german cooperation
DEUTSCHE ZUSAMMENARBEIT

Implemented by:

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

HUMANITARIAN COLLABORATIVE
AT THE UNIVERSITY OF VIRGINIA

Psychosocial Support Interventions in the Context of Displacement

Recommendations for Practice, Research, and Policy from a Review of the Global Literature

PROJECT OVERVIEW

Since the onset of the crisis in Syria, the German Federal Ministry for Economic Cooperation and Development (BMZ) has been a strong supporter of promoting an increased focus on mental health and psychosocial support (MHPSS) for conflict- and crisis-affected populations. MHPSS is defined as "any type of local or outside support that aims to protect or promote psychosocial wellbeing and/or prevent or treat mental disorder"[1]. This includes medication and psychotherapeutic treatments for mental disorders, but also a broad range of psychological and social interventions – such as cultural or recreational activities, peer groups, and positive youth development programs – for instance focused on building coping skills or social support systems to promote wellbeing.

While there is growing evidence for the effectiveness of a relatively narrow set of mental health treatments, **there are still major gaps in knowledge on understanding what comprises the most effective psychosocial interventions.** Even though these psychosocial interventions comprise the bulk of MHPSS programming, there is much less consensus around best practices and the extent to which these interventions work, how they work, and where, when, and for whom they best work. To address these questions, in 2020, GIZ, on behalf of BMZ, **launched a study led by the University of Virginia to review and synthesize the global evidence regarding the effectiveness of psychosocial interventions for populations affected by displacement.** The review **included 162 evaluation reports**, sourced from academic journals and organizational reports, covering interventions from anywhere in the world and for people who had experienced any type of involuntary displacement. It also included a wide range of study designs to gain a fuller picture of what types of interventions have been evaluated, important information about participant perspectives, and how or why interventions might or might not have worked. A more comprehensive description of the study and its findings is available from alena.mehlau@giz.de.

The following brief presents the various practice, research, and policy recommendations from the full study. They are the result of careful review of the study findings in consultation with members of an MHPSS expert steering committee, representatives of GIZ and BMZ, and stakeholders who participated in two dissemination workshops. For each recommendation, a brief rationale is provided, followed by specific suggestions of avenues for progress.

CONTENTS:

AUTHORS:

Amanda Nguyen, PhD
University of Virginia School of Education & Human Development
ajnguyen@virginia.edu

Lucy Bassett, MA, MS
University of Virginia Humanitarian Collaborative
lucybassett@virginia.edu

Kirsten Gelsdorf, MA
University of Virginia Humanitarian Collaborative
Kg8v@virginia.edu

Catherine Lee, PhD
Johns Hopkins School of Public Health
Clee147@jhu.edu

Molly Lasater, PhD
Johns Hopkins School of Public Health
Mlasater@jhu.edu

This study was commissioned by the GIZ Regional Project "Psychosocial Support for Syrian/Iraqi Refugees and IDPs" on behalf of the German Federal Ministry for Economic Cooperation and Development.

1/ Inter-Agency Standing Committee (2007). IASC Guidelines for mental health and psychosocial support in emergency settings. Available at https://www.who.int/mental_health/emergencies/9781424334445/en/

Special thanks to the steering committee accompanying the study: Nancy Baron, The Psycho-Social Services and Training Institute in Cairo; Theresa Betancourt, Boston College; Felicity Brown, War Child Holland; Rabih El Chammay, National Mental Health Program Lebanon; Ananda Galappatti, MHPSS.net; Sarah Harrison, IFRC PS Centre; Zeinab Hijazi, UNICEF; Patrick Onyango Mangen, REPSSI; Alena Mehlau, GIZ.

Recommendations for Practice

1. Clearly develop a theory of change for each intervention type

For many of the reported interventions, the rationale for how the intervention was thought to work was not clearly stated. This theoretical underpinning, or theory of change, is critical to:

- Guide intervention design
- Link intervention activities to assumptions of potential impact
- Inform subsequent decisions, such as: who the intervention should be offered to; how the intervention should be delivered (e.g., in groups, by peers, etc.); what types of outcomes should be evaluated; what unexpected outcomes could happen; what components of the intervention can or should be adapted in order to improve fit for different groups; and timing and order of activities (e.g., if trying to improve both social connectedness and distress, how do these relate to each other? Does the program aim to decrease distress first in order to increase social interaction or increase social interaction in order to decrease distress?)

THEORIES OF CHANGE

1. Connect intervention activities to expected outputs and outcomes
2. Map out relationships between different pieces of the model
3. Identify critical assumptions that need to be met (e.g., assuming that living conditions are safe and stable)
4. Can be tested and iteratively refined

Making the theory of change more explicit can help improve consistency across similar types of psychosocial interventions, supporting sharing of effective program models and lessons learned, which will reduce redundancy in intervention development and make it easier to draw conclusions across interventions.

KEY TAKEAWAY

Having a clear idea of how an intervention is thought to work helps to guide every step of intervention development, testing, and delivery. It is best to develop this from the start, but it is never too late to do so! Besides, the theory should be adapted over time as new findings and experiences emerge, helping to support an organization's learning process.

Developing a theory of change is not a quick or easy process, but there are a number of tools and guidance documents available to help.[2,3] Critically, the theory of change should be developed collaboratively by a range of stakeholders – implementers, researchers, community members and potential participants, etc. Doing so not only ensures different perspectives and knowledge are reflected but can also facilitate the development of joint understandings and expectations and can increase transparency. Ideally a theory of change is developed before an intervention begins but can also be developed for interventions that are already in use to help inform future evaluation. Besides, even after a theory of change is established, it will need to continue to be tested and revised based on new learnings. **Given the challenges identified in this review, a particularly impactful future activity would be to retroactively develop theories of change for existing programs, and to revise evaluation strategies accordingly.**

2. Better match intervention approaches to participants' needs

Interventions can be targeted to specific participants based on their psychosocial needs or provided to a broader group in a generally supportive effort to promote community and family resilience. In this review, **only ten percent of interventions were targeted based on participants' psychosocial needs** (e.g., self-reported or perceived distress). However, **some studies that looked at intervention impacts among people with different levels of distress found differences in whether and how well the intervention worked**. In some cases, the interventions appeared to be more appropriate for participants with higher distress or trauma, whereas in other cases the opposite appeared true.

WHY MATCHING MATTERS

People whose needs are poorly matched to an intervention may not benefit as much. In some cases, a poorly matched intervention could even make their problems worse.

There are several potential explanations for this: A person's ability to engage with and benefit from generally supportive interventions can be impacted by their current functioning and symptom experiences. For example, people experiencing substantial distress may not be able to integrate and engage with intervention content focused on skill building or social integration until their distress is first mitigated. On the other hand, people who are generally coping well may not need intervention activities that are more "treatment-like", such as those focused on processing traumatic experiences. At best, participants with more psychosocial needs may just not be prepared to benefit as much from these types of interventions; at worst, a poorly matched intervention could exacerbate problems or impair recovery. This highlights a need for more careful consideration of how decisions are made regarding intervention eligibility, and how interventions may be tailored to match needs. Tailoring, however, should be balanced against being overly specific, requiring a novel intervention for each specific group in a way that may drain resources and increase the potential for stigma. For example, one of the benefits of a more broadly applicable approach that does not require screening is that it can be offered to a large group while avoiding the potential for stigma. It may be useful to explore ways to maintain a combined group with differing levels of need while still tailoring to the individual needs of participants. Increased use of measures that assess other psychosocial indicators beyond distress (e.g., wellbeing, functioning, coping skills) may also assist with screening while minimizing stigma. These decisions should link to the intervention's theory of change and be made collaboratively between practitioners familiar with the intervention, and representatives familiar with the needs and circumstances of the target groups.

Better understand the impact of each activity or component of an intervention

Many interventions have multiple activities or components (e.g. art, drama, communication, trauma exposure, skill development,

2/ https://www.poverty-action.org/publication/goldilocks-deep-dive-guiding-your-program-build-theory-change#:~:text=Theory%20of%20change%20is%20the,change%20outcomes%20and%20deliver%20impact.

3/ Breuer E., De Silva M, & Lund C (2018). Theory of change for complex mental health interventions: 10 lessons from the programme for improving mental healthcare. Global Mental Health (2018), 5, e24.

memory processing, social events), but are generally described – and evaluated – as a whole package, even in cases where participants may opt into only a subset of activities. Identifying the value add of each piece of an intervention could help to make psychosocial interventions shorter, more useful, and potentially less resource intensive. Ways to explore the contribution of specific intervention components could include:

- Building in brief interim measures to track changes in outcomes associated with specific activities
- Offering activities in different orders to different groups or participants to compare trajectories
- Carrying out post-intervention review of attendance, participation, and engagement in different activities to see what participants really seemed to respond to

EXPLORING DIFFERENTIAL IMPACTS IN PROGRAMMING

- Anticipate potential differences in needs and experiences when developing an intervention
- Separately engage individuals from all target audiences in intervention development and piloting to assess acceptability and fit
- Check in with representatives of different subgroups about engagement and fit during implementation
- Look at evaluation data both overall and by subgroup
- Review surprising findings with participants and providers. Try to understand their perspectives on aspects of both the **intervention model** and **implementation factors** that may have contributed to impact. For example:
  - Did both groups have the same level of attendance and engagement?
  - If not, did one group dislike the intervention or find it a poor fit? Or were there other delivery issues such as who the provider was, or what time of day it was offered?
- Use new findings and lessons learned to refine the theory of change and intervention model

3. Increase availability of gender-sensitive programming

While there were more interventions specific to the needs of women and girls, **the majority of interventions were offered without regard to gender, very few interventions were developed and implemented specifically for men or boys, and no interventions offered targeted support for LGBTQI+ persons.** However, some study findings highlight a need for further attention to gender in programming. For example, findings from classroom-based interventions and child friendly spaces showed differential impacts in boys vs. girls, either in terms of size of the intervention impact (often favoring girls) or the outcomes that were impacted, suggesting that the same intervention may work differently for boys and girls. We also noted that many parenting studies tended to target mothers, and in at least one case proved to be a poorer fit for fathers until further adaptations were made to increase the perceived relevance of intervention activities for men. Another study found an intervention to increase social support for women, but not for men.

Differences in acceptability or impacts by gender may be due to the appropriateness of activities or intervention approaches to gendered roles, experiences, and interactions. Differences may also be due to gender differences in factors such as trauma exposure, displacement stressors, experiences or expressions of distress, and coping mechanisms. **It is important to understand and explore not only where differences in impacts by gender are found, but why, in order to improve intervention design and implementation.**

4. Adopt a developmental/lifespan approach

For many interventions, it makes sense to cover as wide an age range as possible. However, in doing so, important developmental differences may be overlooked. As with the gender findings described above, in some studies stronger impacts were observed for younger vs. older children/youth, suggesting either different needs or a different level of acceptability or participation in the intervention. In other studies, supports were offered to both adults and children, but data was only collected from adults or findings were not broken down by age. **There was particularly limited data on intervention impacts for the youngest children** (intervention evaluations for this group focused more on parental skill building with less attention to child-specific outcomes).[4] There also appeared to be very few interventions specifically developed for older adults, who may have different support needs or challenges to engaging in interventions, and may respond to interventions differently than others. For example, older adults may have physical challenges (e.g., mobility, hearing) that could impact their ability to participate. They may also have different social roles or emotional experiences, fewer social connections, and more difficulty learning to navigate a new place. Lifespan perspectives that seek to understand roles, priorities, and social and emotional processes at different ages would support interventions that more closely target developmental stages. When offered across multiple stages, exploration of subgroup differences is particularly important.

5. Balance the momentum toward a minimum service package with locally/contextually rooted interventions

There is substantial momentum toward the development of a minimum MHPSS service package for humanitarian settings. **Coming to consensus on a minimum set of activities would enable progress toward the development of standard theories of change and evaluation frameworks to begin comparing "like" approaches across different target groups, contexts, and implementation strategies, which would help to address many of the knowledge and practice gaps highlighted in the review.** However, it is quite likely that even with a standard package, substantial adaptations will always be needed to ensure appropriate fit, feasibility, and cultural and ethical consideration.

There is also the risk that "ownership" of the minimum service package, and thus influence on what is included, is dominated by large international organizations or donors without sufficient representation from local communities. **A robust psychosocial support response must recognize that such a package really is the minimum; to be enriched and complemented by locally developed, community-based interventions.** To facilitate an inclusive, culturally, and contextually informed process, the development

4/The lack of data for young children is likely in part due to measurement challenges; refer to this mapping exercise (http://www.cpcnetwork.org/wp-content/uploads/2015/06/Measuring-Child-MHPSS-in-Emergencies_CU_Mapping-Report_March-2014.pdf) and compendium of tools (http://www.cpcnetwork.org/wp-content/uploads/2014/06/Measuring-Child-MHPSS-in-Emergencies_CU_Compendium_March-2014-.pdf) for measuring child MHPSS in humanitarian contexts.

of a minimum service package must be carried out with extensive stakeholder consultation and be accompanied by clear guidance on what and how to adapt programs to local circumstances.

KEY TAKEAWAY
Even with a minimum service package that ensures a similar set of interventions, substantial adaptations will always be needed.
Moreover, a robust psychosocial support response must recognize that such a package really is the minimum; to be enriched and complemented by locally developed, community-based interventions.

Recommendations for Research and Evaluation

1. Whenever possible, include a comparison condition in research and evaluation – even if that means getting creative!
A key finding of this review is that **qualitative perceptions of interventions are generally positive. However, when comparing change measures within a group of people receiving an intervention to a similar group of people not receiving an intervention, the difference between the two groups was often no longer meaningful.** These findings suggest that identified changes may not be a result of the intervention itself, but potentially other individual or contextual factors, for example, natural improvements over time or due to improved security or living conditions. To know for sure, it is necessary to have comparison conditions whenever possible.

There is a clear tension between a need to provide services – or to avoid withholding services that might be helpful – and a need to scientifically test the impact of these services. Certainly, there are concerns about the fit of the historically "gold standard" randomized controlled trial (RCT) with the realities of displacement, and recognition that RCTs alone often do not provide the rich qualitative and implementation data that helps provide insight into important questions around mechanisms of action, stakeholder perceptions, unintended impacts, etc. Some interventions also lend themselves more readily to the traditional RCT, whereas particularly with regard to the less clinical, "low threshold" interventions (e.g., integrated MHPSS, psychological first aid, etc.) there can be substantial barriers to randomization.

These challenges highlight the need not only to increase use of comparison conditions when feasible, but also to think creatively about alternative ways to evaluate interventions when randomization is not an option. As an encouraging finding, the inclusion of studies in our meta-analysis that were not fully experimental (e.g., included allocation to groups that was not fully by chance) did not appear to change the results. With caution, this provides support for the feasibility of using well thought out alternative designs[5] that move away from the traditional RCT. However, it is important to clearly state the reasons for choosing to go with alternative study approaches, and test assumptions about how comparable non-randomized groups are (e.g., by comparing baseline characteristics, comparing access to other services during the intervention period, etc.). Moreover, certain comparison approaches cannot be used to draw conclusions that observed differences were due to an intervention. For example, when looking at people who seek help and people who do not, even if they appear to be similar, there are quite possibly differences between these groups explaining their help-seeking behavior.

GETTING CREATIVE WITH EVALUATION
Randomly assigning participants to receive an intervention or not is not always feasible, or ethical. But there may be other creative approaches to move beyond single group evaluations, like:

- Offering more vs. less intensive supports
- Comparing groups across areas where an intervention is and is not [yet] available, with purposeful selection or matching to make the groups look similar in other key characteristics
- Giving different groups of people different parts of an intervention, or in a different order
- Comparing groups in which everyone receives the intervention, but other aspects are changed such as how it is delivered, or by whom, or with different session frequency
- Delivering the intervention in sections or modules, with break periods in between, and looking at participants'changes during active participation vs during break periods
- Whether these approaches, or others, would make sense or not depends on how the intervention is packaged and thought to work.

2. Standardize reporting of key study features
Consistent, standard reporting is critical for evidence reviews that try to combine lessons across studies. For many of the variables we sought to examine we encountered challenges with inconsistent reporting across studies. There were also concerns that study-level characteristics may not reflect the range of individual participant experiences; for example, that even if we were able to quantify something such as level of discrimination in a particular context, individuals themselves would have unique discrimination experiences that would make the overall characterization less meaningful. **A substantial contribution to the field would be generation of consensus-based guidance around standard reporting to be included in future research.** Regarding impact indicators, this could follow the approach and guidance laid out in the IASC M&E Framework.[6] Beyond impact indicators, this could also include more standardized collection and reporting of context and population characteristics, implementation factors, intervention components, and analyses. Such standardized reporting for even small collections of shared features across studies would make it more feasible in the future to combine data from multiple studies to learn more about individual differences in experiences. It would also be useful for creating databases that focus on specific intervention activities. Standardization in reporting need not undermine context-specific reporting, for instance by combining consistent reporting of key features with flexible

5/ Stern E, Stame N, Mayne J, Forss K, Davies R, Befani B (2012). Broadening the range of designs and methods for impact evaluations (Department for International Development (DFID) Working Paper 38). London, UK: DFID. https://www.oecd.org/derec/50399683.pdf.

6/ Inter-Agency Standing Committee. (2017). Mental Health and Psychosocial Support in Emergency Settings: A common monitoring and evaluation framework. https://interagencystandingcommittee.org/iasc-reference-group-mental-health-and-psychosocial-support-emergency-settings/iasc-common

reporting of features specific to an intervention or context (e.g., for all interventions delivered through the health sector, include a description of how the health services are operated).

3. Link outcomes and measures to the stated theory of change

More clarity is needed for many studies around issues of **which outcomes are chosen, and why** (i.e., how the selected indicators are appropriate choices against which to evaluate whether the intervention "works"?), **which outcomes we definitely would NOT want or expect to see from an intervention** (we can then study those outcomes to check against unintended consequences) and **priority outcomes** (studies often evaluate many different outcomes; where do we definitely want to or expect to see change, whereas what outcomes are being measured more on an exploratory basis?), and **timing** (which outcomes are thought to come first, which later?).

The most commonly measured outcomes were those focused on indicators of distress (e.g., symptoms of post-traumatic stress), many of which fall outside what would be expected to be impacted by a psychosocial intervention. In some instances, it may be reasonable to include such indicators: For example, early changes that may lead to later changes, such as an intervention that aims to improve parent-child interactions by reducing parental distress and improving parental functioning. In other instances, indicators of distress may be more downstream effects, such as in a peer support group for which the primary goal is to increase social connectedness, through which the longer-term impact is thought to be lower symptoms of depression. Still other reasons to measure distress would be to assess different impacts of an intervention on subgroups that may need different levels of support. Those considerations should ideally be laid out in the theory of change.

KEY TAKEAWAY

More clarity is needed around which outcomes are chosen, and why; which are primary or priority outcomes vs. exploratory outcomes; and anticipated timing of expected outcomes.

4. Improve testing of perceived intervention mechanisms

Research and evaluations should be designed to test and iteratively refine theories of change. This could include:

- Use of study designs that build in midway data collection to test assumptions about the timing of changes. For example, if the theory is that the intervention decreases distress by increasing coping skills, perhaps knowledge and use of coping skills should be measured repeatedly over the intervention period, with distress measured at the end.
- Use a mixture of data collection methods to shed light on quantitative findings through the inclusion of stakeholder perspectives such as with focus groups or interviews.
- For more complex or multi-component interventions, evaluation designs that break the intervention into different components, or only offer certain components can be helpful for exploring questions about how various parts of the intervention work together to produce impacts, or whether certain activities are not actually needed.
- Designs that combine outcome evaluations with the study of different delivery approaches (e.g., group lessons vs. self-study) are also increasingly relevant to distinguish between intervention- and implementation characteristics contributing to effectiveness. For example, to what extent is it the relationship or interpersonal interaction that makes the difference, vs. the actual content being delivered?
- Another useful approach when reviewing evidence across studies could be analysis of intervention components focused on identifying critical intervention features;[7] **a particularly impactful area for future research would be to revisit the interventions in this review from a practice- or activity-level perspective to clearly map out shared and unique activities across interventions.** This would be a substantial undertaking, requiring extensive communication with intervention developers and careful coding by mental health practitioners familiar with how distinct therapeutic approaches manifest in various intervention activities.

KEY TAKEAWAY

Research to study mechanisms of an intervention may:

- Collect data at more points in time
- Collect both qualitative and quantitative data
- Break interventions into parts
- Study both intervention and delivery strategies

5. Include analysis and reporting on important subgroup characteristics and impacts

As described above, interventions may work more or less well for groups of people with different characteristics, such as age, gender, or psychosocial need. Most papers included in this review did not explore differences in perceptions and impacts across subgroups. The studies that examined subgroup impacts tended to be quantitative studies that allowed for statistical exploration, but this approach alone was not always able to provide insights into what led to these different outcomes. In other cases, studies may not have explored these questions due to challenges such as small sample size or the need for advanced statistical methods.

TIPS FOR QUALITATIVELY EXPLORING SUBGROUP DIFFERENCES

- Ask the same questions in focus group discussions with different subgroups, and compare responses
- Analyze qualitative data separately by subgroup, even if collected together
- Ask a range of participants targeted questions about how an intervention might fit or work for different groups
- Share subgroup findings with participants, providers, and other stakeholders to get feedback on what might have led to different results

As much as is possible, it would be helpful for future research to examine and report on subgroup differences. Even if taking an exploratory approach, building evidence in this area can contribute to hypothesis development and further testing, and over time a growing body of subgroup findings can potentially be examined together to get a better understanding of differences in impact. With regard to disaggregating age-related data, it would be most helpful to do so according to standard age groupings that could then be more easily compared across studies; for example, 0-6, 7-12, 13-18, 19-25, etc.

7/ Sutcliffe, K., Thomas, J., Stokes, G., Hinds, K., & Bangpan, M. (2015). Intervention Component Analysis (ICA): A pragmatic approach for identifying the critical features of complex interventions. Systematic Reviews, 4(1), 1–13.

Exploration of subgroup differences is also an area in which qualitative data collection approaches could be helpful to shed light on observed differences. For example, although potentially more resource intensive, it would be particularly useful to collect and analyze qualitative data separately for men and women (and non-binary), boys and girls, different ethnic groups or age ranges, and other unique characteristics such as LGBTQI+, persons with disabilities, etc. Even in the absence of differential quantitative findings, by posing similar questions to different groups, answers can then be compared to help develop hypotheses about differential mechanisms of impact. **For example, asking boys and girls separately about what they liked most about an intervention, what activities they found most helpful, and what changes they have experienced due to the intervention could highlight differences in how the two groups perceived, engaged with, and responded to the intervention.** Questions can also be posed that specifically solicit stakeholders' impressions about the extent to which interventions would fit, be acceptable to, and be effective for different subgroups. When the timeline and resources are feasible, quantitative analyses can first be completed, and then separate or joint feedback sessions held with participants from the subgroups in question to elicit their perceptions and experiences that may have led to these differential outcomes.

6. Improve measurement of non-distress-oriented outcome indicators

Measuring outcomes other than distress, such as subjective wellbeing, coping, and social connectedness, is important because these types of outcomes often fit better into the underlying framework in which psychosocial interventions operate. This is supported by our finding that the largest intervention impacts seemed to be in some of these domains. However, a challenge in our research was that **indicators of outcomes such as coping, subjective wellbeing, and social connectedness were not only less commonly measured but also measured less consistently in terms of tools used.** Because of this, the meta-analysis of psychosocial wellbeing, for example, included fewer studies. For outcomes like coping skills, social connectedness, and family processes, there was simply too much variation in how the outcome was conceptualized and measured to be able to collect a large enough group of studies for meta-analysis. In short, research on mental distress and disorder has a much larger footprint in the field — and well-tested measures — than research on non-distress-oriented outcomes. A longer history of studies in mental distress likely contributes to continued overuse of distress measures and less consistency in the use of other tools.

The IASC M&E Framework has been helpful for identifying priority domains of impact (e.g., wellbeing, coping, connectedness, behavior).[8] **Now what is critical for future research is to address measurement challenges and produce valid, consensus-based, user-friendly tools and approaches to measuring a more diverse range of outcomes, and to test these measures across multiple different kinds of settings.** This will likely need to include framework tools or approaches with guidance for local adaptation and testing. As an example, the Bolton approach to developing locally relevant functioning assessments has been widely adopted and was leveraged in many of the papers that included functioning as an outcome. A similar approach may be helpful for making progress toward measuring constructs such as happiness, subjective wellbeing, and coping across contexts.

7. Intentionally explore unanticipated and negative intervention impacts

Few studies reported unanticipated impacts, either in terms of negative change or as observed impacts that fell outside stated expectations. Worsening outcomes were almost entirely based on quantitative measures in which change or movement in a negative direction could be observed. But, as noted above, quantitative findings alone cannot provide a clear explanation for unanticipated findings.

A lack of clarity around theories of change and what impacts would or would not be expected makes it difficult to identify surprising or unanticipated findings. Many quantitative studies measured a large number of outcomes, with little indication of which were considered the primary focus vs. which ones were included to answer more exploratory questions. Qualitative studies reported perceived impacts but did not distinguish between those that were and were not expected. **It is critical that future research improve on the exploration and reporting of all unanticipated and potentially negative impacts, both to avoid doing harm and also to help refine intervention targets and theories of change.**

To address this gap, both qualitative and quantitative research is needed. At the intervention design phase qualitative research that specifically seeks to explore, through clear questioning, all positive and negative changes experienced during the intervention period and the extent to which participants attribute these to the intervention itself would be useful. At the testing phase, plausible intervention impacts must be clearly specified in advance and measured accordingly, with additional hypothetical or exploratory impacts separately noted and measured. For example, an intended outcome of a women's support group may be increased empowerment, which should be measured; however, if there is a concern that empowerment may also contribute to relationship conflict, this should also be evaluated as an exploratory outcome. Post-intervention debriefing that presents unanticipated findings back to stakeholders to help understand the potential causes is also critical. To be able to effectively communicate negative findings, it is important to build and foster networks of practitioners that support a culture of sharing failures, from which everyone could learn.

Finally, dropout and participant tracking is a pressing challenge with data collection in unstable environments, and yet such loss must be carefully explored to determine the extent to which people who complete the interventions are different – perhaps more favorably orientated toward the intervention – than those who choose not to complete the program. It is quite possible that unintended outcomes may be a reason why some people choose to stop participating.

TIPS FOR EXPLORING UNINTENDED IMPACTS

- Ask participants directly about all positive AND negative changes experienced during the intervention period…AND explore what they attribute these changes to
- Using the theory of change, clearly specify which measured outcomes are exploratory, and why
- Ask stakeholders their thoughts and ideas about what contributed to any unintended impacts
- Follow up with people who do not complete the intervention to understand how their experiences might be different

8. Increase focus and reporting of implementation factors

Beyond building evidence for effectiveness, there is a strong need for greater focus and reporting of implementation strategies and factors

8/ Bolton P, Tang AM. An alternative approach to cross-cultural function assessment. Soc Psychiatry Psychiatr Epidemiol. 2002 Nov;37(11):537-43. doi: 10.1007/s00127-002-0580-5. PMID: 12395144.

in MHPSS research,[9] and the extent to which impacts (or lack thereof) may be attributed to intervention components themselves vs. implementation factors that influence intervention impact. For example, if an intervention did not produce great change, is this because the activities were not (and would not be) helpful, or because they weren't carried out as intended? If many people dropped out, is this because they did not like the intervention, or because other issues made it difficult to attend? If a report was highly positive, does this reflect a majority experience or the experience of only those who chose to stay and complete it? As described above, in future research implementation features should be included in reporting standards.

9. Support replication of studies

Many of the included studies were descriptive, feasibility, or pilot studies reporting on early intervention development and initial evaluation. With a few exceptions of well-tested interventions, there was a lack of repeat studies /evaluations across contexts, as well as in most cases a lack of large, controlled evaluations of promising interventions. Given the preventive nature of psychosocial programs, we would anticipate many of the effects to be relatively small, which means studies often need to be quite large to be able to measure the change. In many of the included studies, too few participants may have contributed to lack of clear answers about effectiveness. To address many of the outstanding questions highlighted above, there is a strong need to advance pilot studies with promising findings to full scale evaluations, and then to replicate these evaluations in different settings.

Recommendations for Policy and Funding Agencies

1. Coordinate and align funding for psychosocial programming

Future funding for mental health and psychosocial programming should be used to reinforce many of the above recommendations to develop better theories, refine interventions, and improve fit for different groups. Investment must also include rigorous evaluation of funded programs to strengthen the available evidence for effectiveness. An example of coordination to foster progress in this area is the Health Evaluation and Applied Research Development (HEARD) MHPSS program.[10] The program currently provides funding for five psychosocial evaluation projects, four funded by USAID and one funded by GIZ's Regional Project "Psychosocial support for Syrian/ Iraqi refugees and IDPs." While the interventions are all different, and focused on different regions, project team members participate in a joint learning collaborative through which theories of change are developed and shared, outcome indicators aligned across projects, and intervention adaptations documented using a standard framework.

FUNDING DRIVES PROGRESS

Funding should be used to incentivize programming that reinforces many of the above recommendations to develop better theories, refine and streamline interventions, and improve fit for different groups to optimize impact, with investment for rigorous evaluation of funded programs to further develop the evidence base.

2. Strengthen evaluation and reporting requirements tied to programmatic funding

Funders can greatly advance progress in the field by including requirements for implementers of psychosocial programming that align with the above practice and research recommendations. Specifically, building-in requirements about stating and testing intervention theories, building-in comparison conditions and means of exploring unintended outcomes, and requiring standard reporting elements. Complementary to these requirements, sufficient budget must be allocated to support the background work and M&E capacity necessary to meet these requirements.

Critically, in the current funding environment continued resources depend on achieving, and reporting, positive results. This contributes to evaluation approaches that highlight positive findings while avoiding exploration of potential harm or unintended results. **This risk for presenting overly positive findings is a particularly concerning problem in program evaluations that collect largely qualitative data at a single point in time after an intervention.** The current review found that such data was almost entirely positive, compared to quantitative evaluations that measured outcomes before and after intervention implementation. Funders need to explicitly require exploration and reporting of not only positive impacts, but also lessons learned and opportunities for improving programming. They must also make reporting this information less risky. Certainly, identification of substantial, unintended harm would be reason to cease funding a program, but in a field where so much is still unknown, it is likely that many program evaluations will identify flaws, challenges, or undesired impacts that may be used to contribute to organizational learning and intervention improvement.

KEY TAKEAWAY

To improve evaluation and reporting, funders need to explicitly require exploration and reporting of not only the positive impacts but also lessons learned and opportunities for improving programming. They must also make reporting this information less risky.

3. Enact funding cycles that reflect the iterative nature of intervention development and testing

One of the challenges of the current review is that many of the reports were of descriptive or small-scale pilot and feasibility evaluations, often of one-off interventions without larger follow-up studies and replication. This likely reflects a tension between the brief, crisis-driven funding cycles of many policy and donor organizations, especially in humanitarian settings, and the many years that are required to take a new intervention from initial development through testing, implementation, and replication. This is especially important given the contextual challenges in crisis contexts. For example, rapidly changing situations increase the likelihood that observed changes could be due to contextual changes (e.g., improved living situation) rather than the intervention itself, which makes it even more important to have a comparison condition to sort out the intervention's effects. At the same time, populations with a high level of need may make it more difficult to include a comparison condition, either due to ethical concerns or even just social acceptability of allocating people to receive different services. Likewise, situations in flux will likely result in chal-

10/ https://www.heardproject.org/mental-health-and-psychosocial-support/

9/ Betancourt TS, Chambers DA. Optimizing an Era of Global Mental Health Implementation Science. JAMA Psychiatry. 2016 Feb;73(2):99-100.

lenges for consistent implementation, greater drop-out rates, etc. These issues take time and resources to understand. Moreover, once an intervention has some evidence there is likely a strong motivation to move away from resource-intensive evaluation activities and into standard implementation, but there is still much to be learned about transferring research and evaluation findings into sustainable practice.

Note that while humanitarian funding cycles may be very brief, the longer funding cycles in development work may be a special opportunity to support knowledge building around effective MHPSS; however, this cannot stand alone, as it is also unclear how evidence generated in one context, population, or set of circumstances would extend elsewhere (e.g., does an intervention that is effective in a more stable development context have similar impact in an acute crisis response?).

GIVE IT TIME
Taking an intervention from initial development through field testing, effectiveness evaluation, and replication in different settings takes many years and a great deal of coordination. This often does not align with short-term humanitarian response priorities. Longer development-focused funding cycles provide an opportunity to build progress toward knowledge generation, but do not negate the need to replicate interventions and evaluations in different settings.

4. Provide funding and support for intervention development, measurement, and methods research

While prioritizing funding for direct service delivery is essential, it is equally essential that funders and policy makers prioritize and recognize the need for improved methods and measures to support intervention development and evaluation. Specifically, these efforts should support theoretical and foundational work such as development of theories of change, development and testing of different data collection and measurement tools (validated for different groups and assessing a broader range of outcomes), and collaboratively building out guidance documents and frameworks with input from a wide range of stakeholders (implementers, researchers, community members, etc.). Moreover, given the challenges in conducting rigorous evaluations that are both feasible and responsive to the real-world challenges of displacement settings, more progress in research and evaluation methods is needed, both at the level of study design innovations as well as statistical approaches to addressing design limitations during analysis. Some examples of alternative study designs have been described above, yet these types of alternative designs must continue to be developed and tested.

5. Support evaluation and intervention delivery capacity in community organizations

Evaluations of psychosocial programming are still largely dominated by large-scale international organizations and academic partners, with less representation from grassroots and community organizations. To improve program evaluation efforts, ensure intervention delivery aligns with research goals and strategies, empower community organizations to have an influence on the global research agenda, and increase community involvement in generating evidence for promising psychosocial programs, there is a need to support community organizations to deliver and evaluate programs directly. This requires development and support of authentic, sustained partnerships to build intervention and M&E capacity. In other words, a large organization should not simply identify a local partner for the purposes of meeting a grant requirement, but instead work with a local partner to address both specific programmatic goals as well as build knowledge and skills within the organization. Funders can support this by requiring not only involvement of a local partner, but also requiring demonstration of increased MHPSS and evaluation capacity within the partner organization. For example, building monitoring and evaluation tools that can be maintained by a local partner and adapted to future projects, providing training opportunities for local staff, and encouraging co-authorship on reports and publications. Beyond authentic partnership on individual projects, funding mechanisms that specifically target and build up regional research capacity and research networks is needed.

KEY TAKEAWAY
To improve program evaluation efforts, ensure intervention delivery aligns with research goals and strategies, empower community organizations to have an influence on the global research agenda, and increase community involvement in generating evidence for promising psychosocial programs, there is a need to support community organizations to deliver and evaluate programs directly.

6. Support integration of psychosocial interventions into crisis response and humanitarian operations

Only a small subset of the psychosocial supports included in this review were delivered as integrated within general humanitarian programming. This appears to be because even though basic services and supports are certainly thought to have an impact on wellbeing, the psychosocial aspects of these programs are often not clearly described, nor are psychosocial outcomes commonly measured. In addition, most humanitarian aid practitioners still lack training on MHPSS. Support for better integration of psychosocial considerations in all relevant cluster-specific humanitarian programming is needed. This should include everything from advocacy efforts to specific training and guidance documents that non-MHPSS specialists can understand.

Published by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
Registered offices Bonn and Eschborn, Germany

Programme Psychosocial Support for Syrian/Iraqi Refugees and IDPs

As at February 2021

On behalf of Federal Ministry for Economic Cooperation and Development (BMZ)

In cooperation with University of Virginia

GIZ is responsible for the content of this publication