Preventing Child Mental Health Problems in Southeastern Europe: Feasibility Study

(Phase 1 of MOST framework)

Elena Jansen[1], Inga Frantz[2], Judy Hutchings[3], Jamie Lachman[4, 5], Margiad Williams[3], Diana Taut[6], Adriana Baban[6], Marija Raleva[7], Galina Lesco[8], Cathy Ward[9], Frances Gardner[4], Xiangming Fang[10], Nina Heinrichs[2*], Heather M. Foran[1*]

[1] Institute of Psychology, University of Klagenfurt, Klagenfurt, Austria. [2] Department of Psychology, Clinical Psychology and Psychotherapy, University of Bremen, Bremen, Germany.[3] Centre for Evidence Based Early Intervention, School of Psychology, Bangor University, UK [4] Department of Social Policy and Intervention, University of Oxford, United Kingdom [5] MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK [6] Department of Psychology, Babeș-Bolyai University, Cluj-Napoca, Romania [7] Institute for Marriage, Family and Systemic Practice – ALTERNATIVA, Skopje, Macedonia [8] Health for Youth Association, Chișinău, Republic of Moldova [9] Department of Psychology, University of Cape Town, Rondebosch, South Africa [10] School of Public Health, Georgia State University, Atlanta, Georgia, USA

*Shared senior authorship: Heather M. Foran & Nina Heinrichs.

Nina Heinrichs: Department of Psychology, Clinical Psychology and Psychotherapy, University of Bremen, Bremen, Germany. Questions concerning the RISE study should be addressed to Heather M. Foran, Institute of Psychology, University of Klagenfurt, Universitaetsstr. 65-67, 9020 Klagenfurt, Austria. Heather.foran@aau.at

Abstract

The prevalence of child emotional and behavioral problems is an international problem but is higher in low and middle-income countries (LMIC) where there are often less mental health supports for families. Parenting programs can be an effective means of prevention, but must be low-cost, scalable, and suitable for the local context. The RISE project aims to systematically adapt, implement and evaluate a low-cost parenting program for preventing/reducing child mental health problems in three middle-income countries in Southeastern Europe. This small pre-post pilot study is informed by the Reach, Efficacy, Adoption, Implementation, and Maintenance (RE-AIM) framework and tested the feasibility of the intervention, the implementation and evaluation procedures: Phase 1 of the three-phase Multiphase Optimization Strategy (MOST) for program adaptation. Local facilitators delivered the Parenting for Lifelong Health (PLH) for Young Children program to parents of children aged 2 to 9 in North Macedonia, the Republic of Moldova, and Romania in 2018. Parents completed assessments pre- and post-program. Results demonstrated positive pre-post change for participating families ($N = 140$) on various outcomes including child externalizing and internalizing symptoms and parenting behavior, in all three countries, all in the expected direction. Program participation was associated with positive outcomes in participating families. Based on the experiences of this pilot study, we outline the practical implications for the successful implementation of parenting programs in the three countries that will inform our next study phases, factorial experiment and RCT.

Key words: child behavior problems, child maltreatment, parenting program, PLH, LMIC

The prevalence of child mental health problems remains high despite advances in mental health research and delivery. Globally, where data are available, rates of emotional (e.g., anxiety, depression) and/or behavioral problems (e.g., attention problems, aggression) for children and adolescents range between 10-20 percent (e.g., Kieling et al., 2011). These individuals are also at risk of additional problems such as learning difficulties or experiencing child maltreatment (Stith et al., 2009). Left untreated, children's mental health problems (especially externalizing symptoms) tend to persist (Kessler et al., 2005), leading to later social problems such as antisocial behavior, juvenile delinquency, and intimate partner violence (Patterson, DeBaryshe, & Ramsey, 1989).

Prevention of child mental health problems is critical for reducing the burden on individual children and families (through primary prevention and preventing exacerbation of existing mental health problems), and for reducing associated costs carried by states. This is especially important in low- and middle-income countries (LMIC) with a greater number of risk factors for child mental health problems (e.g., parental stress, poverty) (Kieling et al., 2011; Knerr, Gardner, & Cluver, 2013; Petersen et al., 2016), and limited access to specialist care (Murray et al., 2014; Ordonez & Collins, 2015). Risk factors include biological (e.g., physical health), and social (e.g., school environment) factors, as well as parental (e.g., mental health issues) and child (e.g., personality traits) characteristics, and interactions within the family (e.g., violence, parenting practices; Wlodarczyk et al., 2017). Amongst risk factors, parent-child interactions seem crucial to target, because they 1) impact daily on the child, 2) are changeable, and 3) modifying these interactions has the capacity to also change other family interaction patterns such as couple and sibling relationships. Parenting programs increase positive parenting practices, reduce dysfunctional parenting and are effective in reducing child mental health problems and enhancing children's quality of life. Some studies show effect maintenance months or years after program participation (Furlong et al., 2012;

Leijten et al., 2018). In particular, a strong body of research shows the effectiveness of parenting programs based on social learning theory derived from the Hanf two-stage model of promoting positive parenting behaviors prior to learning effective, nonviolent behavior management (Gardner, Montgomery, & Knerr, 2016; Kaehler, Jacobs, & Jones, 2016; Leijten et al., 2018). Ideally these programs change dysfunctional parent-child interactions and reduce child mental health problems, but also enhance relationships and functioning within the whole family (e.g., less conflicts amongst the couple about parenting, less stress and anxiety in parents). However, most of this research (using gold standard randomized-controlled trials with large samples) was conducted in high-income countries (HIC) (Kieling et al., 2011). Nevertheless, although there is increasing evidence of effectiveness in LMIC (Knerr et al., 2013; Pedersen et al., 2019), and of transportability across cultures (Leijten, Melendez-Torres, Knerr, & Gardner, 2016), considerable barriers to program reach, adoption, implementation, and maintenance in low-resource settings remain (Glasgow, Vogt, & Boles, 1999; Mikton, 2012). Barriers include a) issues of cultural acceptability and contextual relevance, b) limited financial and human resources to implement and sustain programs, c) resistance of developers to adaptation, and d) challenges in integrating programs into existing service delivery systems (Parra-Cardona et al., 2018). As a result, it is important that formative research is conducted when transporting interventions to new settings in order to identify strategies that enhance the effective and sustainable dissemination in local contexts.

The Parenting for Lifelong Health for Young Children (PLH-YC) program is a promising candidate for transferability across different countries and contexts. Originally developed and tested in two RCTs in South Africa (N = 68, Lachman et al., 2017; N = 296, Ward et al., 2020), PLH-YC demonstrated initial promising effects on improving child behavior and reducing harsh parenting. However, PLH-YC faces considerable barriers to

scale-up due to its length, complexity, and overall cost suggesting that it requires additional optimization in order to increase its effectiveness and scalability.

The overall aim of the RISE project is to adapt, pilot, optimize, and evaluate a low-cost parenting program in three LMICs in Southeastern Europe: North Macedonia, the Republic of Moldova, and Romania (Frantz et al., 2019). The study involves one of the largest systematic cross-cultural adaptations and evaluation of the program with the objective of producing an effective and cost-effective intervention that can be taken to scale. The study uses the Multiphase Optimization Strategy as a methodological framework to guide optimization (MOST, Collins, 2018). This framework aims to identify and test an optimized version of the program in three phases: The Preparation Phase (including a small pre-post feasibility trial – the subject of this paper), the Optimization Phase (in which different components are selected as candidates for inclusion in the optimized version) and the Evaluation Phase (in which the program components included in the Optimization Phase are tested in a randomized controlled trial RCT). Additionally, the reach, efficacy, adoption, implementation, and maintenance (RE-AIM) framework (Glasgow et al., 1999) is used to establish factors associated with the successful and sustainable implementation of the program within existing infrastructures.

The goal of the feasibility study (Preparation Phase of MOST) is to examine the (1) implementation feasibility including acceptability, fidelity, and quality of delivery, (2) indicative efficacy of the intervention in reducing child behavior problems and other risk factors (parenting behavior, parent health, family violence) based on differences between baseline and post-intervention assessment, and (3) feasibility and utility of the evaluation measures and indicators for subsequent phases. In this paper, we explore preliminary evidence to support the efficacy of the program in the three countries in a pre-post feasibility trial ($N = 40$ parents per country). More in-depths assessment of the implementation

feasibility including key challenges and facilitators for the program implementation in the three countries (using qualitative and quantitative measures) are described elsewhere (for details see Williams et al., submitted).

This paper focuses on the reach and preliminary efficacy of the intervention on primary, secondary and other-specified outcomes when delivered in North Macedonia, the Republic of Moldova and Romania. Previous research on parenting trainings in LMIC (Knerr et al., 2013; Petersen et al., 2016) including one RCT from Southeastern Europe (Dybdahl, 2001) suggests efficacy (e.g., effects on parenting behavior, parent-child relationship, child behavior problems). Thus, we also expect pre-post improvements in child behavior problems and parenting practices in our study. We additionally expect improvements in broader family functioning including relationships within (e.g., couple relationship) and outside the family (e.g., social support), as well as the wellbeing of other family members (e.g., parent's individual distress, child internalizing problems). Although a pre-post study cannot establish causality, pilot studies are useful in establishing potential intervention efficacy as well as in examining any potential risk for harm, such as differences between baseline and post-test in the opposite direction to that expected, or other adverse events (Thabane et al., 2010). Additionally, this paper examines how post-assessment outcomes are related to implementation characteristics (e.g., attendance rate), after controlling for baseline levels.

Method

Participants and Procedure

This is a single arm pre-post multisite trial. Data for the Pilot Feasibility Phase were simultaneously collected in North Macedonia, Republic of Moldova and Romania in 2018 (pre: April to June; post: September to December). Details of the RISE study protocol, recruitment and intervention have been described elsewhere (Frantz et al., 2019; Jansen et al., 2020). Parents of children aged 2-9 years were recruited through a number of strategies (e.g.,

flyers, non-governmental organizations, teachers) and pre-screened for eligibility. Before

commencing with the face-to-face assessment by trained interviewers, participants gave

consent and final eligibility was determined. Eligibility criteria included that the parent was

aged 18 years or older, lived in the same household as the target child (at least four nights per

week), planned to do so during the study, reported elevated levels of behavioral problems in

the target child, and agreed to participate in the PLH for Young Children program. Eligible

parents underwent baseline assessment, 22 (13.6%) were found to be ineligible after giving

consent. Assessments were completed using electronic tablets (computer-assisted self-

interviewing 'CASI'). Audio-CASI was used for sensitive questions (Phillips, Gomez, Boily,

& Garnett, 2010). Paper-and-pen questionnaires were available for participants who preferred

to not use the tablets. Parents received a food voucher or hygiene products after each set of

assessments (worth approximately €5 per participant), snacks and transportation allowance at

each program session that they attended, a certificate of participation, and an attendance

award (if ≤ 1 program session was missed). RISE was approved by the Human Research

Ethics Committee of the University of Klagenfurt and local ethics committees in North

Macedonia, the Republic of Moldova and Romania. The trial is registered on

ClinicalTrials.gov (ID: NCT03552250).

Intervention

PLH is an initiative of universities, in partnership with UNICEF and the WHO, to

develop a suite of parenting programs that meet the needs and constraints of LMICs (C. L.

Ward et al., 2014). This study examined the feasibility of two versions of the PLH program

for parents of children from 2 to 9 years (PLH for Young Children, or PLH-YC), varying in

the number of sessions and length. The program is based on social learning theory principles

with core session activities including illustrated comics, group discussions, practicing

parenting skills in role-plays, collaborative problem solving, and home activities. The

original version of PLH-YC program comprised 12 sessions (https://rise-plh.eu/work-packages/work-package-2/).

This pilot feasibility study involved adapting the program with input from various stakeholders in each country (e.g., parents, facilitators, practitioners, local governmental agencies, NGOs). Based on their feedback, PLH-YC manuals were adapted and translated for testing in each country. Local facilitators were trained during an intensive training workshop to deliver the interventions. They also received regular coaching sessions to ensure program fidelity. Facilitators conducted weekly phone calls with participants to support each family individually and to encourage home use of strategies. Facilitators sent weekly booster SMSs to remind parents to complete home activities and to praise them for trying.

Due to limited time to deliver the program before the summer holidays when families and service providers would not be available, PLH-YC was implemented in slightly different ways, in Moldova it was delivered as originally developed over 12 weekly sessions, with the exception of sessions 10 and 11 which were delivered within one week. In North Macedonia, the 12 sessions were delivered twice a week over six weeks. In Romania, a condensed 6-session version of the program was delivered on a weekly basis. This version had been piloted previously in South Africa. Although this meant that program implementation varied across contexts and deviated from the study protocol, it was delivered in all three countries with the planned number of parents without risking too much drop-out due to vacation.

Measures

First, we translated and back-translated all measures into the three local languages. This was an iterative process and done by expert child psychiatrist and psychologist partners in each country (e.g., after the translation, the MINI-KID was tested by assessors in each country, wording revisions made, and the measure then back translated). We also worked

with the developers of questionnaires (e.g., for CBCL). This led to an official version of the

CBCL now available in Moldovan, and new versions in Macedonian and Romanian.

Primary Outcome Measures

Child externalizing behavior problems were measured using the Child Behavior

Checklist (CBCL) (Achenbach & Rescorla, 2001). For the present study, the parent-report

versions for children aged 1½-5 (i.e., young) and 6-18 (i.e., old) were employed. Since

normative data, and hence T-scores, are not available in the languages and countries of this

study, the CBCL raw scores are presented separately for the two versions of the scale. The

externalizing symptoms subscale (CBCL 1½-5 version [y] = 24 items, range 0-48; CBCL 6-

18 version [o] = 35 items, range 0-70) and the *aggressive behavior* subscale ($CBCL_y$ = 19

items, range 0-38; $CBCL_o$ = 18 items, range 0-36) were used, both indicating more problems

with higher scores. Reliability and validity have been reported for multiple countries (Weisz,

Sigman, Weiss, & Mosk, 1993). Reliability for aggressive behaviors and externalizing

symptoms in the current study were α_y = .86 and α_o = .87, and α_y = .43 and α_o = .65

respectively. Although not mentioned in the trial registry, we used the aggressive behavior

subscale separately and in addition to the externalizing subscale post-hoc, because of the low

internal consistency of the externalizing symptoms subscale in our sample.

Additionally, the Mini International Neuropsychiatric Interview for Children and

Adolescents – Parent Version (MINI-KID-P) (Sheehan et al., 2010) was used. The MINI-

KID is a structured interview to evaluate the presence of current psychiatric disorders (based

on DSM-5 with corresponding ICD-10CM codes; using a binary "yes/no" format). The

interview is organized in disorder-specific modules and uses screening questions for each

disorder. Data assessors, trained in administering the clinical interview, conducted the MINI-

KID. According to the interview guidelines, the assessor decided if each criterion for the

diagnosis of ODD or CD were met or not (e.g., ODD criterion A: at least 4 out of 8 behaviors

met yes/no). Based on the parent's response on the initial screening question for externalizing disorders, the assessor asked additional questions (referring to the diagnostic criteria) or skipped the section, consequently administration time varied greatly and could take up to 45 minutes. For the current study, results of the conduct disorder and oppositional defiant disorder modules were combined to reflect the prevalence of externalizing disorders.

Secondary Outcome Measures

The CBCL was also used to assess *child emotional problems* using the *internalizing symptoms* subscale (CBCL$_y$ = 31 items; range 0-62; CBCL$_o$ = 32 items; range 0-64) with higher scores indicating more emotional problems. Reliability for internalizing symptoms in the current study was α_y = .79 and α_o = .78. We were not able to assess emotional problems in children using the structured interview MINI-KID as originally planned (Frantz et al., 2019) since the total assessment was too long (ca. 1.5-2h). To prevent parents dropping out because of assessment length as well as invalid data due to assessor/parent fatigue, we decided to assess child emotional outcomes with one indicator (CBCL internalizing subscale). Also, while the Parenting Stress Index is listed in the trial registration, it was not used, because US licensing restrictions made its use in LMIC too expensive.

Parenting was assessed with the Parenting Scale (PS; i.e., *dysfunctional discipline practices*) (Arnold, O'leary, Wolff, & Acker, 1993) and the Parenting of Young Children Scale (PARYC; *positive parent behavior and effective discipline)* (McEachern et al., 2012). The total PS score (30 items, α = .61) and the three subscales – laxness (11 items, α = .66), over-reactivity (10 items, α = .72) and verbosity (7 items, α = .42) – were used. Similarly, the total PARYC score (21 items, α = .85) and the three subscales – positive parenting (7 items, α = .73), setting limits (7 items, α = .75) and proactive parenting (7 items, α = .76) – were used.

Harsh Parenting was assessed using 14 items based on the ISPCAN Child Abuse Screening Tool-Intervention scale (ICAST-I) (Runyan, Dunne, & Zolotor, 2009; Runyan,

Dunne, Zolotor, et al., 2009) and the child maltreatment screener by Slep et al. (2013). Four

items assessed physical abuse (e.g., "In the past 4 weeks, how often did you discipline your

child by slapping, spanking, or hitting with your hand?"), seven items assessed emotional

abuse (e.g., "In the past 4 weeks, how often did you shout, yell or scream at your child?") and

three items assessed neglect (e.g., "How often in the past month did your child not get the

food or drink that he/she needed even when there was money to pay for it?"). The overall

score for harsh parenting is used based on summing all items (possible range 0-112), as well

as for each individual subscale (possible ranges: physical 0-32, emotional 0-56, neglect 0-24).

Psychological distress in parents was assessed using the Depression, Anxiety and

Stress Scale (DASS, 21 items) (Lovibond & Lovibond, 1995). The overall score was used

and higher scores indicate more psychological distress. Previous studies have shown good

internal reliability (Henry & Crawford, 2005). Reliability in the current study was α = .91.

Parental well-being was measured with the WHO-5 Well-Being Scale (Staehr, 1998).

We used the percentage score (World Health Organization, 1998). Higher scores indicated

better well-being. Previous studies have shown good internal reliability (Topp, Ostergaard,

Sondergaard, & Bech, 2015). Reliability in the current study was α = .77.

Other Pre-Specified Outcome Measures

Intimate partner violence was assessed using 29 questions (15 victimizing and 14

perpetrating behaviors) from the Brief Screening Instrument for Partner Maltreatment by

Heyman and colleagues (2013) and adapted from the revised Conflict Tactics Scale (CTS2S)

(Straus & Douglas, 2004). Overall victimization and perpetration frequencies in the past

month were used as a measure of severity, with ranges from 0-112 and 0-104 respectively.

Family functioning was assessed with the general functioning subscale of the Family

Assessment Device short form (FAD, 12 items) (Epstein, Baldwin, & Bishop, 1983).

Responses on each item (ranging from 1 = "Strongly agree" to 4 = "Strongly disagree") were

averaged after reverse coding where appropriate. Higher mean scores indicated more problems in family functioning. Reliability in the current study was α = .81.

Social support was measured using the emotional support subscale of the Medical Outcome Study Social Support Survey (MOS-SSS, 8 items) (Sherbourne & Stewart, 1991). Parents report how often they receive emotional support (e.g., "Someone you can count on to listen to when you need to talk") using a 5-point Likert scale (1 = "None of the time" to 5 = "All of the time"). Higher mean scores indicate more emotional support. Reliability in the current study was α = .96.

Parental relationship quality was assessed with the three items of the Kansas Marital Satisfaction Scale (Grover, Paff-Bergen, Russell, & Schumm, 2016). Items were rated on a scale from 1 (extremely dissatisfied) to 7 (extremely satisfied). An overall partner satisfaction score was calculated by summing up all items. Reliability in the current study was α = .98.

Implementation Characteristics

Four implementation characteristics were assessed based on facilitator or process monitor-report. These included weekly session attendance, frequency of phone calls to parents, frequency of SMS messages to parents, and weekly report of home activities completion. Percentages (based on planned delivery and varying by number of sessions and delivery density within each country) were calculated for all four implementation characteristics and used in current analysis.

Adverse Event Assessment

Feasibility of the adverse events (AE) assessment procedure was tested. We aimed to assess the occurrence of potential adverse events (AE; new symptom or worsening of an existing symptom), serious adverse events (SAE; e.g., life-threatening), or Unanticipated Problems Involving Risk to Subjects or Others (i.e., AE that are related to the study, unexpected, and involve a greater risk of harm). At post-assessment, AE were assessed using

an open question (*How are you doing – have you or your child had any problems since the last contact with anyone from the project team?*). If a parent answered "*yes*", the research team asked for additional information and rated the severity of the parent-reported AE and potential relatedness to the study. In a second step, the AE report was adjudicated by the local PI. The AE procedures were informed by the procedures of the STRONG STAR (Peterson, Roache, Raj, & Young-McCaughan, 2013) and the GROW&TREAT study (Job et al., 2020).

Data Analysis

Patterns of missingness were examined to determine missing data treatment. Comparisons across dropouts ($n = 46$) and completers ($n = 94$) at pre-assessment were conducted using ANOVA and chi-square tests. Only one difference (higher well-being total score in non-completers, $p = .02$) was found. There were no significant differences in socio-demographic characteristics, primary, secondary or other pre-specified outcomes. Therefore, analysis with full information maximization likelihood (FIML) was deemed appropriate. For purposes of full transparency, pre-post comparisons using ANCOVA with listwise deletion of cases with missing data (i.e., remaining total $n = 94$) are provided in Supplemental Materials. Study country was entered as a covariate (i.e., entered as two dummy variables with North Macedonia [largest sample size] as references group: Dummy1 compares Moldova to North Macedonia; Dummy2 compares Romania to North Macedonia). Mean differences between pre- and post-assessment on primary, secondary and other outcomes were examined next using the equivalence of paired samples t-test in a structural equation modeling context and conducted in Mplus Version 8.0 (Muthén & Muthén, 1998-2017). This allowed including country as two covariates, running analysis with FIML ($N = 140$), and utilizing the robust maximum likelihood (MLR) estimator to account for issues with normality. The Wald test was used to test whether the difference between post- and pre-assessment was significantly

different from zero. Estimation of Cohen's d effect sizes for all outcome measures was based on: small = 0.2–0.5, medium = 0.5–0.8, and large = 0.8 or higher (Cohen, 1988).

To explore whether post-assessment outcomes were related to implementation characteristics, associations between the four characteristics (percentage attendance, SMS, phone calls and home activity; adjusted for the 12 or 6 sessions offered) and the primary, secondary and other outcomes were examined next. Bivariate correlations were examined and those that were significant ($p < 0.01$) were then followed up with regression analysis including the primary/secondary/other outcome at post-assessment as dependent variable and the implementation characteristics as independent variable, while adjusting for the pre-assessment value of the dependent variable as well as study country.

Results

In total, 140 parents provided baseline data and enrolled in the study (see Figure 1 for Participant flow diagram). Sample characteristics are presented in Table 1. Of the enrolled parents, 21 dropped out before allocation to groups, reducing the number to 119. There were no significant differences between those who dropped out between the pre-assessment and group allocation in terms of education level, country, or literacy. There was a trend for dropouts to be older (mean age: 38 vs. 35 years, $p = 0.056$). Ninety-four parents were assessed at post-test (67% study retention, below the pre-defined acceptable cut-off of 80%).

Pre-Post Comparisons

Table 2 shows pre-post comparisons for all participants. Mean differences of primary, secondary and other pre-specified outcome variables were significantly different from zero, adjusting for study country (Cohen's $d = 0.20$ to 1.27). The only exception was parent-report of social support. All mean differences were in the expected direction. After the program, parents reported improvements in positive parenting behavior, well-being and relationship satisfaction and decreased child aggressive behaviors, externalizing symptoms and

internalizing symptoms, parental psychological distress, use of harsh parenting, dysfunctional discipline practices, intimate partner violence and family dysfunction.

In total, 78 children aged between 6-9 years were assessed with the MINI-KID at pre-assessment, while 51 (65.4%) provided data at post-assessment. The prevalence of externalizing disorders (i.e., conduct disorder and/or oppositional defiant disorder) was 12/78 (15.4%) at pre- and 2/51 (3.9%) at post-test.

Implementation Characteristics and Post-Assessment Outcomes

To combine results for the three countries, percentages are shown, acknowledging variation in number and frequency of sessions (i.e., North Macedonia: 12 sessions in 6 weeks, Moldova: 12 sessions in 11 weeks, Romania: 6 sessions in 6 weeks). Overall enrolment was 79%, with 110 families attending at least one session out of 140 families that were recruited into the program. Of the 119 participants allocated to groups, one dropped out before sessions commenced. The participation rate was $M = 72\%$, $SD = 27\%$ (median 83%; mean/median attendance rate of those parents attending at least one session). Facilitators sent fewer SMS messages and made fewer phone calls than planned: The mean percentage of SMS messages was $M = 61\%$, $SD = 36\%$ (median 73%); 16% of participants did not receive any text messages. The mean percentage of phone calls was $M = 44\%$, $SD = 35\%$ (median 38%), 22% did not receive any phone calls. The mean percentage of completed home activities was $M = 55\%$, $SD = 35\%$ (median 64%), with 18% not completing any (including the 7% of participants that did not attend any session where home activities were planned).

Table 3 shows the unadjusted correlations between implementation characteristics and post-assessment outcomes. After adjusting for baseline values and countries no relationship remained significant at $p < 0.01$. However, using $p < 0.05$, five relationships remained: higher attendance was associated with lower internalizing symptoms for children aged 2-5 years ($\beta = -.463$, $p = .027$) at post-assessment; higher numbers of SMS sent by the facilitator

were associated with higher scores on the aggressive behavior and the externalizing symptoms subscales of the CBCL for children aged 2-5 years (β = .343, p = .033 and β = .357, p = .030 respectively); higher numbers of completed home activities were associated with lower scores on the externalizing symptoms subscale of the CBCL for children aged 6-9 years (β = -.327, p = .015) and higher scores on the PARYC overall score (β = .327, p = .030) at the post-assessment. No correlations were found with phone calls.

Adverse Events

Twenty adverse events were reported by 17 participants with one reporting two events and one three events. Eight AE were reported during the intervention and 12 during the post-assessment. Fifteen events were rated as adverse, (e.g., newly occurring physical symptoms associated with parental hypertension or newly occurring depressive symptoms in response to the death of a relative), five as serious adverse events (usually require hospital admission, e.g., child hospitalization because of a severe flu). No event was study-related.

<div align="center">**Discussion**</div>

This is the first multi-country regional feasibility pilot of a parenting intervention aimed at reducing child mental health problems in Southeastern European LMIC. Preliminary pre-post differences in child mental health and parenting outcomes suggest large to medium effect sizes for all outcomes and are comparable to findings from other similar trials (Pedersen et al., 2019). This suggests that the program could be helpful in reducing children's risk for behavioral and emotional problems and preventing child maltreatment and dysfunctional parenting practices. There were no indications of negative effects based on pre-post analyses; none of the encountered AE were related to study participation. Larger differences were observed with primary outcomes (child externalizing behaviors, including aggression) than secondary (e.g., parenting behavior) or other outcomes (e.g., intimate partner violence) as would be expected given the program focus. However, before drawing

conclusions it is crucial to replicate this in the Optimization and Evaluation Phases using a larger sample, a follow-up assessment point and a control condition comparison.

Implementation characteristic results support the benefit of regular program attendance and completion of home activities since higher attendance and completion rates were associated with improved child behavior and parenting outcomes. Results relating to the SMS and phone calls were less conclusive and future empirical studies that examine engagement boosters as a causal mechanism are needed. We will test in Phase 2, if an enhanced engagement package (weekly text messages and phone consultation after each session) results in higher attendance rates, and better results on primary outcomes (parenting practices, child behavior problems) compared to a basic package (no communication boosters). Implementation characteristics and potential facilitators and barriers for parent engagement are described in more depth in a second paper (Williams et al., submitted).

Demographic and pre-assessment data show that participating families were in need of help, suggesting that the varied recruitment and implementation strategies were successful in reaching vulnerable families in the three LMIC. In sum, the results suggest that delivery of the program is feasible in North Macedonia, the Republic of Moldova and Romania. This is line with other findings showing that parenting programs can be successfully implemented in new cultural contexts (Knerr et al., 2013).

Strengths and Limitations

There are several limitations. The sample size is quite small, so results should be considered as preliminary. In addition, the samples are not representative of all parents in the three countries in which the program was tested, but rather a selected sample of parents, many of whose children had elevated behavioral symptoms. Further, the sample showed high socioeconomic disadvantage and is not necessarily representative of all parents with children at elevated risk for poor outcomes in Southeastern Europe. The most important limitation is

that the design tested feasibility and was not a randomized controlled study. Thus, it is not possible to attribute differences in pre-post assessments to the program. Another limitation, common in parenting program studies, is that few fathers participated. Finally, some implementation difficulties occurred and it is unclear how these impacted the results (e.g., number/timing of the sessions and program length across countries), in spite of controlling for country differences.

Despite these limitations, there are several strengths. This study forms part of a planned three-phase optimization trial and paves the way for the next stages. Results across outcomes were quite consistent and promising. Another strength is the rigorous approach to assessment in which measures not previously available in the study regions were adapted, (back-)translated, and psychometrically evaluated. The data analysis strategy was also more comprehensive than often used in pre-post feasibility studies, including both quantitative and qualitative data, intention-to-treat analyses with full information maximum likelihood estimation, as well as analysis of implementation characteristics and their relation to pre-post differences. This is the first multisite pilot feasibility study of a parenting program in Southeastern Europe. Given that cost-effective parenting programs in LMICs are sorely needed, it is essential that researchers systematically evaluate both program implementation and effectiveness. The RISE project, in which this study serves as the first phase, is designed exactly for that purpose by using an implementation science approach to examine both implementation and intervention outcomes, how they inform each other, and the cost-effectiveness of the intervention.

Lessons Learned and Implications For The Next Phases

Results of this study are valuable in informing the next phases in various ways: Although all eligible parents completed the informed consent, the information sheet was too complex for some of the participants. We thus plan to adapt the informed consent sheet using

plain language in subsequent phases. Some of the measures – although extensively used in other countries – did not work optimally across the three sites and internal consistencies were low for some (sub)scales (e.g., CBCL externalizing symptoms subscale). Consequently, the plan for the Optimization Phase is to replace or delete questionnaires that did not perform well in this study. Also, feedback from the country implementation teams revealed large heterogeneity between countries in the attractiveness of participation incentives, and difficulties in applying communication boosters (SMS, phone consultations) in the way originally planned due to preference for phone calls over SMS and lower numbers of (personally owned) cell phones, despite generally good internet reception and network availabilities. However, the overall participation rate (72%) was on a par with other behavioral parenting studies (Chacko et al., 2016). Given the association between higher attendance rates and completion of home activities and positive child and/or parent outcomes, it may be beneficial to enhance parental program participation and home activity reminders. In the next study phase, we plan to examine whether or not the amount of participation incentives and communication boosters impact both participation rate of the program and program effectiveness (Lachman et al., 2019). The surprise finding of a positive relationship between number of SMS sent by facilitators and child aggressive behavior and externalizing behavior seen here underscores this need to verify any potential negative or positive impact of SMS delivery during a parenting program. We examined potential country-specific effects (relationships remained positive and significant for Romania, North Macedonia at $p = .059$, though not for Moldova), however, this was not a statistically robust finding and should be interpreted with caution. When data were split by country, only 17-25 parents were left in each and the SMS variable was not normally distributed. Internal consistency of the CBCL externalizing behavior subscale was low. Tentative explanations are limited by the pre-post design. Parents who had children with higher levels of aggressive behavior possibly received

more SMS due to increased engagement with facilitators. The higher numbers of SMS may have increased parent intrusiveness resulting in increased child oppositional behavior or report thereof. A third variable might be responsible for the association seen between the two variables (e.g., higher parenting distress or parents spending more time on their phone in general).

Facilitator feedback suggests that the amount and intensity of facilitator training and supervision is highly relevant in the three countries. To allow future scalability it is important to have a low-cost program, yet facilitator training and supervision are relatively cost intensive. It is therefore imperative to understand which level of training and supervision intensity is most cost-effective. Therefore, we plan to test a fidelity booster package in the Optimization Phase, including weekly supervision sessions vs. supervision on-demand (Lachman et al., 2019).

Finally, this pilot study revealed difficulties in delivering a 12-session program in the countries given the timetable that was imposed by the research design. Implementation challenges included public holidays, fit with parents' working schedule, and some parents' preference for a shorter program. This insight is also a strength, as it allowed us to explore three different variations of program delivery (the original 12 weekly sessions, 12 sessions delivered 2x a week, and six weekly sessions). To identify the optimal length of the PLH-YC program in the three countries, we will test whether a shorter or longer program version is more efficient and cost-effective in the local context in the next phase (Lachman et al., 2019).

Experiences with the AE assessments suggested that it was challenging for the country teams to implement these due to the group format of the intervention, since assessment required one-to-one contact. In the Optimization Phase, we will use a structured AE checklist that can be filled in by parents individually with one-to-one adjudication administered in a follow-up call, if needed.

The retention rate at post-assessment was relatively low compared to other studies (e.g., Ward et al., 2020). Some of the encountered barriers and challenges included a) timing of assessments during or after the summer holidays, and b) retaining participants who did not enrol in or engage with the programme. To increase the participation rate at post- and follow-up assessments in the next phase, we plan to better time assessments (schedule assessments before or after the summer, Orthodox Easter and Christmas holidays), increase the financial incentives in addition to non-material incentives such as certificates of completion for post- and follow-up assessments, and send more reminders (e.g., PLH facilitators and data assessors send SMS reminders to parents). In addition, data assessors will be instructed to more clearly stress the study procedures and the importance of completing all study components during the informed consent stage.

Conclusions

The results of this pilot study suggest the feasibility of an intervention aimed at reducing child mental health problems in high-risk families in three LMIC in Southeastern Europe. In line with others (Thabane et al., 2010), we strongly recommend that researchers test the feasibility of interventions, the implementation framework and the assessment methods in a pilot study before conducting a comprehensive RCT – especially in LMIC where fewer research studies have been conducted and where things might work slightly differently compared to HIC. This process allows for further refinement of the intervention structure, delivery, and content taking into consideration key constrains as well as facilitators in the local context. As a result, researchers and program implementers can avoid wasting limited human and financial resources to make sure that the most cost-effective and locally relevant studies and interventions are implemented in order to achieve the common goal of improving the mental health of children and adolescents across the globe.

References

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth and Families.

Arnold, D. S., O'leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment, 5*(2), 137-144.

Chacko, A., Jensen, S. A., Lowry, L. S., Cornwell, M., Chimklis, A., Chan, E., . . . Pulgarin, B. (2016). Engagement in Behavioral Parent Training: Review of the Literature and Implications for Practice. *Clin Child Fam Psychol Rev, 19*(3), 204-215.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Collins, L. M. (2018). *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST)*. Cham, Switzerland: Springer International Publishing AG.

Dybdahl, R. (2001). Children and mothers in war: an outcome study of a psychosocial intervention program. *Child Dev, 72*(4), 1214-1230.

Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster family assessment device. *Journal of Marital and Family Therapy, 9*(2), 171-180.

Frantz, I., Foran, H. M., Lachman, J. M., Jansen, E., Hutchings, J., Baban, A., . . . Heinrichs, N. (2019). Prevention of child mental health problems in Southeastern Europe: a multicentre sequential study to adapt, optimise and test the parenting programme 'Parenting for Lifelong Health for Young Children', protocol for stage 1, the feasibility study. *BMJ Open, 9*(1), e026684.

Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database Syst Rev*(2), Cd008225.

Gardner, F., Montgomery, P., & Knerr, W. (2016). Transporting Evidence-Based Parenting Programs for Child Problem Behavior (Age 3-10) Between Countries: Systematic Review and Meta-Analysis. *J Clin Child Adolesc Psychol, 45*(6), 749-762.

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health, 89*(9),

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol, 44*(Pt 2), 227-239.

Heyman, R. E., Slep, A. M. S., Snarr, J. D., & Foran, H. M. (2013). Practical tools for assessing partner maltreatment in clinical practice and public health settings. In H. M. Foran, S. R. H. Beach, A. M. S. Slep, R. E. Heyman, & M. Z. Wamboldt (Eds.), *Family Problems and Family Violence: Reliable Assessment and the ICD-11* (pp. 43-70). New York: Springer.

Jansen, E., Lachman, J. M., Heinrichs, N., Hutchings, J., Baban, A., & Foran, H. M. (2020). Hunger in Vulnerable Families in Southeastern Europe: Associations With Mental Health and Violence. *Frontiers in Public Health, 8*(115).

Job, A.-K., Ehrenberg, D., Hilpert, P., Reindl, V., Lohaus, A., Konrad, K., & Heinrichs, N. (2020). Taking Care Triple P for Foster Parents With Young Children in Foster Care: Results of a 1-Year Randomized Trial. *J Interpers Violence, Online ahead of print*.

Kaehler, L. A., Jacobs, M., & Jones, D. J. (2016). Distilling Common History and Practice Elements to Inform Dissemination: Hanf-Model BPT Programs as an Example. *Clin*

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 593-602.

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., . . . Rahman, A. (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet, 378*(9801), 1515-1525.

Knerr, W., Gardner, F., & Cluver, L. (2013). Improving positive parenting skills and reducing harsh and abusive parenting in low- and middle-income countries: a systematic review. *Prev Sci, 14*(4), 352-363.

Lachman, J. M., Cluver, L., Ward, C. L., Hutchings, J., Mlotshwa, S., Wessels, I., & Gardner, F. (2017). Randomized controlled trial of a parenting program to reduce the risk of child maltreatment in South Africa. *Child Abuse Negl, 72*, 338-351.

Lachman, J. M., Heinrichs, N., Jansen, E., Brühl, A., Taut, D., Fang, X., . . . Foran, H. M. (2019). Preventing child mental health problems through parenting interventions in Southeastern Europe (RISE): Protocol for a multi-country cluster randomized factorial study. *Contemporary Clinical Trials, 86*, 105855.

Leijten, P., Melendez-Torres, G. J., Gardner, F., van Aar, J., Schulz, S., & Overbeek, G. (2018). Are Relationship Enhancement and Behavior Management "The Golden Couple" for Disruptive Child Behavior? Two Meta-analyses. *Child Dev, 89*(6), 1970-1982.

Leijten, P., Melendez-Torres, G. J., Knerr, W., & Gardner, F. (2016). Transported Versus Homegrown Parenting Interventions for Reducing Disruptive Child Behavior: A Multilevel Meta-Regression Study. *J Am Acad Child Adolesc Psychiatry, 55*(7), 610-617.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales*. Sydney: Psychology Foundation of Australia.

McEachern, A. D., Dishion, T. J., Weaver, C. M., Shaw, D. S., Wilson, M. N., & Gardner, F. (2012). Parenting Young Children (PARYC): Validation of a Self-Report Parenting Measure. *J Child Fam Stud, 21*(3), 498-511.

Mikton, C. (2012). Two challenges to importing evidence-based child maltreatment prevention programs developed in high-income countries to low-and middle-income countries: Generalizability and affordability. In Dubowitz H (Ed.), *World perspectives on child abuse and neglect* (Vol. 10, pp. 97). Aurora, CO: International Society for the Prevention of Child Abuse and Neglect.

Murray, L. K., Tol, W., Jordans, M., Zangana, G. S., Amin, A. M., Bolton, P., . . . Thornicroft, G. (2014). Dissemination and implementation of evidence based, mental health interventions in post conflict, low resource settings. *Intervention (Amstelveen), 12*(Suppl 1), 94-112.

Muthén, L., & Muthén, B. (1998-2017). *Mplus user's guide, eight edition*. Los Angeles, CA: Muthen & Muthen.

Ordonez, A. E., & Collins, P. Y. (2015). Advancing Research to Action in Global Child Mental Health. *Child Adolesc Psychiatr Clin N Am, 24*(4), 679-697.

Parra-Cardona, R., Leijten, P., Lachman, J. M., Mejía, A., Baumann, A. A., Amador Buenabad, N. G., . . . Domenech Rodríguez, M. M. (2018). Strengthening a Culture of Prevention in Low- and Middle-Income Countries: Balancing Scientific Expectations and Contextual Realities. *Prev Sci.*

Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *Am Psychol, 44*(2), 329-335.

Pedersen, G. A., Smallegange, E., Coetzee, A., Hartog, K., Turner, J., Jordans, M. J. D., . . . Studies, F. (2019). A Systematic Review of the Evidence for Family and Parenting

Interventions in Low- and Middle-Income Countries: Child and Youth Mental Health Outcomes. *28*(8), 2036-2055.

Petersen, I., Evans-Lacko, S., Semrau, M., Barry, M. M., Chisholm, D., Gronholm, P., . . . Thornicroft, G. (2016). Promotion, prevention and protection: interventions at the population- and community-levels for mental, neurological and substance use disorders in low- and middle-income countries. *Int J Ment Health Syst, 10*, 30.

Peterson, A. L., Roache, J. D., Raj, J., & Young-McCaughan, S. (2013). The need for expanded monitoring of adverse events in behavioral health clinical trials. *Contemp Clin Trials, 34*(1), 152-154.

Phillips, A. E., Gomez, G. B., Boily, M. C., & Garnett, G. P. (2010). A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. *Int J Epidemiol, 39*(6), 1541-1555.

Runyan, D. K., Dunne, M. P., & Zolotor, A. J. (2009). Introduction to the development of the ISPCAN child abuse screening tools. *Child Abuse Negl, 33*(11), 842-845.

Runyan, D. K., Dunne, M. P., Zolotor, A. J., Madrid, B., Jain, D., Gerbaka, B., . . . Youssef, R. M. (2009). The development and piloting of the ISPCAN Child Abuse Screening Tool-Parent version (ICAST-P). *Child Abuse Negl, 33*(11), 826-832.

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., . . . Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *J Clin Psychiatry, 71*(3), 313-326.

Sherbourne, C. D., & Stewart, A. L. (1991). The MOS social support survey. *Social Science and Medicine, 32*(6), 705-714.

Slep, A. M. S., Heyman, R. E., Snarr, J. D., & Foran, H. M. (2013). Practical tools for assessing child maltreatment in clinical practice and public health settings. In H. M. Foran, S. Beach, A. M. S. Slep, R. E. Heyman, & M. Wamboldt (Eds.), *Family problems and violence: Reliable assessment and the ICD-11* (pp. 159-184). New York: Springer.

Staehr, J. K. (1998). *The use of well-being measures in primary health care-the DepCare project*. Geneva: World Health Organization.

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., . . . Dees, J. E. M. E. G. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior, 14*(1), 13-29.

Straus, M. A., & Douglas, E. M. (2004). A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence and Victims, 19*(5), 507-520.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., . . . Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC Medical Research Methodology, 10*(1), 1.

Topp, C. W., Ostergaard, S. D., Sondergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: a systematic review of the literature. *Psychother Psychosom, 84*(3), 167-176.

Ward, C. L., Mikton, C., Cluver, L., Cooper, P., Gardner, F., Hutchings, J., . . . Wessels, I. M. (2014). *Parenting for Lifelong Health: From South Africa to other low-and middle-income countries. Early Childhood Matters: Responsive Parenting: A Strategy to Prevent Violence.*

Ward, C. L., Wessels, I. M., Lachman, J. M., Hutchings, J., Cluver, L. D., Kassanjee, R., . . . Gardner, F. (2020). Parenting for Lifelong Health for Young Children: a randomized controlled trial of a parenting program in South Africa to prevent harsh parenting and child conduct problems. *J Child Psychol Psychiatry, 61*(4), 503-512.

Weisz, J. R., Sigman, M., Weiss, B., & Mosk, J. (1993). Parent reports of behavioral and emotional problems among children in Kenya, Thailand, and the United States. *Child Dev, 64*(1), 98-109.

Williams, M. E., Foran, H. M., Hutchings, J., Frantz, I., Taut, D., Lachman, L., . . . Heinrichs, N. (Submitted). Exploring predictors of parent engagement in a parenting programme in Southeastern Europe. *Journal of Child and Family Studies*.

Wlodarczyk, O., Pawils, S., Metzner, F., Kriston, L., Klasen, F., & Ravens-Sieberer, U. (2017). Risk and protective factors for mental health problems in preschool-aged children: cross-sectional results of the BELLA preschool study. *Child Adolesc Psychiatry Ment Health, 11*, 12.

World Health Organization. (1998). *Well-Being Measures in Primary Health Care – the DepCare Project*. Copenhagen: WHO Regional Office for Europe.

Table 1: Participant sociodemographic characteristics by study country

Variable	North Macedonia		Moldova		Romania	
	n	*%*	*n*	*%*	*n*	*%*
Child						
Age (in years) – M and SD	5.7	1.8	6.3	2.1	5.5	2.1
Gender (girl)	26	52.0	26	60.5	25	53.2
Caregiver						
Age – M and SD	36.7	4.3	34.3	7.5	34.6	9.8
Relationship with child						
Biological mother	46	92.0	39	90.7	41	87.2
Biological father	3	6.0	0	0	0	0
Stepmother/Stepfather	0	0	1	2.3	0	0
Grandmother/Grandfather	0	0	1	2.3	4	8.5
Aunt/Uncle	0	0	0	0	1	2.1
Sister/Brother	0	0	0	0	1	2.1
Cousin	0	0	1	2.3	0	0
Foster parent	0	0	1	2.3	0	0
Other*	1	2.0	0	0	0	0
Relationship status						
Single/widowed/separated/ divorced	6	12.0	7	16.3	4	8.5
In a relationship	2	4.0	5	11.6	14	29.8
Married	42	84.0	31	72.1	29	61.7
Number of children living in the household	2.2	1.7	1.9	0.9	2.6	1.7
Education level (no university/college)	18	36.0	15	34.9	35	74.5
Literacy level (cannot/only read with difficulty)	6	12.0	4	9.3	22	46.8
Experienced at least one form of hunger in past 30 days (yes)	9	18.0	12	27.9	23	48.9

* Adoptive mother

Table 2: Results – comparisons of mean differences between pre-and post-assessment for primary and secondary outcomes (standardized estimates are presented), with country included in the models (N = 140)

	Mean (SD) at pre	Mean (SD) at post	Wald test estimate	Wald test p-value	Effect size (Cohen's d)
Primary outcomes					
CBCL externalizing – 2-5yr[a]	19.0 (8.2)	9.3 (7.2)	26.10	<.001	1.26
CBCL externalizing – 6-9yr[b]	14.9 (8.8)	6.6 (5.1)	18.58	<.001	1.15
CBCL aggressive behavior – 2-5yr[a]	15.8 (6.9)	7.7 (5.8)	28.54	<.001	1.27
CBCL aggressive behavior – 6-9yr[b]	10.8 (6.6)	4.9 (4.0)	13.83	.001	1.08
Secondary outcomes					
CBCL internalizing – 2-5yr[a]	12.6 (8.2)	5.6 (5.4)	19.95	<.001	1.01
CBCL internalizing – 6-9yr[b]	12.1 (8.4)	5.9 (6.4)	5.87	.015	0.83
DASS overall	31.0 (19.3)	22.1 (15.8)	11.38	.001	0.50
WHO wellbeing overall	53.1 (18.9)	59.4 (17.1)	13.09	.001	-0.35
Physical abuse – frequency	3.2 (4.0)	1.1 (1.9)	19.27	<.001	0.67
Emotional abuse – frequency	9.4 (7.3)	5.1 (4.3)	25.51	<.001	0.72
Neglect – frequency	0.6 (2.4)	0.1 (0.5)	7.24	.007	0.29
Overall harsh parenting - frequency	13.3 (10.9)	6.2 (5.5)	32.45	<.001	0.82
PS overall score	3.7 (0.5)	3.2 (0.6)	16.15	<.001	0.91
PS laxness	3.6 (1.0)	3.1 (1.0)	12.70	.001	0.50
PS over-reactivity	3.2 (1.0)	2.7 (0.9)	7.80	.005	0.53
PS verbosity	4.8 (1.0)	4.1 (1.1)	13.24	.001	0.67
PARYC overall score	97.0 (15.6)	106.8 (14.8)	20.08	<.001	-0.64
PARYC positive parenting	34.7 (6.0)	37.6 (5.6)	16.92	<.001	-0.50
PARYC monitoring	31.9 (6.2)	35.6 (5.1)	17.84	<.001	-0.65
PARYC proactive parenting	30.5 (7.3)	33.6 (6.3)	7.03	.008	-0.45
Other outcomes					
IPV – victimization[c]	5.9 (11.4)	3.1 (8.5)	12.54	<.001	0.28
IPV – perpetration[c]	4.6 (7.0)	2.2 (5.4)	13.34	<.001	0.38
KMSS – parental relationship quality[c]	16.7 (5.2)	17.9 (4.2)	24.13	<.001	-0.25
FAD – family functioning	1.9 (0.5)	1.8 (0.5)	10.47	.001	0.20
MOS emotional support	3.6 (1.1)	3.8 (1.0)	1.48	.225	-0.19

[a] The total sample size for the CBCL 2-5yr version was *n* = 64

[b] The total sample size for the CBCL 6-9yr version was *n* = 76

[c] The sample size of participants being in a relationship was *n* = 123

Note: all effect sizes were in the desirable direction.

Abbreviations: CBCL = Child Behavior Checklist; DASS = Depression, Anxiety, Stress Scale; PS = Parenting Scale; PARYC = Parenting of Young Children Scale; IPV = Intimate Partner Violence; KMSS = Kansas Marital Satisfaction Scale; FAD = Family Assessment Device; MOS = Medical Outcome Study.

Table 3: Correlations between implementation factors (in percentages) and primary, secondary and other outcomes at post-assessment (N = 140)

	Attendance	SMS	Phone call	Home activity
Primary outcomes				
CBCL externalizing – 2-5yr[a]	-.560*	.355*	-.244	.207
CBCL externalizing – 6-9yr[b]	.366*	.211	-.118	-.469*
CBCL aggressive behavior – 2-5yr[a]	-.573	.318*	-.265	.282
CBCL aggressive behavior – 6-9yr[b]	.300	.178	-.157	-.462*
Secondary outcomes				
CBCL internalizing – 2-5yr[a]	-.632*	.253	-.145	.237
CBCL internalizing – 6-9yr[b]	.262	-.039	-.023	-.251
DASS overall	.276	-.092	.051	-.447*
WHO wellbeing overall	-.272	-.079	-.107	.233
Physical abuse – frequency	.058	.059	.085	-.223
Emotional abuse – frequency	.316*	.183	-.090	-.330*
Neglect – frequency	.162	-.213	-.071	-.195
Overall harsh parenting - frequency	.157	-.009	.055	-.373*
PS overall score	-.106	-.186	-.130	-.054
PS laxness	-.106	-.153	-.048	-.138
PS over-reactivity	-.011	-.152	-.140	.116
PS verbosity	-.198	.044	-.199	-.033
PARYC overall score	-.165	-.120	-.061	.447*
PARYC positive parenting	-.293	-.097	-.075	.397
PARYC monitoring	.003	-.218	.009	.368
PARYC proactive parenting	-.095	-.022	-.080	.385
Other outcomes				
IPV – victimization[c]	-.013	-.093	-.042	-.017
IPV – perpetration[c]	-.013	-.004	.062	-.245
KMSS – parental relationship quality[c]	-.003	-.066	.025	-.026
FAD – family functioning	.067	-.069	.042	-.074
MOS emotional support	-.228	-.058	-.221	.270

[a] The total sample size for the CBCL 2-5yr version was *n* = 64

[b] The total sample size for the CBCL 6-9yr version was *n* = 76

[c] The sample size of participants being in a relationship was *n* = 123

Significant correlations with *p* <.01 are highlighted with an asterisk *.

The same results were found when examining the correlations and follow-up analysis in the 94 participants returning for post-assessment.

Abbreviations: CBCL = Child Behavior Checklist; DASS = Depression, Anxiety, Stress Scale; PS = Parenting Scale; PARYC = Parenting of Young Children Scale; IPV = Intimate

Partner Violence; KMSS = Kansas Marital Satisfaction Scale; FAD = Family Assessment
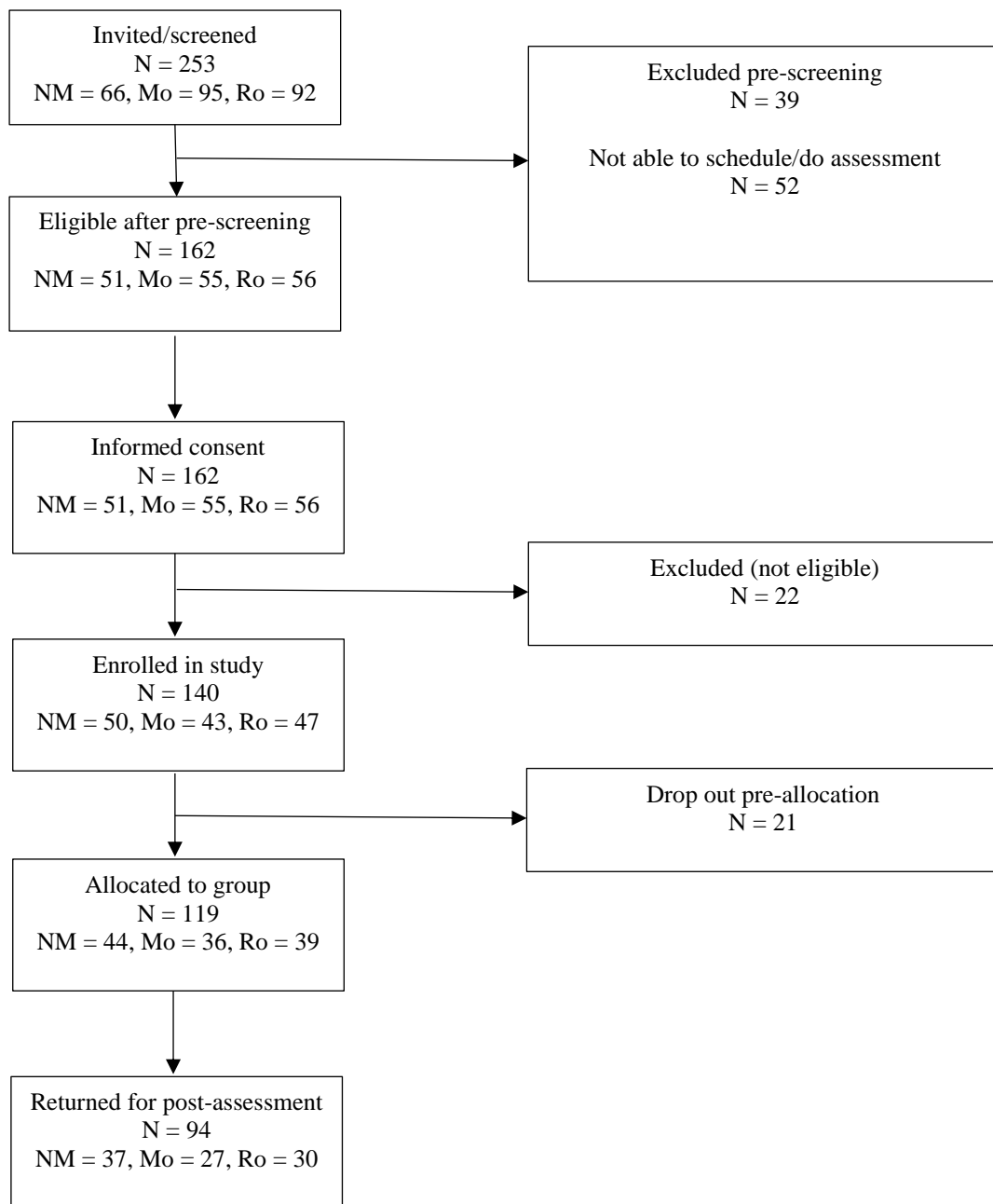
Device; MOS = Medical Outcome Study.

Figure 1. Participant flow diagram

Abbreviations: NM = North Macedonia, Mo = Moldova, Ro = Romania

Supplemental Material

As can be seen from the Table, all differences except for the Parenting of Young Children (PARYC) scales were statistically significant in the direction suggestive of improvements at post-assessment. Differences found between listwise analyses and full sample analyses with regard to the PARYC scales may be due to the larger sample size and adjustment for non-normality.

Table S1: Pre-post comparisons using ANCOVA, adjusted for study country

	Mean (*SD*) at pre (*N* = 94)	Mean (*SD*) at post (*N* = 94)	*p*-value
Primary outcomes			
CBCL externalizing – 2-5yr[a]	19.0 (8.3)	9.2 (7.1)	.002
CBCL externalizing – 6-9yr[b]	14.9 (8.8)	6.6 (5.3)	<.001
CBCL aggressive behavior – 2-5yr[a]	15.8 (7.0)	7.6 (5.8)	.002
CBCL aggressive behavior – 6-9yr[b]	10.7 (6.7)	4.9 (4.1)	.001
Secondary outcomes			
CBCL internalizing – 2-5yr[a]	12.6 (8.3)	5.5 (5.2)	.002
CBCL internalizing – 6-9yr[b]	12.1 (8.4)	6.1 (6.4)	<.001
DASS overall	31.2 (17.9)	22.1 (15.9)	.003
WHO wellbeing overall	50.6 (18.5)	58.7 (17.1)	<.001
Physical abuse – frequency	3.2 (4.0)	1.1 (1.9)	<.001
Emotional abuse – frequency	9.4 (7.4)	4.9 (4.2)	<.001
Neglect – frequency	0.6 (2.4)	0.1 (0.5)	.004
Overall harsh parenting - frequency	13.1 (11.2)	6.1 (5.6)	<.001
PS overall score	3.7 (0.6)	3.2 (0.7)	.006
PS laxness	3.6 (1.0)	3.1 (1.0)	<.001
PS over-reactivity	3.2 (1.0)	2.7 (1.0)	.016
PS verbosity	4.8 (1.0)	4.1 (1.1)	.004
PARYC overall score	96.9 (15.1)	106.9 (14.9)	.138
PARYC positive parenting	34.6 (5.7)	37.6 (5.6)	.048

PARYC monitoring	32.1 (5.8)	35.6 (5.1)	.266
PARYC proactive parenting	30.1 (7.1)	33.6 (6.4)	.162
Other outcomes			
IPV – victimization[c]	5.83 (9.9)	2.96 (8.0)	<.001
IPV – perpetration[c]	4.58 (6.5)	2.15 (5.4)	<.001
KMSS – parental relationship quality[c]	12.25 (3.0)	17.98 (4.2)	.003
FAD – family functioning	1.94 (0.4)	1.80 (0.5)	<.001
MOS emotional support	3.55 (1.1)	3.81 (1.0)	<.001

[a] The total sample size for the CBCL 2-5yr version was $n = 40$

[b] The total sample size for the CBCL 6-9yr version was $n = 49$

[c] The sample size of participants being in a relationship was $n = 80$

Abbreviations: CBCL = Child Behavior Checklist; DASS = Depression, Anxiety, Stress Scale; PS = Parenting Scale; PARYC = Parenting of Young Children Scale; IPV = Intimate Partner Violence; KMSS = Kansas Marital Satisfaction Scale; FAD = Family Assessment Device; MOS = Medical Outcome Study.

Table S2: Lack of household assets at pre-assessment by study country

Household assets not available at pre-assessment	North Macedonia (*n* = 50)	Moldova (*n* = 43)	Romania (*n* = 47)
	No, not available – count (%)		
Water	0	1 (2.3)	6 (12.8)
Electricity	0	0	2 (4.3)
Radio	12 (24.0)	17 (39.5)	18 (8.3)
Tv	2 (4.0)	1 (2.3)	7 (14.9)
Mobile phone	0	0	5 (10.6)
Telephone	24 (48.0)	15 (34.9)	38 (80.9)
Fridge	5 (10.0)	1 (2.3)	6 (12.8)
Dishwasher	17 (34.0)	42 (97.7)	40 (85.1)
Washing machine	4 (8.0)	2 (4.7)	10 (21.3)
Computer	9 (18.0)	11 (25.6)	28 (59.6)
Internet	9 (18.0)	5 (11.6)	19 (40.4)
Watch	0	5 (11.6)	13 (27.7)
Bicycle	16 (32.0)	11 (25.6)	21 (44.7)
Motorbike or scooter	46 (92.0)	40 (93.0)	45 (95.7)
Car	15 (30.0)	17 (39.5)	25 (53.2)
Animal cart	50 (100)	42 (97.7)	38 (80.9)
Boat	48 (96.0)	43 (100)	46 (97.9)
Tractor	50 (100)	43 (100)	46 (97.9)
Books	0	1 (2.3)	15 (31.9)