Effectiveness and cost-effectiveness for the treatment of depressive symptoms in refugees and asylum seekers: A multi-centred randomized controlled trial

Kerem Böge,[1,a] Carine Karnouk,[1,a] Andreas Hoell,[b] Mira Tschorn,[c] Inge Kamp-Becker,[d] Frank Padberg,[e] Aline Übleis,[e] Alkomiet Hasan,[f] Peter Falkai,[e] Hans-Joachim Salize,[b] Andreas Meyer-Lindenberg,[b] Tobias Banaschewski,[b] Frank Schneider,[g,h] Ute Habel,[g] Paul Plener,[i,j] Eric Hahn,[a] Maren Wiechers,[e] Michael Strupf,[e] Andrea Jobst,[e] Sabina Millenet,[b] Edgar Hoehne,[d] Thorsten Sukale,[i] Raphael Dinauer,[j] Martin Schuster,[j] Nassim Mehran,[k] Franziska Kaiser,[f] Stefanie Bröcheler,[f] Klaus Lieb,[l] Andreas Heinz,[m,1] Michael Rapp,[c,1] and Malek Bajbouj[a]*

[a]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Campus Benjamin Franklin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[b]Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
[c]Department of Psychiatry and Psychotherapy, Psychosomatics and Psychotherapy, Faculty of Human Medicine, Philipps-University Marburg, Marburg, Germany
[d]Department of Psychiatry and Psychotherapy, University Hospital, Ludwig Maximilian University Munich, München, Germany
[e]Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, University of Augsburg, BKH Augsburg, Dr.-Mack-Str. 1, 86156, Augsburg, Germany.
[f]Department of Psychiatry and Psychotherapy Rheinisch-Westfälische Technische Hochschule Aachen University and JARA-Institute Brain Structure Function Relationship (INM 10), Research Center Jülich and RWTH, Aachen, Germany
[g]University Hospital Düsseldorf, Düsseldorf, Germany
[h]Department of Child and Adolescent Psychiatry, Medical University of Vienna, Vienna, Austria
[i]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Ulm, Ulm, Germany
[j]Department of Social and Preventive Medicine, University of Potsdam, Germany
[k]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany
[l]Department of Psychiatry and Psychotherapy, University of Mainz, Mainz, Germany
[m]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Campus Mitte, Berlin, Germany

Summary

Background Current evidence points towards a high prevalence of psychological distress in refugee populations, contrasting with a scarcity of resources and amplified by linguistic, institutional, financial, and cultural barriers. The objective of the study is to investigate the overall effectiveness and cost-effectiveness of a Stepped Care and Collaborative Model (SCCM) at reducing depressive symptoms in refugees, compared with the overall routine care practices within Germany's mental healthcare system (treatment-as-usual, TAU).

Methods A multicentre, clinician-blinded, randomised, controlled trial was conducted across seven university sites in Germany. Asylum seekers and refugees with relevant depressive symptoms with a Patient Health Questionnaires score of ≥ 5 and a Refugee Health Screener score of ≥ 12. Participants were randomly allocated to one of two treatment arms (SCCM or TAU) for an intervention period of three months between April 2018 and March 2020. In the SCCM, participants were allocated to interventions tailored to their symptom severity, including watchful waiting, peer-to-peer- or smartphone intervention, psychological group therapies or mental health expert treatment. The primary endpoint was defined as the change in depressive symptoms (Patient Health Questionnaire-9, PHQ-9) after 12 weeks. The secondary outcome was the change in Montgomery Åsberg Depression Rating Scale (MADRS) from baseline to post-intervention.

Findings The intention-to-treat sample included 584 participants who were randomized to the SCCM (n= 294) or TAU (n=290). Using a mixed-effects general linear model with time, and the interaction of time by randomisation group as fixed effects and study site as random effect, we found significant effects for time ($p < .001$) and time by group interaction ($p < .05$) for intention-to-treat and per-protocol analysis. Estimated marginal means of the PHQ-9

The Lancet Regional Health - Europe 2022;19: 100413
Published online xxx
https://doi.org/10.1016/j.lanepe.2022.100413

*Corresponding author at: Humboldt-Universität zu Berlin, and Berlin Institute of Health, Hindenburgdamm 30, 12203 Berlin, Germany.
E-mail address: malek.bajbouj@charite.de (M. Bajbouj).

[1] shared authorship

scores after 12 weeks were significantly lower in SCCM than in TAU (for intention-to-treat: PHQ-9 mean difference at T_1 1.30, 95% CI 1.12 to 1.48, $p < .001$; Cohen's d=.23; baseline-adjusted PHQ-9 mean difference at T_1 0.57, 95% CI 0.40 to 0.74, $p < .001$). Cost-effectiveness and net monetary benefit analyses provided evidence of cost-effectiveness for the primary outcome and quality-adjusted life years. Robustness of results were confirmed by sensitivity analyses.

Interpretation The SSCM resulted in a more effective and cost-effective reduction of depressive symptoms compared with TAU. Findings suggest a suitable model to provide mental health services in circumstances where resources are limited, particularly in the context of forced migration and pandemics.

Funding This project is funded by the Innovationsfond and German Ministry of Health [grant number 01VSF16061]. The present trial is registered under Clinical-Trials.gov under the registration number: NCT03109028. https://clinicaltrials.gov/ct2/show/NCT03109028

Keywords: Stepped-care and collaborative model; Refugees; Asylum seekers; Depression; Germany; Mental health care; SCCM; Interventions; Cost-effectiveness

Research in context

Evidence before this study

We searched PubMed and Google Scholar for recently published studies, nationwide surveys and official reports from the World Health Organization (WHO) regarding recent migration movements and the prevalence of mental illness among refugee populations in Europe and the Middle East. Current evidence points towards a high disease burden in Europe's refugee population, contrasting with a scarcity of available resources and culturally adapted treatment options in existing mental health care settings. Several linguistic, systemic, financial and cultural barriers thematically appeared in our research. In a parallel literature review, numerous RCTs had confirmed the efficacy of stepped care and collaborative models (SCCM) for the successful and cost-efficient treatment of psychiatric disorders.

Added value of this study

MEHIRA was tailor-designed to provide a wide range of culturally adapted mental health care services. Matching symptoms to a suitable and innovative psychological intervention showed to be both successful and cost-effective at reducing depressive symptoms in our refugee sample. Our study outcomes reveal evidence for overall high clinical- and cost-effectiveness of this model. To our knowledge, no other studies to date have investigated the efficacy of a culturally adapted SCCM. Findings from our RCT confirm the success of this model in its adaptability to different contexts both in crisis situations, but also in low- and middle-income countries, where the demand for medical care is high, but resources are scarce. The model also provides solutions to various barriers and resource scarcity simultaneously.

Implications of all the available evidence

Our findings suggest an adaptable model that can strengthen mental health care services for groups seeking psychological care in restricted contexts. SCCM's are sustainable, cost-effective and resource saving hybrid models, representing a combination of traditional and digital mental health care services. If adopted, they can provide benefits to both mental health care systems and individuals who are seeking care by filling an existing gap and overcoming different challenges at once.

Introduction

As a consequence of pandemics, natural disasters or armed conflicts such as in the Ukraine or Syria, mental health care systems face tremendous challenges.[1] Financial, structural, and cultural barriers, coupled with an increased disease burden, have led to large treatment gaps in many parts of the world.[2] Across the globe, healthcare systems are often characterised by a scarcity of resources and a lack of specialised expertise,[3] resulting in health care disparities and marginalisation of minority groups.[4] Consequently, intervention models intelligently allocating resources and catering to a large number of people at low cost, but with high effectiveness, are needed.[2,5,6]

A case example highlighting the need for alternative solutions can be currently observed in the context of the large migration movements, for example in the ongoing war in the Ukraine, or recent migration movements of populations of predominantly Farsi or Arab backgrounds (constituting 65% of Germany's refugee population[7]) from the Middle East to Europe in 2015. In that

time, Europea was faced with an unexpected challenge to accommodate over two million refugees who needed protection, shelter and medical assistance.[8] Around 400,000 of the incoming refugees between the years of 2010 and 2019 (with a peak in 2015) were unaccompanied and separated minors, many of whom applied for asylum in Germany.[9] Mental healthcare systems in Europe and Germany[5] were and are still not prepared to meet the healthcare needs of incoming populations[5] that has experienced severe psychological distress[10,11] and in which a significant proportion showed depression, PTSD and anxiety, preventing them from social and economic integration.[12] These patients posed a great challenge to Germany's health care system, due to systemic, organizational, cultural and linguistic barriers, despite its robustness.[5] Other host countries have highlighted similar challenges.[13] For this population, a lack of preparedness became apparent especially concerning culturally-sensitive interventions. This is noteworthy especially because even before the civil war in Syria large minorities already lived in Germany for decades and their specific needs were not addressed sufficiently. As a result, two interrelated challenges have surfaced: first that mental health systems were not tailored to the needs of minorities and second, that institutions lack the flexibility to quickly adapt to changing environments.[14]

Current evidence points towards a frequent and urgent plea for sustainable, accessible, culturally-sensitive and innovative mental health care models that are scalable, resource-saving and cost-effective.[1,2,6] Stepped Care and Collaborative models (SCCM) represent hybrid systems that offer collaborative and individualized treatments ranging from low- to high threshold interventions.[15] SCCM have shown to be effective in the prevention and intervention of anxiety and depression-related symptoms and have been used to systematically integrate and engage different mental health care workers at different stages of treatment,[16,17] insinuating its collaborative nature and capacity for differentiated support.[16] There are several guidelines and delivery versions of stepped care, with operational variations and defining features distinguishing their implementation.[16,18,19] To the best of our knowledge, stepped care models were not used in the mental health care of refugees and asylum seekers and little is known about their effectiveness with this specific population. However, since 2019, several studies have emerged that plan to implement this needs-based model in order to complement existing mental health care structures, such as Refukey[20] and BETTER CARE.[21] Within mental health, there are two common ways of implementing stepped care models: progressive and stratified approaches.[18,19,22] The progressive model focuses on initially assigning all patients to the lowest intensity intervention and then 'self-correcting' by stepping patients up.[16,23] Whereas, the stratified model places patients in either lower or higher intensity interventions depending on assessment, including complexity of disorder and symptom severity.[24] It is also not uncommon to combine both models. In an evaluation of stratified and progressive care in four different mental health care sites, it appeared that each site had a slightly different interpretation to stepped care.[19] Our model strives to develop a culturally-adapted SCCM, filling an existing gap by addressing major barriers in the availability and delivery of tailored psychiatric treatments for refugees and asylum seekers.[25,26]

Therefore, the Mental Health in Refugees and Asylum Seekers (MEHIRA) study, a multi-centre, randomized, controlled trial intended to evaluate the effectiveness of a SCCM, in which interventions were allocated according to disease severity at four levels. Within this study, interventions were developed specifically for refugees with depressive symptoms of different severities. Based on available governmental data about the age range of refugees with a peak age between 14 and 25 years, we developed a study design with interventions specifically tailored to the needs of the population within this age range. i.e. in the transition between adolescence and adulthood. We hypothesize that participants in the active condition (SCCM) will show greater improvements in depressive scores compared to the control condition (TAU) from baseline to post-intervention. As a second hypothesis, we assume more cost-effectiveness of the SCCM compared with the cost-utility of the routine care practices (TAU) in Germany.

Methods

Design

A multicenter, clinician-blinded, randomised controlled trial was conducted between May 05/2018 and 03/2020, including seven university hospitals across Germany. Approval was obtained from the institutional ethics board at each site. All participants provided written informed consent. For participants <18 years of age, either the parents or legal guardians gave consent. Further details with respect to statistical analyses, data collection, and study procedures were conducted according to the original study protocol[27] and are described in more detail in the supplementary material.

Participants and procedure

Trial population included male and female asylum seekers and refugees as defined by the United Nations High Commissioner for Refugees,[8] aged between 14-65 years, Arabic/Farsi native-speakers and/or fluent in English/German, with at least mild depressive symptoms measured by the self-rated Patient Health Questionnaires (PHQ-9[28]; PHQ-A for adolescents[29]) and relevant psychological distress assessed by the Refugee

Health Screener (RHS-15).[30] In the screening phase, participants first needed to display score of ≥ 12 for items 1 to 14 or ≥ 5 for item 15 at the RHS-15, and moreover at least a PHQ-9 score of at least 1 or higher on ≥ 5 items. The cut-off values for the RHS-15 are based on previous studies with refugees from Iraq, Nepal, Bhutan and Burma with ≥ 14 years, while the combination of these values showed high sensitivity (0.87) and specificity (0.79) to identify PTSD in refugees.[28] Moreover, we used the internationally well-established screening instrument PHQ-9 and its cut-offs scores. The PHQ-9 has been translated to more than 70 languages and provides evidence on measurement invariance in multi-ethnic populations[31] and thus, items were similarly used or function similarly in people from different ethnic backgrounds, including South-Asians, Africans, Turks, and Dutch.

Exclusion criteria were absent informed consent, diagnosis of a psychotic disorder assessed by the Mini-International Neuropsychiatric Interview (MINI),[32] a degenerative disorder evaluated by psychiatrist, and current risk of suicidality measured by the clinician-based Montgomery-Åsberg-Depression-Rating Scale (MADRS),[33] with ≥ 4 on item 10 assessed by a psychiatrist and/or psychologist. In cases of severe suicidality or other adverse events as defined by the study protocol, participants were excluded from the present study and admission to the inpatient units of the respective study center was arranged. In contrast to the initial trial protocol, the age range was decreased to 14 years before starting the study to support recruitment.

Participants were recruited through regionally heterogeneous allocation paths, including general practitioners, social workers, central clearing and outpatient clinics, refugee accommodations, language courses and religious institutions. Study teams presented the study, as well as the inclusion criteria, in a variety of lectures, workshops, as well as general information sessions and distributed flyers about the study in the mentioned locations to promote recruitment. However, irrespective of the heterogenous allocation paths, all screenings were conducted by study personal of trained psychologists to ensure consistency either at the respective place or at each study center. Participants indicating relevant symptom burdens, as measured by the PHQ-9 and RHS-15 according to inclusion criteria, were further invited for a complete study inclusion and baseline assessment at each corresponding study center. Screening phase (T_{-1}) was performed -4 weeks to 1-day prior to baseline assessment. Afterwards baseline assessment was performed at each study site respectively and participants were directly randomized displaying T_0.

Randomization was assigned at the individual level in a 1:1 scheme with fixed block size of four ("Extended Stratified Block Without List - secuTrial"), stratified by center, according to a computer-generated electronic Case Report File (eCRF) by an independent and external coordinating centre for clinical trials (Koordinierungszentrum für Klinische Studien - KKS), to receive either a SCCM intervention or treatment-as-usual. Study personnel had to enter the collected data to an online eCRF mask and received the allocation after inclusion inhibiting any prediction about the randomization process.

Treatment phase was initiated for a period of 12-weeks, directly resulting in the post-intervention assessment (T_1). Participants were invited to the study sites for two further assessments after 24 weeks (T_2) and 48 weeks (T_3) for follow-up.

Interventions

Within the SCCM condition, interventions were provided at four levels; participants were allocated to one of the four levels according to the PHQ-9 score at baseline (Figure 1). Composition of interventions differed for the adult and adolescent subgroups. Interventions for the adult subpopulation (age range 18 to 65) were provided by study sites in Munich, Aachen and two sites in Berlin. Interventions for the adolescent group (14 to 21 years) were provided by study sites in Marburg, Ulm, Tübingen, and Mannheim.

Interventions had a duration of 12 weeks for all levels, albeit with increasing treatment intensity and frequency by each level. After 12 weeks, symptoms were reassessed with the PHQ-9 and a second intervention phase was initiated according to the respective PHQ-9 at post-intervention in line with the procedure at baseline. Thus, participants could either be stepped up or down after the first treatment period depending on the initial inclusion criteria for PHQ-9 and RHS-15 (see participant and procedure). If a second treatment was not conducted, participants continue with the assessments leading to the follow-up assessments. All SCCM treatments were given alone and with no addition of any further practices.

Overall, in the first year of the project, the empowerment-, peer-to-peer- and smartphone-based interventions for levels 2 and 3 were developed by the research consortium through a participatory approach with members of the Middle Eastern refugee community (see study protocol for the in-depth elaboration of the development and methodological procedures[27]). A description of the SCCM and a summary of the content can be found in Table S6 in the supplementary material.

Participants assigned to level 1 received no intervention for the treatment period ("watchful waiting"). In level 2, a smartphone-based intervention was provided for adolescent and adult participants. The smartphone application (Balsam) contains 80+ videos and 15 modules covering topics, such as disease models, sleep hygiene, stigma, symptom manifestations, cultural belonging, acculturation, and cognitive-behavioural

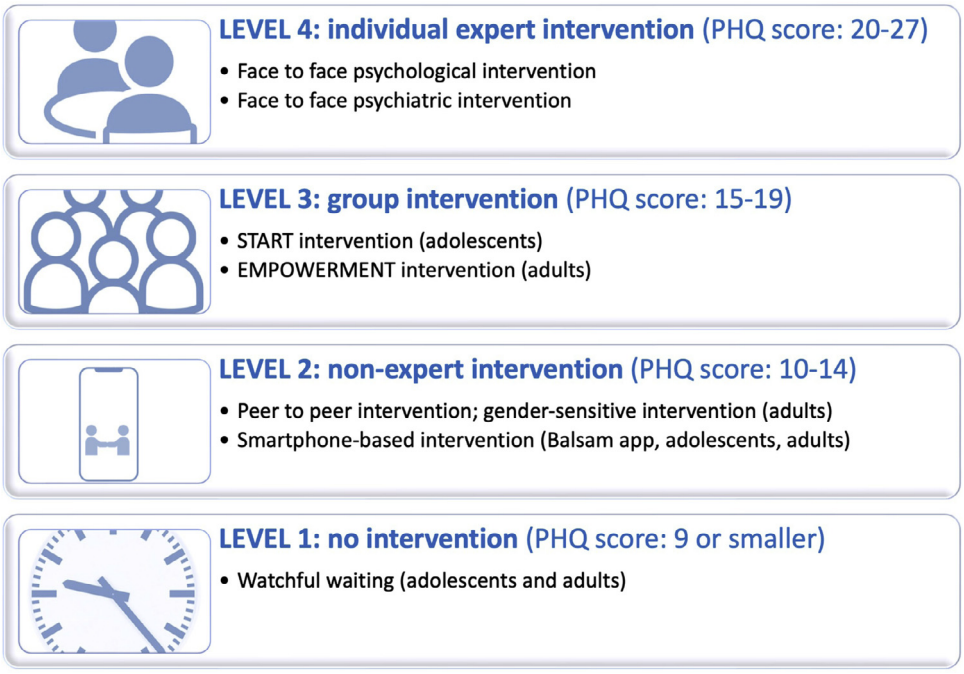

Figure 1. Intervention pyramid of the Stepped Care and Collaborative Model (SCCM) for adults (upper part) and adolescents (lower part of figure)

interventions. Study personal supported the participants in downloading the application at baseline and gave an overview of the content and functions. As another option on level 2, participants received an in-person peer-to-peer group intervention. Members from the local refugee community were trained who had a background in teaching or medicine professions including teachers, nurses, educators and spoke either Arabic or/and Farsi. These peers offered weekly psychosocial support and received weekly supervision by licensed psychotherapists at each study side. The peer-to-peer manual and intervention is group based with 6-10 participants and two peer trainers aimed to strengthen resources, address emotional needs, and improve coping skills by focusing on sharing experiences.[15]

At level 3, two separate in-person group-based interventions were offered by incensed psychotherapists either in fluent in Arabic or Farsi or in assistance of translators. Psychotherapists received regular supervision by the respective principal investigator at each study center. Adolescents participants received an adapted START intervention, a standardised and international well-validated brief therapeutic intervention, including elements of mindfulness, skill training and covering topics centred around stress regulation and resilience.[34] Adult participants received the Empowerment intervention,[35] which was developed to equip participants with the information and resources to cope with stressors and depressive symptoms. The manual includes 16 culturally sensitive modules comprising cognitive and behavioural strategies (behavioural activation, stress management, emotion regulation) as well as specific topics such as homesickness and psychosomatic symptoms. Level 4 consisted of an in-person expert intervention in which participants either received pharmacological agents and/or psychotherapy given either by a licensed psychiatrist or psychotherapist on a weekly basis. Treatment providers received weekly supervision from the respective principal investigator within the mental care structures of the respective study center. When clinicians were not able to speak the participant's preferred language, a professional translator was provided.

In the control condition, participants were allowed to receive all social, psychotherapeutic, psychiatric, and further healthcare services within their region. Accordingly, participants were allowed for receive any kind of treatment without time, frequency, and intensity restriction within the German health care system which displays treatment-as-usual (TAU) in Germany. There were no binding regulations and stipulations regarding the respective treatments, -facilities, or -persons.

Public and patient involvement

The refugee population was involved in the MEHIRA study on four interrelated levels: they contributed (i)

	Mean ± SD; N/Total N (%)		
	SCCM (n=294)	TAU (n=290)	*P* Value
Age (years)	28.63 ± 10.79	28.63 ± 10.36	.99
Female	93/294 (31.63)	93/290 (32.31)	.91
Years of education	8.63 ± 4.03	8.83 ± 4.36	.59
Marital status			.42
Single	156/284 (54.93)	148/270 (54.81)	
Married	104/284 (36.62)	89/270 (32.96)	
Divorced	18/284 (6.34)	27/270 (10.00)	
Widowed	6/284 (2.11)	6/270 (2.22)	
Having children	113/283 (39.93)	106/268 (39.55)	.93
Past SES			.36
Upper class	19/271 (7.01)	26/262 (9.92)	
Upper middle class	44/271 (16.23)	53/262 (20.23)	
Middle class	152/271 (56.09)	129/262 (49.24)	
Lower middle class	33/271 (12.18)	36/262 (13.74)	
Lower class	23/271 (8.49)	18/262 (6.87)	
Current SES			.08
Upper class	1/271 (.37)	4/261 (1.53)	
Upper middle class	16/271 (5.90)	8/261 (3.07)	
Middle class	109/271 (40.22)	94/261 (36.02)	
Lower middle class	52/271 (19.19)	70/261 (26.82)	
Lower class	93/271 (34.32)	85/261 (32.57)	
Current employment			.64
Unemployed	216/269 (80.30)	219/263 (83.33)	
Protected employment	4/269 (1.49)	3/263 (1.14)	
Employee	47/269 (17.47)	41/263 (15.59)	
Military service/community	1/269 (.37)	0/263 (.00)	
Self-employed	1/269 (.37)	0/263 (.00)	
Reason for migration[a]			
War	167/291 (57.39)	158/277 (57.04)	.93
Natural disaster	2/291 (.69)	3/277 (1.08)	.61
Economic crisis	27/291 (9.28)	19/277 (6.86)	.29
Individual situation	49/291 (16.84)	42/277 (15.16)	.59
Political situation	102/291 (35.05)	109/277 (39.35)	.29
Social situation	61/291 (20.96)	59/277 (21.30)	.92
Other	33/291 (11.34)	24/277 (8.66)	.29
Time since arrival in Germany (in years)	3.04 ± 2.29	2.71 ± 4.25	.31
Primary Outcome PHQ-9, estimated marginal means from GLMM (mean ± standard error)			
PHQ-9 T_0 (Week 0)	15.87 ± 0.84	16.62 ± 0.81	.10
PHQ-9 T_1 (Week 12)	13.09 ± 0.88	14.41 ± 0.88	.04
PHQ-9 T_2 (Week 24)	12.80 ± 0.96	13.51 ± 0.94	.42
PHQ-9 T_3 (Week 48)	12.10 ± 1.30	13.05 ± 1.32	.54

***Table 1*: Baseline Characteristics of the ITT sample.**
Abbreviations: SES socioeconomic status, PHQ-9 Patient Health Questionnaire, GLMM generalised linear mixed model.
[a] Multiple answers possible.

within the development procedure of the MEHIRA interventions as official and unofficial advisors, as well as focus group members; (ii) as graphic designers for the BALSAM app; (iii) as facilitators in the peer-to-peer-intervention; and (iv) within the study in the role of interns, translators, psychologists and research assistants.

Outcomes

All assessments were conducted with validated and reliable tools for the specific population (PHQ-9, RHS-15, MADRS, MINI, and WHO-BREF) at baseline (T_0), at 12, (T_1), at 24 (T_2) and 48 weeks (T_3), using official translations in Arabic or Farsi, delivered by our study personnel. Moreover, demographic information was

	Mean ± SD; N/Total N (%)		
	SCCM (n=144)	TAU (n=191)	*P* Value
Age (years)	29.84 ± 10.93	28.92 ± 10.33	.43
Female	53/144 (36.81)	63/191 (32.98)	.47
Years of education	8.48 ± 4.12	8.23 ± 4.39	.60
Marital status			.46
Single	71/143 (49.65)	102/189 (53.97)	
Married	58/143 (40.56)	63/189 (33.33)	
Divorced	11/143 (7.69)	21/189 (11.11)	
Widowed	3/143 (2.10)	3/189 (1.59)	
Having children	62/143 (43.36)	75/190 (39.47)	.48
Past SES			.32
Upper class	5/137 (3.65)	12/185 (6.49)	
Upper middle class	22/137 (16.06)	36/185 (19.46)	
Middle class	80/137 (58.39)	96/185 (51.89)	
Lower middle class	15/137 (10.95)	28/185 (15.13)	
Lower class	15/137 (10.95)	13/185 (7.03)	
Current SES			.04
Upper class	1/138 (.73)	1/184 (.54)	
Upper middle class	11/138 (7.97)	4/184 (2.17)	
Middle class	57/138 (41.30)	69/184 (37.50)	
Lower middle class	26/138 (18.84)	55/184 (29.89)	
Lower class	43/138 (31.16)	55/184 (29.89)	
Current employment			.51
Unemployed	108/137 (78.83)	153/185 (82.70)	
Protected employment	3/137 (2.19)	2/185 (1.08)	
Employee	25/137 (18.25)	30/185 (16.22)	
Self-employed	1/137 (.73)	0/185 (.00)	
Reason for migration[a]			
War	86/144 (59.72)	118/191 (61.78)	.70
Natural disaster	1/144 (.69)	1/191 (0.52)	.84
Economic crisis	13/144 (9.03)	10/191 (5.24)	.17
Individual situation	19/144 (13.19)	27/191 (14.14)	.80
Political situation	50/144 (34.72)	73/191 (38.22)	.51
Social situation	31/144 (21.53)	37/191 (19.37)	.63
Other	14/144 (9.72)	19/191 (9.95)	.95
Primary Outcome PHQ-9, estimated marginal means from GLMM (mean ± standard error)			
PHQ-9 T_0 (Week 0)	14.31 ± 1.28	15.19 ± 1.27	.15
PHQ-9 T_1 (Week 12)	11.78 ± 1.31	13.52 ± 1.29	.01
PHQ-9 T_2 (Week 24)	11.48 ± 1.25	12.99 ± 1.14	.15
PHQ-9 T_3 (Week 48)	10.05 ± 1.69	12.11 ± 1.48	.25

***Table 2*: Baseline characteristics of the PP sample.**
Abbreviations: SES socioeconomic status, PHQ-9 Patient Health Questionnaire, GLMM generalised linear mixed model.
[a] Multiple answers possible.

assessed at baseline which are displayed in Table 1 and Table 2. Here, participants were free to rate for example their socioeconomic status, without being asked about their income. In some cases, assessment procedures were conducted with the support of interpreters. Clinician-based assessments such as the MADRS or MINI were performed by a psychiatrist or psychologist blinded for the type of condition (TAU and SCCM) as well as specific SCCM intervention. Participants were briefed about the purpose of blind-assessments. In the few cases of unblinding, we changed the assessor to ensure blinding throughout the trial.

The primary endpoint was the reduction of the self-rated PHQ-9 score from baseline (T_0) to post-intervention (T_1). Remission was defined as a reduction of PHQ-9 scores to ≤ 8 points, and response was defined as a reduction of PHQ-9 scores by ≥ 50%. A secondary endpoint was the reduction of the blinded clinician-based

MADRS score from baseline (T_0) to post-intervention (T_1) to examine depressive symptoms changes objectified by a clinician. Only results of the primary outcomes depressive symptoms (self-rated PHQ-9 and blind clinician-rated MADRS) as well as the cost-effectiveness outcomes are present in this article. All further assessed variables as depicted in the published study protocol[27] and preregistered at clinical trials (NCT03109028) and their outcomes will be published in future articles.

Cost-effectiveness analyses

We performed a cost-effectiveness analysis alongside the trial. We estimated direct costs associated with the consumption of health care resources under routine conditions and costs of the intervention. We adopted the perspective of the health service providers for both kinds of costs.

We measured routine health service utilization in the sample of refugees and asylum seekers with an adapted version of the validated Mannheim Module Resource Use (MRU).[36] The MRU measured the frequency of resources consumed in the following areas: outpatient medical services, emergencies, mental health specialists, remedies and further outpatient therapies, and counselling or health support services. It also included hospitalization days and medication use. The recall period covered the last three months. We combined data on resource use with specific unit costs that were taken from nationally or regionally available data sources. All prices were calculated for the reference year 2019 in Euros and if necessary were indexed using the German consumer price index.[37] We extended resource uses to one year using follow-up measurements on resource use or last observation carried forward (LOCF) (further specification can be found in the appendix and Tables S1 and S2).

We calculated intervention costs of trial for each treatment step separately, because participants of the intervention had the chance to enter just one specific intervention type based on the individual baseline PHQ-value. We used a micro-costing approach to calculate the costs of each step by interviewing key persons of each intervention. In the structured interview, we collected data on running expenses due to consumables, personnel deployment and operating expenses (premises, equipment). To calculate costs of the intervention we used personnel wages based on gross hourly wages according to the tariff agreement of the federal states in Germany in 2019. We calculated per group and per patient costs using ITT-sample sizes of participants assigned to each step. In addition, we calculated costs per step or type of intervention based on an optimal scenario where all groups and applications had an expected capacity utilization of 100% in consideration of the rate of refugees with depressive symptoms entering Germany in 2019. Hence, we calculated the average costs of [TERM] MEHIRA trial per patient by counting the proportions of participants in each intervention type either divided by the total number of the ITT-sample size (Base Case) or by the total number of theoretically assigned participants (Optimal Case). An in-depth description of this procedure and assigning of average costs of each step and intervention grouped by adolescents and adults is provided in the appendix and table S2. Finally, we combined resource use costs with implementation costs. For reasons of uncertainty and because we did not discount costs due to the short time horizon of the study, we created three alternative scenarios for the comparison of total costs in trial with resource use costs in TAU. We defined a base case (BC) and an optimal case (OC) as mentioned above. In addition, we defined an on-top case (IC), where intervention expenses were compared to zero costs of the TAU condition. The underlying assumption was that resource use patterns were equal between both groups after one year.

In the TAU condition, participants were allowed to receive all social, psychotherapeutic, psychiatric, and further healthcare services within their region. We did not assign any intervention costs to that group, but measured the consumption of (routinely) used health services at any time point.

The primary clinical outcome of the cost-effectiveness analysis is PHQ-9 value at follow up. We adjusted PHQ-9 values for age, gender, study site, and baseline value.

For the cost-utility analysis, the outcome was quality-adjusted life years (QALY). We calculated QALY for each measurement point using results from the self-rated WHO Quality of Life questionnaire, brief version (WHOQOL-BREF),[38] a 26-item questionnaire, where each item is rated on a five-point Likert scale. To obtain utility values we applied the conversion algorithm proposed by Salize & Kilian.[36] We extended calculated QALY from post-intervention to a 12 months period using additional QALY values from further follow-ups or last observation carried forward (LOCF) method. We corrected reported QALYs for age, gender, study site and initial value. Because the time horizon of the study was one year, we did not discount outcomes.

Sample size calculation and statistical analyses of treatment effect

The a priori power analysis for primary outcome of depressive symptoms, taking site variation into account, occurrence yielded initially a planned sample size of 476 participants (238 per arm) for the analysis from T0 to T3, with an anticipated dropout rate of 50% leading to an overall needed sample size of 952. We specified as the intention-to-treat (ITT) sample all randomised participants who provided baseline data on the primary outcome. Furthermore, we pre-specified as per-protocol (PP) sample all participants who took part in at least

50% of the provided intervention sessions (details: supplementary material). To enable a better interpretation of effect sizes for the primary analysis from T0 to T1, we calculated a post-hoc power analysis based on observed effect size (f = .115), α = .05, and β = .80. Considering a correlation between repeated measures of .5, assuming a conventional 2×2 repeated measures ANOVA yielded an overall sample size of 152 for the within between interaction effect. This effect size is comparable to priori interventions in refugee populations (2-4).

For primary and the first set of secondary analyses, repeated measure generalised linear mixed models (GLMM) were used to examine fixed effects of the time by randomisation groups (SCCM vs. TAU) from T_0 to T_1 (week 0 to week 12). For this model, we included time and time × randomisation (treatment effect) as fixed effects and random effect to adjust for study centre (cluster). We ran this model for the primary outcome variable PHQ-9 score and the secondary outcome MADRS score, for both the ITT and PP sample, using data from T_0 and T_1 (12 weeks). For a comprehensive quantification of the analysed treatment effect, we report the difference (SCCM vs. TAU) of the PHQ-9 estimated means at T_1 and the baseline-adjusted estimated mean difference at T_1. Additionally, we calculated Cohen's d by dividing the estimated marginal mean differences (SCCM vs. TAU at T_1) by the pooled T_0 SD. For a clinically relevant measure, we reported and compared remission rates (T_1 PHQ-9 score $\leq$ 8) as well as response rates defined as a reduction of PHQ-9 scores by $\geq$ 50% and imputed missing values at T_1 via LOCF for remission and response analyses. For follow-up analyses, we performed GLMM with both PHQ-9 and MADRS scores from the ITT as dependent variables, randomisation group (TAU vs. SCCM), and the interaction of randomisation group and time point as fixed effects, and centre as random effect for all available data from T_0 to T_2 (24 weeks), and T_0 to T_3 (48 weeks), respectively, and tested for differences in overall model fit between the models. Tests were 2-tailed and statistical significance was set at a P-value of less than 0.05.

Cost-effectiveness - statistical analyses

Statistical analyses were performed strictly according to the ITT principle. Missing data were imputed with the LOCF method, a conservative approach strengthening the null-hypotheses of equal costs and effects between SCCM and TAU. Because of highly right-skewed cost data, we applied generalized linear models (GLM) with gamma distribution and identity link function to estimate differences in health care costs between groups for all three scenarios. We performed a crude model containing randomization group as explanatory variable and an adjusted model containing randomization group, age, gender, study site, and baseline costs as explanatory variables.

We determined the incremental cost-effectiveness ratios (ICER), which we calculated as the ratio between the differences in mean costs, i.e. incremental costs (ΔC) and the differences in mean effects of depressive symptoms and QALY, i.e. incremental effects (ΔE). The ICER represents the additional costs to obtain one additional QALY or to decrease the PHQ-score by one point. To satisfy the condition of statistical uncertainty around the ICER, we performed non-parametric bootstrapping with 10,000 replications, which we plotted on cost-effectiveness planes. We calculated bootstrapped 95% confidence intervals (95%CI) around the ICER. Bootstrapped confidence limits are elusive when the denominator of the ICER approaches zero and in cases bootstrapped ICER spread all over the CE plane.[4] Therefore, and since certain thresholds (λ) are usually unknown, we checked the likelihood of cost-effectiveness with an additional incremental net-monetary benefit (NMB) approach. The NMB approach is a function of λ and we considered different willingness to pay thresholds represented on the horizontal axis with the probability of cost-effectiveness on the vertical axis. The cost-effectiveness acceptability curve (CEAC) shows the probability that SCCM is cost-effective in comparison to TAU for a range of willingness to pay values. To satisfy the condition of parameter uncertainty, we performed all analyses with the three scenarios mentioned above (BC, OC, and IC). All analyses were performed using SPSS (SPSS Inc., Chicago, Illinois, USA) version 26, SAS statistical software (SAS Institute Inc., Cary, North Carolina, USA) version 9.4 and Excel 2016 for Windows.

Results

Sample and recruitment

Between April 2018 and December 2019, 584 participants were randomised in a 1:1 ratio, with 294 assigned to SCCM, and 290 to TAU. Recruitment varied across centres and the number of participants randomised to each centre is shown in Figure 2. ITT participants' characteristics in each study group at baseline are shown in Table 1. PP participants' characteristics in each study group at baseline are shown in supplementary material Tables S1 and S2. Based on the original sample size calculation we were able to recruit a little over 75% of the planned recruitment target.

Effectiveness

ITT analysis. *Primary outcome.* For the ITT sample, primary outcome data were available for 294 participants in SCCM and 290 participants in TAU at T_0, and 174 in SCCM and 186 in TAU at T_1, respectively. There was a significant effect for time (T_0 vs. T_1) ($F_{1,940}$=39.51, p < .001). A time by group (SCCM vs. TAU; $F_{2,940}$=3.35,

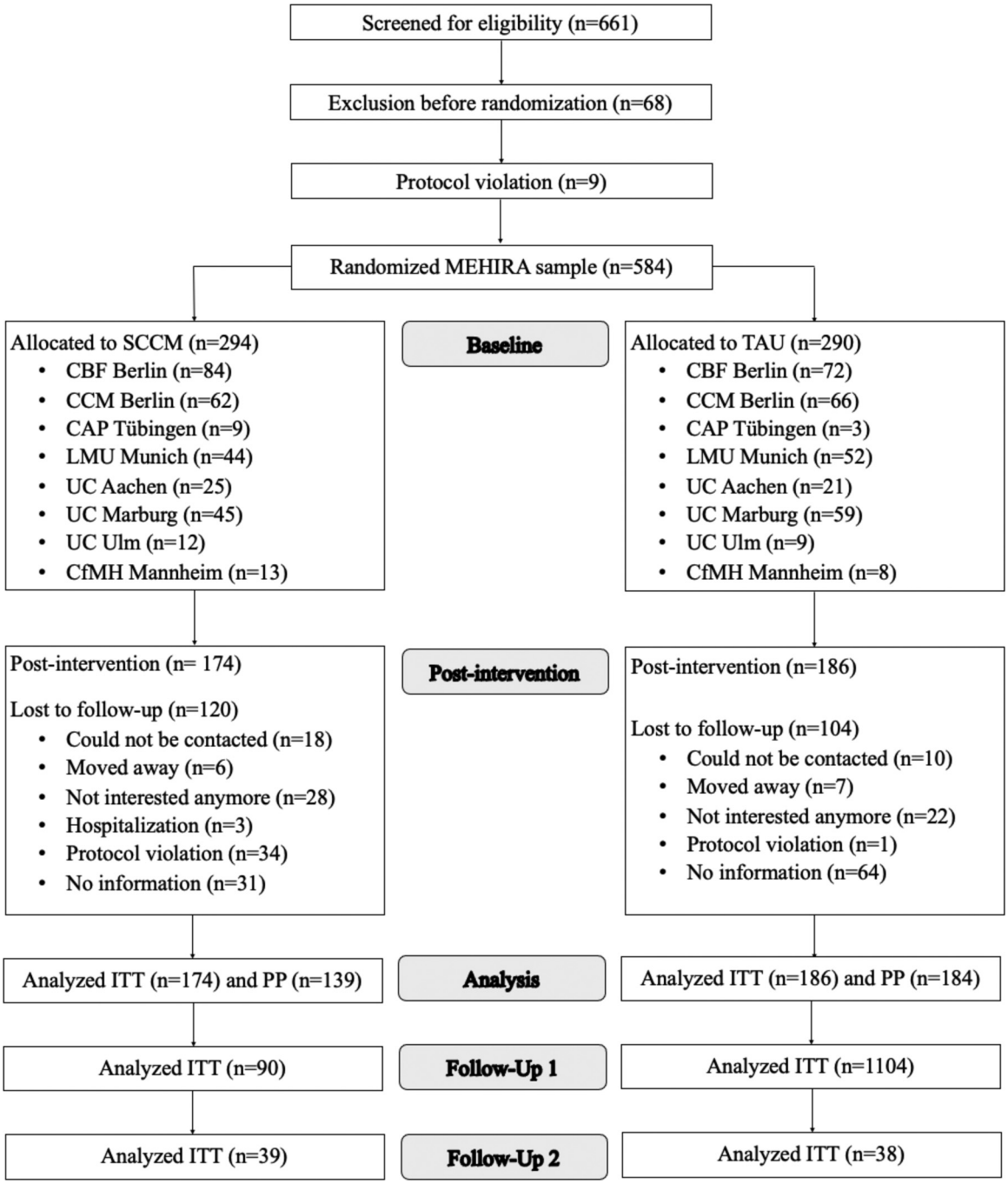

Figure 2. Flow chart of recruitment and randomization allocation

p = .035, d=.23 in favour of SCCM) interaction significantly predicted PHQ-9 scores and characterises the main treatment effect. The intraclass correlation coefficient (ICC) across centres was .11, therefore 11% of PHQ-9 variance were explained by the study centre. PHQ-9 scores as a function of intervention group (SCCM vs. TAU) and time (T_0 vs. T_1) are shown in Figure 3a. Difference of PHQ-9 estimated means at T_1 was 1.3 (95% CI=1.12 − 1.48, p < .005) and baseline-adjusted PHQ-9 estimated mean difference at T_1 was 0.57 (95% CI 0.40 to 0.74, p < .001). Therefore, in the GLMM the T_0-T_1 decrease of PHQ-9 scores in SCCM was 0.57 higher compared to TAU.

Rate of remission (PHQ-9 score ≤ 8 at T1) in SCCM was 19% (95%CI: 14.7%-24%) and 12% (8.6%-16.4%) in TAU (p = .020). Rate of response (PHQ reduction of

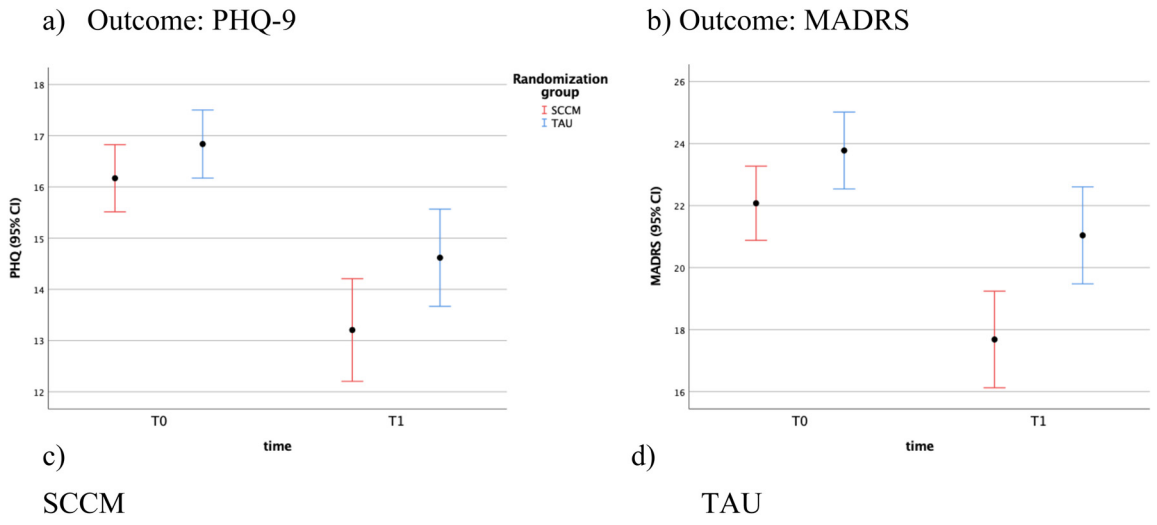

Figure 3. Scores on the PHQ-9 (primary outcome) and the MADRS scale (secondary outcome) as a function of randomization group (TAU vs. SCCM) and time (T0 vs. T1) for the ITT sample.

Note. Error bars represent 95% confidence intervals.

≥ 50%) was 12.9% (9.3%-17.3%) in SCCM and 9.6% (6.5%-13.7%) in TAU (p = .212).

Secondary outcome. MADRS data were available for the ITT sample for 273 participants in SCCM and 258 participants in TAU at T_0, and 177 in SCCM and 189 in TAU at T_1, respectively. For MADRS scores as the dependent variable, GLMM revealed a main effect for time (T_0 vs. T_1; $F_{1,893}$=32.78, p < .001), together with significant time by randomization group (SCCM vs. TAU; $F_{2,893}$=4.77, p = .009) interaction. The ICC across centres was .29. MADRS scores as a function of intervention group (SCCM vs. TAU) and time (T_0 vs. T_1) are shown in Figure 3b.

All ITT analyses stayed stable when repeated in the PP sample. As expected, we found a higher effect size (d=.30 in favour of SCCM) and significantly higher response rates in SCCM (23.6%, 16.9%-31.4%) compared to TAU (14.1%, 9.5%-19.9%, p = .026). See further details in Table S5.

Follow-up analysis. For the ITT sample, primary outcome data were available for 294 participants in SCCM and 290 participants in TAU group at T_0, 174 in SCCM and 186 in TAU at T_1, 90 in SCCM and 104 in TAU at T_2, and 39 in SCCM and 38 in TAU at T_3. Using GLMM from T_0 to T_2 with PHQ-9 scores as the dependent variable and a fixed effect for time (T_0 vs. T_1 vs. T_2) and a time by randomisation group (SCCM vs. TAU) interaction, we found a main effect for time (T_0 vs. T_1 vs. T_2; $F_{2,1132}$=30.88, p < .001), together with a marginal time by randomisation group (SCCM vs. TAU; $F_{3,1132}$=2.38, p = .068) interaction. Model fit, as indicated by the -2log likelihood criterion was 7228.27. Running the same analysis from T_0 to T_3, we found a main effect for time (T_0 vs. T_1 vs. T_2 vs. T_3; $F_{3,1207}$=23.8154, p < .001) but no time by randomisation group (SCCM vs. TAU; $F_{4,1207}$=1.86, p = .116) interaction. As indicated by -2log likelihood criterion, model fit was 7735.63, indicating a poorer model fit in the more extended follow-up model. PHQ-9 scores over all time points are shown in Supplementary Figure S1.

Cost-effectiveness

Health care costs. Resource use and cost data were available for 561 participants (SCCM=283 and TAU=278). One year after baseline, per capita resource use costs averaged €1,688.5 (SD=€571.3). Compared to TAU, per capita costs were significantly lower in SCCM (€-456.0, 95%CI=€-789.8 to €-122.2), due to significantly reduced inpatient and outpatient psychological or psychiatric treatment (Table S3).

Per capita BC intervention costs varied greatly between intervention types (Table S2). Overall, mean BC program costs for all 283 SCCM participants were €312.1 (SD=€57.5).

Total health care costs in SCCM were the sum of resource use costs and SCCM intervention costs. The fully adjusted GLM revealed no differences in total health care costs between SCCM and TAU after one year in BC (Table 3, Table S3). Additional results are provided in the supplementary material.

Clinical outcome. The intervention had a significant effect on depressive symptom scores displayed by adjusted mean differences in PHQ-values (Table S4). The ICER for BC was estimated at €-129.5 with

	Cost-effectiveness (PHQ)SCCM vs. TAU		Cost-effectiveness (QALY)SCCM vs. TAU	
	M (95%CI)	P	M (95%CI)	P
			Base Case	
Incremental Costs	-205.3 (-690.7 to 252.6)	0.44	-202.7 (-703.4 to 272.5)	0.44
Incremental Effect	1.59 (.85 to 2.32)	0.03	0.08 (-.01 to 0.18)	0.1
ICER (€/Effect)	-129.53 (-507.0 to 171.0)		-2,401.9 (-22.521.6 to 11,873.7)	
			Optimal Case	
Incremental Costs	-391.86 (-874.0 to 79.5)	0.07	-388.0 (-897.1 to 83.3)	0.07
Incremental Effect	1.59 (.83 to 2.32)	0.03	0.08 (-.01 to 0.18)	0.1
ICER (€/Effect)	-247.21 (-680.4 to 52.2)		-4,597.8 (-36,317.8 to 16,308.1)	
			On-Top Case	
Incremental Costs	315.1 (135.9 to 320.9)	<.001	315.1 (309.9 to 320.1)	<.001
Incremental Effect	1.59 (.84 to 2.32)	0.03	0.08 (-.01 to 0.18)	0.1
ICER (€/Effect)	199.31 (135.9 to 374.9)		3,733.7 (-16,284.5 to 28,005.7)	

***Table 3*: Adjusted Results of GLM Analyses for incremental costs and effects (PHQ and QALY) of SCCM scenarios compared to TAU and calculated ICER with bootstrapped confidence limits.**

Notes: n= 382 participants were included in cost-effectiveness for clinical effects (PHQ) and n= 529 participants were included in cost-effectiveness for patient reported outcome (QALY). GLM adjusted for age, gender, study site and resource use costs at baseline. GLM = Generalized Linear Model, PHQ = Patient Health Questionnaire, QALY = Quality Adjusted Life Year, ICER = Incremental Cost-Effectiveness Ratio, SCCM = Stepped and Collaborative Care Model, TAU = Treatment as Usual.

bootstrapped 95%CI of €-507.0 to €171.0. We found that 80% of the bootstrapped ICER were located in the southeast quadrant of the CE plane (Table S4 and Figure S2). This quadrant is associated with greater effects and lower costs of SCCM in relation to TAU. Thus, SCCM showed greater cost-effectiveness than the TAU condition. Figure 4 showed that the incremental CEAC for BC intersects the Y-axis at at 0.8. That means

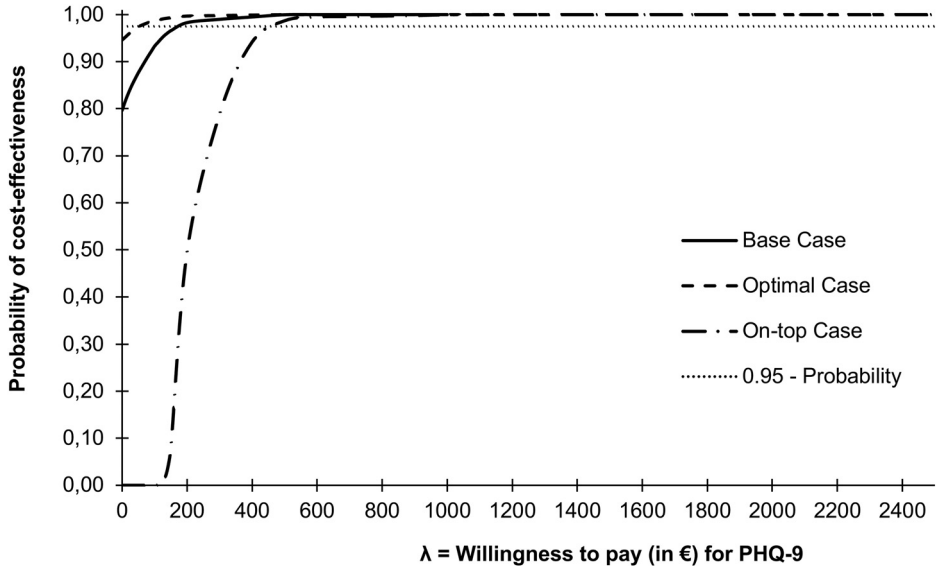

Figure 4. Net-Monetary Benefit - cost-effectiveness acceptability curves for all scenarios of SCCM vs TAU on PHQ.

Note. Probability that SCCM intervention is acceptable (values on the vertical axis) in relation to TAU on the willingness to pay for a reduction of PHQ values by one point, given varying thresholds for willingness to pay (horizontal axis) based on 10,000 bootstrapped ICER replications. The small dotted line (0.95 − probability) indicates the upper 95%CI, i.e. the maximum amount that has to be invested to be confident that SCCM is cost-effective. Intersections of CEAC with the confidence line represents cost-effectiveness for a specific scenario. Thus, these λ were €171 for Base Case, €52 for Optimal Case, and €375 for On-top Case (representing cost of intervention only), respectively.

SCCM=Stepped and Collaborative Care, TAU=Treatment as Usual, PHQ=Patient Health Questionnaire, CEAC=Cost-Effectiveness Acceptability Curve.

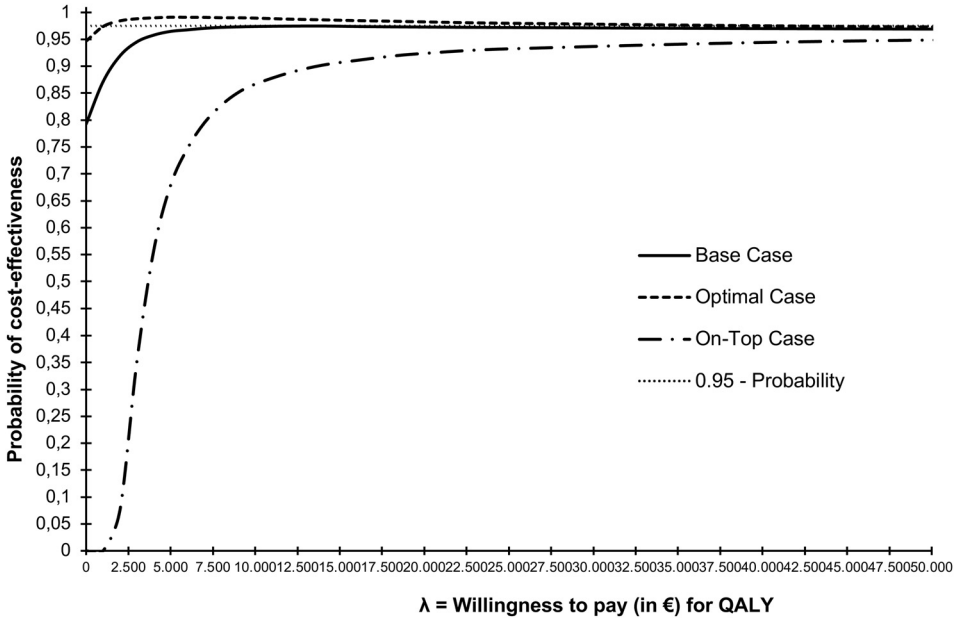

Figure 5. Net-Monetary Benefit - cost-effectiveness acceptability curves for all scenarios of SCCM vs TAU concerning QALY.

Note. Probability that SCCM intervention is acceptable (values on the vertical axis) in relation to TAU on the willingness to pay for an additional quality adjusted life year, given varying thresholds for willingness to pay (horizontal axis) based on 10,000 bootstrapped ICER replications. The small dotted line (0.95 − probability) indicates the upper 95%CI, i.e. the maximum amount that has to be invested to be confident that SCCM is cost-effective. Intersections of CEAC with the confidence line represents cost-effectiveness for a specific scenario.These λ were €11,874 and €20,000 for Base Case, and €1,100 and €30,000 for Optimal Case. CEAC for the On-top Case (representing cost of intervention only) asymptotically approximates the upper 95%CI, the higher the chosen WTP, but did not intersect the confidence line.

SCCM=Stepped and Collaborative Care, TAU=Treatment as Usual, QALY=Quality Adjusted Life Years, CEAC=Cost-Effectiveness Acceptability Curve.

that SCCM obtained an additional effect without any additional costs. Using different willingness to pay thresholds for BC, we found that the incremental CEAC intersects the upper confidence limit of 97.5% at €171, i.e. the maximum amount that had to be invested to make sure that SCCM is cost effective in relation to TAU (Figure 4). The robustness of BC results was confirmed with calculated ICER and incremental NMB for OC and IC (Figure 1 and Table S4). In-depth descriptions are provided in the supplementary material.

Utility. Average QALY of the SCCM conditions were slightly but insignificantly higher than average QALY of the TAU condition after one year; .501 (SD=.178) in SCCM and .480 (SD=.161) in TAU (Table S4). The ICER resulted in €-2,401.9 with large bootstrapped 95% CI ranging from €-22,521.6 to €11,873.7 (Table S4). The majority of bootstrapped ICER of BC were located in the southeast quadrant (76.2%) (Figure S3). This quadrant is associated with greater effects and lower costs of SCCM in relation to TAU. The CEAC of BC intersected the vertical axis at 0.8 in Figure 5. That means that SCCM probably obtained an additional QALY without any additional costs in relation to TAU. The CEAC of SCCM intersected the upper confidence interval twice at WTP values of €11,874 and €20,000. After the second intersection and thus, at WTP-values larger than €20,000, a prolonged decrease of the CEAC was observed with an asymptotic approximation to an imagined probability-line of 96% (Table S5). Hence, a maximum amount of €11,874 had to be invested to make sure that SCCM is cost-effective, but at amounts larger than €20,000 this assurance slowly decreased. We changed parameters with OC and IC, whereby the CEAC of OC revealed a similar pattern at different WTP values and the CEAC for the IC did not intersect the 97.5% confidence limit but asymptotically approximated an imaginary 96% probability line (Figure 5). In-depth descriptions are provided in the supplementary material.

Discussion

We found that patients treated within a Stepped and Collaborative Care Model (SCCM) showed a larger reduction of depressive symptoms and superior cost-effectiveness compared with TAU. Outcomes were mainly driven by effects of interventions provided for more severely affected participants, although a

reduction in depressive scores was also present in the TAU condition. Our study is in line with previous studies,[16,17] providing evidence for an intervention model that can strengthen healthcare systems and address challenges, such as scarcity of resources, shortage of staff and specialized expertise. A key strength of the present study lies in its representative sample involving several German study sites with unique clinical characteristics being examined at different time points in a randomized controlled trial. Another major strength lies in the development of five diverse and culturally sensitive digital, group and community-based interventions.

As a consequence of large migration movements, mental health institutions in Europe have faced vast challenges and burdens to their health care systems.[5] Against this background, the present model was developed as an approach to support refugees who face psychological distress.[12,27] SCCM was tailored to simultaneously overcome several obstacles by offering a set of culturally-sensitive psychological interventions. Although conflict exposure can result in a wide range of psychiatric disorders (depression, PTSD, anxiety and psychosis),[39] we followed a diagnostic approach in our study focusing on depressive symptoms, which are highly prevalent among refugee groups in Germany.[12]

While Germany's mental health care system is relatively robust and well-organized, it still has shortcomings, such as an ineffective allocation of existing psychotherapeutic resources, fragmentation and lack of coordinated care, and a scarcity of available flexible and integrative treatment models.[40,41] This became particularly evident in recent years, where the health care system was additionally challenged by a vast number of incoming asylum seekers and refugees. Therefore, our model might provide one possible approach and evidence on how to address challenges in situations where increased numbers of migrants are relocated, such as in the case of the Syrian civil war or the Ukrainian war. In agreement with our findings, previous stepped care studies demonstrated to save resources and to be clinically and cost-effective when compared with standard care.[16,42,43] As a part of England's Improving Access to Psychological Therapies (IAPT)[42] initiative, SCCM have also been incorporated with the goal of evaluating its organizational potential as an evidence-based promising approach that can improve and expand on the quality and access to available mental health care services. So far, preliminary investigations have demonstrated an improved use of available resources and increased recovery rates. Our SCCM was clinically effective at a lower cost and so we suggest that our model might be explored not only in settings with other ethnic minorities in high income countries, but also in low-and-middle-income-countries (LAMICs). Adaptations as indicated in already existing evidence (MANAS in India,[44] STEPCARE in Nigeria[42]) might be necessary since health systems in LAMICs are likely to be different to those in academic settings in Germany.

There are potential indicators that clinical efficacy of the SCCM might be driven by either pharmacological- and psychotherapeutic interventions that were offered by trained clinicians (psychiatrists and psychologists) at intervention levels 3 and 4. In a meta-analysis by Khan et al. (2010), including 45 trials, patients with higher depressive scores at baseline who were treated with antidepressants, showed greater improvements, compared with participants with lower initial symptom severity.[45] These findings suggest that interventions tailored to the needs of more severely depressed participants are more likely to be more effective than those for less severely ill. This is of importance since the effectiveness of regularly used internet-based CBT (often used in patients with less severe depression) has been demonstrated to be both clinically and cost effective[43] and a means to overcome stigma and language barriers.[45] Especially since web-based and community-based interventions have been reported to be less stigmatizing for individuals seeking psychiatric care,[49] making treatments more accessible. Future research (with sufficiently powered samples) is necessary to assess further beneficial dimensions besides ameliorating depressive symptoms, such as effects in reducing stigma around mental health treatment or increasing social support. before group interventions in SCCM could be filled and initialed.

Participants were initially assessed for symptom severity and their placement within the stepped model followed a more stratified stepped care approach, with only some participants being stepped up occasionally, depending on the availability of resources and time. There are several factors that play a role in the limited number of patients who were stepped up in this study. As other studies have already indicated,[16,18,19] several operational and organizational complications can arise in the implementation of stepped care models. Some of them included group interventions taking a long time to be filled up, participant motivation to continue in the study, dropouts due to the nature of this mobile population, difficulty reaching participants and time constraints due to a longer recruitment phase. Our results, thus, confirm the operational complexity in the service delivery of this model and shed light on the reality of how organizational pathways can influence the performance of stepped care systems, often leading to an alteration of initial plans. For this reason, a replication of the results including a larger sample size, more resources and a longer treatment period for interventions would be beneficial to confidently solidify the study outcomes.

The number of dropouts and the heterogeneity of dropout rates across all sites reflect a limitation of the present study. Building upon our findings, future research should consider reasons for dropout and ways

in which drop out and response rates can be improved in order to achieve better research outcomes. To account for differential effects, a PP analysis was conducted (see supplement), reflecting a reasonable sensitivity analysis that clearly indicates that, adjusting for site as a random effect in our models, treatment effects remain stable when considering dropouts. Given those different regions in Germany offered refugee placements at different times during the migration process in 2015 and onward, we also tested whether time since arrival in Germany (see Table 1) differed between sites, which was not the case ($p = .55$), however future research should consider range of time since arrival and its specific effects on intervention success. Further limitations are regional differences in health care between study sites, which may have led to considerable variability in recruitment, assessment, and intervention levels. Recruitment across all study sites was reported to be challenging due to several factors, i.e. mental health stigma, regional differences in health care provision and an uneven spread of refugee communities across Germany, leading to an oversampling of refugees in the larger cities. Furthermore, setting up a new SCCM will likely cause temporary disruptions in already existing workflows. Although this data was not systematically collected in the trial, we believe that it would be useful to gather information regarding organizational complexities and any noteworthy challenges related to the delivery of services in future implementations of similar projects. Nonetheless, since treatment gaps are not as pronounced in the German health care system compared with other regions, we expect that the model needs to be adapted if used in other international and humanitarian-aid setting. We initially planned and documented in our published study protocol to conduct a cluster-randomization, however as we could only conduct the intervention levels at two different sites inhibiting a cluster randomization. Finally, interpretations with respect to efficacy on the level of interventions or treatment steps can only be made very cautiously since the study was not sufficiently powered to address this issue.

In conclusion, the proposed SCCM intervention demonstrates clinical efficacy in reducing depressive syndromes in a sample of adult and adolescent refugees. Our study provides evidence for higher cost-effectiveness of the overall model. Further directions should investigate the effectiveness of SCCM's for other prevalent diseases, medical settings, and cultural contexts. Another promising context is the provision of the stepped care model in form of a digital stepped care model, i.e. a model in which most of the interventions are provided digitally. It can be assumed -albeit evidence is lacking- that this form of SCCM may increase resource efficacy. Our findings contribute to the development of models that can improve clinical productivity, decrease disease burden, and include marginalized communities.

Contributors

Kerem Böge: project management, project execution, investigation, development of interventions, data collection, data analysis, data interpretation, writing, review and editing
Carine Karnouk: literature review, project execution, data collection, development of interventions, writing, review and editing
Andreas Hoell: literature search, cost-effectiveness data collection, cost-effectiveness analysis, figures, writing,
Inge Kamp-Becker: design, funding acquisition, investigation, project supervision
Frank Padberg: study design, project supervision, development of intervention, review and editing throughout
Aline Übleis: study design, project supervision, development of intervention, review and editing throughout
Alkomiet Hasan: study design, project supervision, review and editing throughout
Peter Falkai: funding acquisition,study design, project supervision, review and editing throughout
Hans-Joachim Salize: design, funding acquisition, software, project supervision, review and editing
Andreas Meyer-Lindenberg design, funding acquisition, conceptualisation, review and editing
Tobias Banaschweski: design, funding acquisition, investigation, project supervision, review and editing
Frank Schneider: design, funding acquisition, conceptualisation, review and editing
Ute Habel: design, funding acquisition, development of intervention, investigation, project supervision, review and editing
Paul Plener: design, funding acquisition, conceptualisation, review and editing
Eric Hahn: investigation, project supervision, development of intervention, original paper draft outline, review and editing throughout
Mira Tschorn: software, data analysis, data interpretation und writing, review and editing
Maren Wiechers: project management, investigation, development of interventions, data analysis, review and editing
Michael Strupf: project management, investigation, development of interventions, data analysis, review and editing
Andrea Jobst: design, funding acquisition, investigation
Sabina Millenet: design, funding acquisition, investigation
Edgar Hoehne: project management, investigation, development of interventions, data analysis, review and editing
Thorsten Sukale: investigation, review and editing
Raphael Dinauer: investigation, review and editing
Martin Schuster: investigation, review and editing
Nassim Mehran: investigation, review and editing
Franziska Kaiser: project management, investigation, development of interventions, review and editing
Stefanie Bröcheler: project management, investigation, development of interventions, review and editing

Klaus Lieb: investigation, review and editing
Andreas Heinz: design, funding acquisition, development of intervention, investigation, conceptualisation, review and editing
Michael Rapp: design, funding acquisition, software, formal analysis, methodology, review and editing
Malek Bajbouj: design, funding acquisition, investigation, development of intervention, project supervision, original paper draft outline, review and editing throughout.

Declaration of interests

Dr. Banaschewski served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Takeda, and Infectopharm. He received conference support or speaker's fee from Lilly, Medice, and Takeda. He received royalties from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press; the present work is unrelated to these relationships. Alkomiet Hasan has been invited to scientific meetings by Lundbeck, Janssen, and Pfizer, and he received paid speakerships from Desitin, Janssen, Otsuka, and Lundbeck. He was a member of Roche, Otsuka, Lundbeck, and Janssen advisory boards. Paul Plener was involved in clinical trials of Lundbeck and Servier. He received a speaker's honorarium from Shire and Infectopharm. Frank Padberg is a member of the European Scientific Advisory Board of Brainsway Inc., Jerusalem, Israel, and has received speaker's honoraria from Mag&More GmbH and the neuroCare Group. His lab has received support with equipment from neuroConn GmbH, Ilmenau, Germany, and Mag&More GmbH and Brainsway Inc., Jerusalem, Israel. The other authors declare no competing interests.

Data sharing statement

The trial data can be requested deidentified and anonymised by researchers for future usage in independent scientific research projects. These requests should be addressed to the corresponding author to negotiate a data-sharing agreement with the Charité – Universitätsmedizin Berlin.

Acknowledgements

This project is funded by the Innovationsfond and German Ministry of Health [grant number 01VSF16061]. We thank members of the steering committee, including Dr. Frank Montgomery (German Medical Association), Dr. Thomas Zahn (AOK health insurance), Dr. Volker Westerbarkey (Medicine sans Frontiers), Dr. Matthias Mohrmann (AOK) and all participant, for taking part in the trial. We would also like to take the opportunity to thank all of members of the refugee population for their contribution and invaluable work for the MEHIRA study as official and inofficial advisors, as graphical designers for BALSAM, as peers in the peer-to-peer-intervention, as well as interns, translators, psychologists and research assistants.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.lanepe.2022.100413.

References

1 Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. *N Engl J Med*. 2020;383(6):510–512. https://doi.org/10.1056/NEJMp2008017.
2 Tol WA, Leku MR, Lakin DP, et al. Guided self-help to reduce psychological distress in South Sudanese female refugees in Uganda: a cluster randomised trial. *Lancet Glob Health*. 2020;8(2):e254-e263. https://doi.org/10.1016/S2214-109X(19)30504-2.
3 Berwick DM, Shine K. Enhancing private sector health system preparedness for 21st-century health threats: foundational principles from a national academies initiative. *JAMA*. 2020;323(12):1133–1134. https://doi.org/10.1001/jama.2020.1310.
4 World Health Organization. *The European Mental Health Action Plan 2013–2020*. 2015.
5 Jefee-Bahloul H, Bajbouj M, Alabdullah J, Hassan G. Barkil-Oteo A. Mental health in Europe's Syrian refugee crisis. *Lancet Psychiatry*. 2016;3(4):315–317. https://doi.org/10.1016/S2215-0366(16)00014-6.
6 King JS. Covid-19 and the Need for Health Care Reform. *N Engl J Med*. 2020;382(26):e104. https://doi.org/10.1056/NEJMp2000821.
7 Schneider R. *BAMF-Kurzanalyse "Volljährige Asylerstantragsteller in Deutschland im Jahr 2016 – Sozialstruktur, Qualifikationsniveau und Berufstätigkeit", Stand April 2017*. 2017.
8 *United Nations High Commissioner for Refugees (UNHCR)*. Operational Portal: Refugee Situations; 2022. https://data2.unhcr.org/en/situations/mediterranean#_ga=2.127672137.595711140.1594397605-1306335004.1593538956&_gac=1.19467978.1593538956.CjwKCAjwxev3BRBBEiwAiB_PWNEoeg-iK2ESaX6sjXNyqVoYYDSf8fMJx6wjpiG82pgE9fhn2WxqHxoCZ90QAvD_BwE.
9 UNHCR. *Global Trends Forced Displacement in 2019*. 2020. https://www.unhcr.org/5ee200e37.pdf.
10 Georgiadou E, Zbidat A, Schmitt GM, Erim Y. Prevalence of mental distress among syrian refugees with residence permission in Germany: a registry-based study. *Front Psychiatry*. 2018;9:393. https://doi.org/10.3389/fpsyt.2018.00393.
11 Walther L, Kroger H, Tibubos AN, et al. Psychological distress among refugees in Germany: a cross-sectional analysis of individual and contextual risk factors and potential consequences for integration using a nationally representative survey. *Bmj Open*. 2020;10:(8) e033658. https://doi.org/10.1136/bmjopen-2019-033658.
12 Hoell A, Kourmpeli E, Salize HJ, et al. Prevalence of depressive symptoms and symptoms of post-traumatic stress disorder among newly arrived refugees and asylum seekers in Germany: systematic review and meta-analysis. *BJPsych Open*. 2021;7(3):e93. https://doi.org/10.1192/bjo.2021.54.
13 Kiselev N, Morina N, Schick M, Watzke B, Schnyder U, Pfaltz MC. Barriers to access to outpatient mental health care for refugees and asylum seekers in Switzerland: the therapist's view. *Bmc Psychiatry*. 2020;20(1):378. https://doi.org/10.1186/s12888-020-02783-x.
14 Bhugra D, Gupta S, Bhui K, et al. WPA guidance on mental health and mental health care in migrants. *World Psychiatry*. 2011;10(1):2–10. https://doi.org/10.1002/j.2051-5545.2011.tb00002.x.
15 Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. *Br J Psychiatry*. 2005;186(1):11–17. https://doi.org/10.1192/bjp.186.1.11.
16 Harter M, Watzke B, Daubmann A, et al. Guideline-based stepped and collaborative care for patients with depression in a cluster-randomised trial. *Sci Rep*. 2018;8(1):9389. https://doi.org/10.1038/s41598-018-27470-6.

17 Ho FY, Yeung WF, Ng TH, Chan CS. The Efficacy and Cost-Effectiveness of Stepped Care Prevention and Treatment for Depressive and/or Anxiety Disorders: A Systematic Review and Meta-Analysis. *Sci Rep*. 2016;6:29281. https://doi.org/10.1038/srep29281.

18 Boyd L, Baker E, Reilly J. Impact of a progressive stepped care approach in an improving access to psychological therapies service: An observational study. *PLoS One*. 2019;14:(4) e0214715. https://doi.org/10.1371/journal.pone.0214715.

19 Richards DA, Bower P, Pagel C, et al. Delivering stepped care: an analysis of implementation in routine practice. *Implement Sci*. 2012;7(1):3. https://doi.org/10.1186/1748-5908-7-3.

20 Trilesnik B, Altunoz U, Wesolowski J, et al. Implementing a Need-Adapted Stepped-Care Model for Mental Health of Refugees: Preliminary Data of the State-Funded Project "RefuKey". *Front Psychiatry*. 2019;10:688. https://doi.org/10.3389/fpsyt.2019.00688.

21 Rosner R, Sachser C, Hornfeck F, et al. Improving mental health care for unaccompanied young refugees through a stepped-care approach versus usual care+: study protocol of a cluster randomized controlled hybrid effectiveness implementation trial. *Trials*. 2020;21(1):1013. https://doi.org/10.1186/s13063-020-04922-x.

22 Nicholas J, Ringland KE, Graham AK, et al. Stepping Up: Predictors of 'Stepping' within an iCBT Stepped-Care Intervention for Depression. *Int J Environ Res Public Health*. 2019;16(23):4689. https://doi.org/10.3390/ijerph16234689.

23 Haaga DA. Introduction to the special section on stepped care models in psychotherapy. *J Consult Clin Psychol*. 2000;68(4):547–548.

24 Richards D. United Kingdom perspective. In: Gilbody S, Bower P, eds. *Depression in primary care: evidence and practice*. Cambridge University Press; 2010:127–136.

25 Hinton DE, Patel A. Culturally sensitive CBT for refugees: key dimensions. In: Morina N, Nickerson A, eds. *Mental Health of Refugee and Conflict-Affected Populations*. Springer International Publishing; 2018:201–219. chap 10.

26 Rathod S, Gega L, Degnan A, et al. The current status of culturally adapted mental health interventions: a practice-focused review of meta-analyses. *Neuropsychiatr Dis Treat*. 2018;14:165–178. https://doi.org/10.2147/NDT.S138430.

27 Böge K, Karnouk C, Hahn E, et al. Mental health in refugees and asylum seekers (MEHIRA): study design and methodology of a prospective multicentre randomized controlled trail investigating the effects of a stepped and collaborative care model. *Eur Arch Psychiatry Clin Neurosci*. 2020;270(1):95–106. https://doi.org/10.1007/s00406-019-00991-5.

28 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

29 Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *J Adolesc Health*. 2002;30(3):196–204. https://doi.org/10.1016/S1054-139x(01)00333-0.

30 Hollifield M, Verbillis-Kolp S, Farmer B, et al. The Refugee Health Screener-15 (RHS-15): development and validation of an instrument for anxiety, depression, and PTSD in refugees. *Gen Hosp Psychiatry*. 2013;35(2):202–209. https://doi.org/10.1016/j.genhosppsych.2012.12.002.

31 Galenkamp H, Stronks K, Snijder MB, Derks EM. Measurement invariance testing of the PHQ-9 in a multi-ethnic population in Europe: the HELIUS study. *Bmc Psychiatry*. 2017;17(1):349. https://doi.org/10.1186/s12888-017-1506-9.

32 Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59(Suppl 20(20)):22–33. quiz 34-57.

33 Montgomery S, Åsberg M. *A New Depression Scale Designed to be Sensitive to Change*. Acad. Department of Psychiatry, Guy's Hospital; 1977.

34 Dixius A, Möhler E. START—entwicklung einer intervention zur erststabilisierung und arousal-modulation für stark belastete minderjährige flüchtlinge. *Praxis der Kinderpsychologie und Kinderpsychiatrie*. 2017;66(4):277–286.

35 Wiechers M, Übleis A, Padberg F. *Empowerment für Menschen mit affektiven Erkrankungen und Migrationserfahrungen: Therapiemanual für Einzel- und Gruppensettings*. 2019. Klett-Cotta.

36 Salize HJ, Kilian R. *Health Economics in Psychiatry. Concepts, Methods, Analyses*. 2010. Kohlhammer.

37 Statistisches Bundesamt. *Prices. Consumer Price Index for Germany. Long Series from 1948 to January 2021*. 2021.

38 World Health Organization. The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. *Social Sci Med*. 1998;46(12):1569–1585.

39 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(10194):240–248. https://doi.org/10.1016/S0140-6736(19)30934-1.

40 Johne J, von Peter S, Schwarz J, Timm J, Heinze M, Ignatyev Y. Evaluation of new flexible and integrative psychiatric treatment models in Germany- assessment and preliminary validation of specific program components. *Bmc Psychiatry*. 2018;18(1):278. https://doi.org/10.1186/s12888-018-1861-1.

41 Neumann A, Hense H, Baum F, et al. Evaluation of a flexible and integrative psychiatric care model in a department of child and adolescent psychiatry in Tubingen, Germany: study protocol (EVA_TIBAS). *BMC Health Serv Res*. 2021;21(1):1262. https://doi.org/10.1186/s12913-021-07226-1.

42 Gureje O, Oladeji BD, Montgomery AA, et al. Effect of a stepped-care intervention delivered by lay health workers on major depressive disorder among primary care patients in Nigeria (STEPCARE): a cluster-randomised controlled trial. *Lancet Glob Health*. 2019;7(7):e951–e960. https://doi.org/10.1016/s2214-109x(19)30148-2.

43 Mohr DC, Lattie EG, Tomasino KN, et al. A randomized noninferiority trial evaluating remotely-delivered stepped care for depression using internet cognitive behavioral therapy (CBT) and telephone CBT. *Behav Res Ther*. 2019;123: 103485. https://doi.org/10.1016/j.brat.2019.103485.

44 Patel V, Weiss HA, Chowdhary N, et al. Effectiveness of an intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomised controlled trial. *Lancet*. 2010;376(9758):2086–2095. https://doi.org/10.1016/S0140-6736(10)61508-5.

45 Sijbrandij M, Acarturk C, Bird M, et al. Strengthening mental health care systems for Syrian refugees in Europe and the Middle East: integrating scalable psychological interventions in eight countries. *Eur J Psychotraumatol*. 2017;8:(sup2) 1388102. https://doi.org/10.1080/20008198.2017.1388102.