement

Over the past 20 years, important gains have been made in measuring outcomes in mental health and psychosocial support-related research. Recent developments include clear steps for adapting DSM-based and ICD-based questionnaires to local contexts,[20] integrating cultural constructs of distress,[21] and developing assessment tools in target languages.[22] Despite this progress, we note the persistence of several basic misunderstandings and problematic practices. Here, we discuss the most salient.

Inappropriate reporting of psychometrics

Statistics that indicate reliability and validity are properties of samples, not of specific measures. Psychometrics do not belong to questionnaires, but to responses to those questionnaires. These data vary from sample to sample, and by extension from population to population. However, many researchers still do not report indicators of reliability or validity from their own data, relying instead on other studies that report such indicators, on the basis of samples drawn from different populations, often populations with substantively different concepts of mental health. This practice is psychometrically unsound and precludes meaningful interpretation of trial outcomes.

Translating and adapting outcome measures

The settings for mental health and psychosocial support interventions are usually far outside the cultural contexts in which most mental health outcome tools were and are developed (ie, North America and Europe). The adaptation of existing measures and the development of culturally specific measures are both viable options to redress this reality. Essential to both is considering cultural constructs of distress: culturally based idioms of distress, cultural syndromes, and explanatory models. In their adaptation of the Child PTSD Symptom Scale in

Nigeria, Kaiser and colleagues found that several items were better expressed using idioms of distress rather than the original items.[23] For example, the Hausa term "having a dry heart" better captured emotional numbing than the original "unable to have strong feelings". In developing the Afghan Symptom Checklist, Miller and colleagues[22] used a storytelling task to identify culturally salient indicators of distress, including the depression-like *jigar khun* and the stress-related *asabi*. Cultural constructs of distress are central to how people understand distress, and represent clinically important information.

Conventional translation and back translation are important steps in sound measure adaptation, but are insufficient to ensure that measures are appropriate for use in a novel population. A crucial step is cognitive interviewing, in which participants are asked about their interpretation of item content and response categories.[24] This process can identify gaps between what is intended and what is understood. Cognitive interviewing can also identify items that are culturally inappropriate to ask about, or that do not have a local meaning. In research with Syrian refugees, Miller and colleagues discovered, through cognitive interviewing, that items exploring harsh parenting, included in two widely used and carefully back-translated parenting questionnaires, were perceived as provocative and would probably evoke defensive responses.[25] Kasujja and colleagues adapted the nine-question Patient Health Questionnaire for Congolese refugees[26] using a process combining expert English-to-Swahili translation, review by local mental health experts, group review, and cognitive interviewing. The adapted nine-question Patient Health Questionnaire included items that incorporated more familiar Swahili phrases than the original translation.

After cognitive interviewing, measures should be piloted to assess their psychometrics. Assessment of internal consistency can be especially helpful at this stage. If item-total correlations are small, this might indicate that participants understand items differently than intended by the researchers. Pilot testing should also include assessment of test–retest reliability, which is notably absent from many reports of mental health and psychosocial support trials,[27] since outcome measures used in randomised controlled trials are meant to capture change over time. Kohrt and Kaiser[20] provide a detailed description of adaptation and psychometric evaluation of measures in humanitarian settings.

Implementation

Implementation quality has a powerful influence on the outcomes of psychological interventions.[17] Numerous factors can affect the quality of implementation; namely, inadequate competence of delivery agents, low fidelity to programme design, insufficient exposure to the intervention, or minimal engagement with intervention activities.

Attendance and fidelity are routinely reported in mental health and psychosocial support trials. We encourage the use of planned analyses examining the relationship of these variables to trial outcomes. Useful information would be, for example, what degree of fidelity and what frequency of attendance are required to achieve desired or at least acceptable outcomes. Such information is especially useful as interventions move from what Jordans and Kohrt have called the research space to the implementation space, that is, scaled up implementation in real-world settings.[28]

In contrast to attendance and fidelity, participant engagement has received noticeably less attention in trials of mental health and psychosocial support interventions. People can attend an intervention without engaging in key activities theoretically linked to desired outcomes. In formative research on the caregiver support intervention with Syrian refugees in Lebanon, 75% of participants attended at least seven of the nine sessions. However, half of the men included in the study reported that they had not practised the stress management exercises between sessions (a theoretically key component); nearly all the women reported practising the exercises regularly. Following programme modification on the basis of participant feedback (eg, incorporation of participants' own ways of dealing with stress), men's engagement increased markedly with corresponding increases in their reports of psychological wellbeing and the frequency of positive interactions with their children, and a reported decrease in harsh parenting.[29]

Low engagement might contribute to suboptimal intervention outcomes, and can indicate the need for programme modification. Monitoring engagement should be a routine element of trial design and extent of engagement should be included in findings. Monitoring would also allow for the assessment of mediation by key intervention components that require participant engagement (eg, the practice of cognitive behavioural therapy techniques, positive parenting methods, or mindfulness activities).

Mental health and psychosocial support interventions in humanitarian settings are commonly implemented by non-specialist staff who are often members of the target community who receive brief training and some form of ongoing support. This approach to implementation, referred to as task sharing or task shifting, has been found to be effective in numerous trials.[19,30] However, the selection criteria for intervention staff and the assessment of their competence have received little attention in reports of mental health and psychosocial support trials. We know from psychotherapy research that therapist attributes are strongly associated with therapy outcomes.[31] This finding suggests that greater attention should be paid to the attributes of the individuals recruited to deliver mental health and psychosocial support interventions, and the elements of training and support to ensure competence.

In 2020, the Ensuring Quality in Psychological Support (EQUIP) platform was launched—a joint WHO–UNICEF project—to improve the competence of intervention staff and the consistency and quality of training and service delivery. EQUIP includes instruments that support the assessment of competencies, which can be used to strengthen the selection and training of providers of mental health and psychosocial support interventions.[32] The development of competency assessment tools represents an important shift towards standardising the assessment of provider competence in mental health and psychosocial support interventions.

Trial design

By investigating the degree to which implementation variables are associated with outcomes and whether improved implementation results in optimisation of care, we enter the domain of implementation science. With increasing evidence for the effectiveness of task-shifted interventions, when designing trials researchers should also consider addressing various implementation science questions. For example, what effect do different levels of fidelity, supervision, implementer competence, attendance, and engagement have on outcomes? What effect do booster sessions have on the maintenance of intervention effects? The use of effectiveness–implementation hybrid designs provides a dual focus on evaluating effectiveness and implementation outcomes (appendix p 3).[33]

Examining mechanisms of change

Research on psychotherapy has focused extensively on identifying mechanisms of change that account for client improvement.[34] This research has led to the identification of common factors that function across different therapeutic modalities to foster positive change[31] and has also helped us to understand the contribution of model-specific methods to therapy outcomes. In contrast, trials of mental health and psychosocial support interventions have primarily focused on assessing the overall effectiveness of different approaches, with only a small number of studies examining putative mechanisms of change within those approaches.[35]

Although we have no wish to discourage researchers from testing novel approaches to strengthen mental health in humanitarian settings, at this point several interventions have shown promising effects. A crucial next step is examining how existing evidence-based interventions lead to change, as this will allow fine tuning of interventions or optimisation of results. Especially in trials of novel interventions, researchers should investigate the hypothesised mechanisms of action, ideally with sufficient statistical power. This step might require the development of assessment tools that assess specific mechanisms of action to get as close as possible to assessing the contribution of intervention-specific elements to outcomes.

There are various methods for examining mechanisms of change in clinical and preventive trials. These methods include dismantling studies, in which an intervention is compared with a pared down version of itself, with the comparator omitting a theoretically key component of the full intervention. Another approach is temporal mediation analysis, in which measurement of putative mediators precedes measurement of key outcomes. Although simultaneous mediation analysis is more common in clinical trials, the assessment of temporal mediation allows for greater confidence in the inference of causality.[36] Simultaneous assessment leaves open the possibility that changes in the outcome, perhaps due to some exogenous variable, are the cause, rather than the result of changes in the theoretical mediator. Jordans and colleagues offer an illustrative example of temporal mediation in their cluster randomised controlled trial of Problem Management Plus in Nepal.[37]

Assessing the competence of delivery agents in mental health and psychosocial support interventions to ensure quality implementation would also allow researchers to examine the role of implementer attributes in accounting for intervention outcomes. Given the robust evidence for psychotherapist competence as a powerful mediator to improve therapy, it seems reasonable to hypothesise that the competence of implementation staff could have an important role in influencing outcomes of mental health and psychosocial support programming.

Discussion

In this Personal View, we have highlighted strategies to strengthen the evidence base for mental health and psychosocial support interventions for populations affected by humanitarian crises. Key points include adding metrics of clinical or real-world significance to trial reports; increasing transparency regarding the rationale for the selection of control conditions and monitoring and reporting the help-seeking behaviour of participants in control groups; ensuring that assessment tools are psychometrically and culturally sound; monitoring and reporting implementation factors, including implementor competence, implementation fidelity, and participant attendance and engagement, and examining the effect of these factors on trial outcomes; and assessing mechanisms of change in existing and novel interventions.

This set of methodological topics is by no means exhaustive. For example, trials of mental health and psychosocial support interventions in humanitarian settings (like other randomised controlled trials) are routinely powered for their primary outcomes, with subgroup analyses, either planned or post-hoc, typically underpowered and therefore unreliable for establishing effectiveness.[38] However, findings of underpowered subgroup analyses are frequently presented as evidence of effectiveness, when they are inherently exploratory (or hypothesis-generating).[38]

Search strategy and selection criteria

References were identified from the authors' personal files, and through searches for English language articles in Google Scholar using various combinations of the following search terms: "measuring impact", "Reliable Change Index", "NNT", "number needed to treat", "limits of effect size", "understanding effect sizes", "control conditions", "randomized controlled trial", "RCT", "behavioral science", "psychological interventions", "systematic review", "MHPSS", "refugees", "humanitarian", "disaster", "conflict-affected", "armed conflict", "measurement", "assessment", "psychometrics", "cross-cultural", "implementation", "fidelity", "engagement", "task-shifting", "task-sharing", "non-specialist", "mechanisms of change", and "sub-group analysis". The search was limited to English language publications from Jan 1, 2000, to April 30, 2023.

A crucial yet largely overlooked challenge is establishing the costs of implementing effective interventions. By incorporating implementation science questions into effectiveness trials (the effectiveness–implementation hybrid designs discussed earlier), researchers can assess changes in effect associated with different levels of implementation variables; this in turn would allow identification of the minimum adequate amount of training, support, fidelity, attendance, and engagement needed to achieve meaningful results, essential information for organisations operating with scarce resources. A related issue is cost-effectiveness. To secure the political will and financing required to bring interventions to scale, and thereby contribute to sustainability of service provision, researchers will need to establish the cost-effectiveness of mental health and psychosocial support interventions. Fortunately, there is a nascent trend in this direction. Examples include the PREMIUM trials by Patel and colleagues,[39] which examined the effectiveness and cost-effectiveness of task-shifting interventions for substance misuse and depression in India, and the SHARE trials by Sikander and colleagues,[40] which examined a peer-delivered intervention for perinatal depression in Pakistan and India. Establishing the cost-effectiveness of interventions can help to ensure their sustainability. Other approaches to promote sustainability include partnering with local communities, local and national non-governmental organisations, health centres, and policy makers throughout the research process, to ensure buy-in and uptake of effective interventions.

Finally, because our focus in this Personal View is on methodological issues, we have not addressed the growing discussion of ethical issues in the design and conduct of randomised controlled trials in humanitarian settings. We regard that discussion as extremely important, and refer interested readers to several excellent papers on the topic.[41–43]

The evidence base for mental health and psychosocial support interventions is steadily accruing, yet the quality of randomised controlled trials remains limited by factors that can be readily addressed. Addressing these factors is essential if we are to have a sufficient foundation for scaling up programmes to strengthen mental health in populations affected by humanitarian crises.

Contributors
This Personal View was conceptualised, drafted, and revised by KEM, AR, and MJDJ.

Declaration of interests
We declare no competing interests.

Acknowledgments
The authors thank Dr Gabriela Koppenol-Gonzalez for her contribution to the initial conceptualisation of this paper.

References

1 Bangpan M, Dickson K, Felix L, Chiumento A. The impact of mental health and psychosocial support interventions on people affected by humanitarian emergencies. March, 2017. https://fic.tufts.edu/publication-item/evidence-brief-mental-health-and-psychosocial-support/ (accessed April 15, 2018).
2 Turrini G, Purgato M, Acarturk C, et al. Efficacy and acceptability of psychosocial interventions in asylum seekers and refugees: systematic review and meta-analysis. *Epidemiol Psychiatr Sci* 2019; **28:** 376–88.
3 Purgato M, Gross AL, Betancourt T, et al. Focused psychosocial interventions for children in low-resource humanitarian settings: a systematic review and individual participant data meta-analysis. *Lancet Glob Health* 2018; **6:** e390–400.
4 United Nations High Commissioner For Refugees. Figures at a glance. Geneva: United Nations High Commissioner For Refugees, 2022.
5 World Meteorological Organization. Weather-related disasters increase over past 50 years, causing more damage but fewer deaths. Aug 31, 2021. https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer (accessed April 30, 2023).
6 Charlson F, van Ommeren M, Flaxman A, Cornett J, Whiteford H, Saxena S. New WHO prevalence estimates of mental disorders in conflict settings: a systematic review and meta-analysis. *Lancet* 2019; **394:** 240–48.
7 Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018; **210:** 2–21.
8 Bakker A, Cai J, English L, Kaiser G, Mesa V, Van Dooren W. Beyond small, medium, or large: points of consideration when interpreting effect sizes. *Educ Stud Math* 2019; **102:** 1–8.
9 Cohen J. Statistical power analysis for the behavioral sciences, second edition. Hillsdale: Lawrence Erlbaum Associates, 1988.
10 Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. *Adv Methods Pract Psychol Sci* 2019; **2:** 156–68.
11 Grice JW, Medellin E, Jones I, et al. Persons as effect sizes. *Adv Methods Pract Psychol Sci* 2020; **3:** 443–55.
12 Jacobson NSYT, P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; **59:** 12–19.
13 Nosè M, Ballette F, Bighelli I, et al. Psychosocial interventions for post-traumatic stress disorder in refugees and asylum seekers resettled in high-income countries: systematic review and meta-analysis. *PLoS One* 2017; **12:** 1–16.
14 Mohr DC, Spring B, Freedland KE, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychother Psychosom* 2009; **78:** 275–84.
15 Gold SM, Enck P, Hasselmann H, et al. Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. *Lancet Psychiatry* 2017; **4:** 725–32.
16 Furukawa TA, Noma H, Caldwell DM, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand* 2014; **130:** 181–92.

17 Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol* 2008; **41:** 327–50.

18 Bangpan M, Felix L, Dickson K. Mental health and psychosocial support programmes for adults in humanitarian emergencies: a systematic review and meta-analysis in low and middle-income countries. *BMJ Glob Health* 2019; **4:** e001484.

19 Barbui C, Purgato M, Abdulmalik J, et al. Efficacy of psychosocial interventions for mental health outcomes in low-income and middle-income countries: an umbrella review. *Lancet Psychiatry* 2020; **7:** 162–72.

20 Kohrt BA, Kaiser BN. Measuring mental health in humanitarian crises: a practitioner's guide to validity. *Confl Health* 2021; **15:** 72.

21 Kohrt BA, Rasmussen A, Kaiser BN, et al. Cultural concepts of distress and psychiatric disorders: literature review and research recommendations for global mental health epidemiology. *Int J Epidemiol* 2013; **43:** 365–406.

22 Miller KE, Omidian P, Quraishy AS, et al. The Afghan symptom checklist: a culturally grounded approach to mental health assessment in a conflict zone. *Am J Orthopsychiatry* 2006; **76:** 423–33.

23 Kaiser BN, Ticao C, Anoje C, Minto J, Boglosa J, Kohrt BA. Adapting culturally appropriate mental health screening tools for use among conflict-affected and other vulnerable adolescents in Nigeria. *Glob Ment Health* 2019; **6:** e10.

24 Willis GB. Analysis of the cognitive interview in questionnaire design. Oxford: Oxford University Press, 2015.

25 Miller KE, Arnous M, Tossyeh F, et al. Supporting Syrian families displaced by armed conflict: a pilot randomized controlled trial of the caregiver support intervention. *Child Abuse Negl* 2020; **106:** 104512.

26 Kasujja R, Bangirana P, Chiumento A, et al. Translating, contextually adapting, and pilot testing of psychosocial and mental health assessment instruments for Congolese refugees in Rwanda and Uganda. *Confl Health* 2022; **16**: 17.

27 Christy S, Siriwardhana C, Lohmann J, Roberts B, Smith S. Quality of mental health questionnaires in conflict-affected adult populations in low and middle income countries: a systematic review. *J Migr Health* 2021; **4:** 100068.

28 Jordans MJ, Kohrt BA. Scaling up mental health care and psychosocial support in low-resource settings: a roadmap to impact. *Epidemiol Psychiatr Sci* 2020; **29:** e189.

29 Miller KE, Ghalayini H, van den Broek M, et al. Strengthening parenting among conflict-affected parents: development of the caregiver support intervention. *Glob Mental Health* 2020; **7:** e14.

30 Singla DR, Kohrt BA, Murray LK, Anand A, Chorpita BF, Patel V. Psychological treatments for the world: lessons from low- and middle-income countries. *Annu Rev Clin Psychol* 2017; **13:** 149–81.

31 Norcross JC. The therapeutic relationship. In: Dunca BL, Miller SD, Wampold BE, Hubble MA, eds. The heart and soul of change: delivering what works in therapy. 2nd edn. Washington, DC: American Psychological Association, 2009: 113–41.

32 Kohrt BA, Schafer A, Willhoite A, et al. Ensuring quality in psychological support (WHO EQUIP): developing a competent global workforce. *World Psychiatry* 2020; **19:** 115–6.

33 Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care* 2012; **50:** 217–26.

34 Cuijpers P, Reijnders M, Huibers MJH. The role of common factors in psychotherapy outcomes. *Annu Rev Clin Psychol* 2019; **15:** 207–31.

35 Bosqui T, Mayya A, Al-Kadasi A. Mechanisms of change for child mental health and psychosocial support in conflict settings: a systematic review. *Intervention* 2022; **20:** 161–69.

36 Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol* 2007; **3:** 1–27.

37 Jordans MJD, Kohrt BA, Sangraula M, et al. Effectiveness of group problem management plus, a brief psychological intervention for adults affected by humanitarian disasters in Nepal: a cluster randomized controlled trial. *PLoS Med* 2021; **18:** e1003621.

38 Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015; **351:** h5651.

39 Patel V, Chisholm D, Rabe-Hesketh S, Dias-Saxena F, Andrew G, Mann A. Efficacy and cost-effectiveness of drug and psychological treatments for common mental disorders in general health care in Goa, India: a randomised, controlled trial. *Lancet* 2003; **361:** 33–39.

40 Sikander S, Ahmad I, Atif N, et al. Delivering the Thinking Healthy Programme for perinatal depression through volunteer peers: a cluster randomised controlled trial in Pakistan. *Lancet Psychiatry* 2019; **6:** 128–39.

41 Carlson C, Sweetland A, Wainberg M. Ethical challenges in global mental health clinical trials. *Lancet Psychiatry* 2018; **5:** 866–67.

42 Deps PD, Rezende I, Andrade MAC, Collin SM. Ethical issues in research with refugees. *Ethics Med Public Health* 2022; **24:** 100813.

43 Makhoul J, Chehab RF, Shaito Z, Sibai AM. A scoping review of reporting "Ethical Research Practices" in research conducted among refugees and war-affected populations in the Arab world. *BMC Med Ethics* 2018; **19:** 36.