Supplementary file

Measures

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
1.1 What is the feasibility of delivery of the adapted version of PLH for Young Children?	Reach: Enrolment rate	Number of families who attend at least one session of the programme divided by the number of families recruited into the programme	Facilitator report form, 1
	Reach: Participation rate	Mean attendance rate for programme sessions based on those families who enrolled in the programme (i.e., parents who attended at least one session). Percentage of families who enrolled in the programme who attended 50% (e.g., 6 sessions) and 75% (e.g., 9 sessions) or more	Facilitator report form, 1
	Implementation Fidelity	Overall fidelity score: Percentage of number of session activities delivered by facilitators per session (by facilitator group, implementing agency, and participating country site)	Implementation monitors, fidelity check-list, video recording, 1
	Implementation: Dosage	Average number of hours delivered by facilitators (time for pre-programme consultation plus session plus phone consultations per participant)	Facilitator, implementation monitors, report form, 1
		Total number of points of contact by facilitators	Facilitator report form, 1
	Implementation: Quality	PLH-Facilitator Assessment Tool (PLH-FAT): 7 standard behaviour categories are grouped into 2 scales: core activities (quality of delivery during home activity review, illustrated story discussions, practicing skills) and process skills (modelling skills, collaborative facilitation approach, encouragement of participation, leadership skills)	Facilitator questionnaire, 1
	Implementation: Acceptability and appropriateness of programme materials, delivery, and key programme components	Interviews*	Qualitative interviews with parents and facilitators, 2
	Implementation: Participant reported observed change in parenting practices and child behaviour at home	Interviews with intervention participants and focus groups with the facilitators in order to explore programme acceptability*	Qualitative interviews with parents and facilitators, 2
1.2 Are the evaluation methods	Reach: Recruitment rate	Number of families who were eligible for inclusion and provided consent to	Tablets, 4

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
appropriate and feasible?		participate in the programme divided by the number of target population who were exposed to recruitment activities	
	Informed Consent	Percentage of parents giving informed consent from those who are eligible	Data assessor report, 4
	Eligibility	Percentage of families being eligible (meeting inclusion criteria) from those who are screened (completed screening on tablet). Acceptable: 70%	Data assessor report, 4
	Measurement reliability	Internal consistency Acceptable: Cronbach's alpha ≥ 0.70	Questionnaires, 3
	Study retention	Percentage of parents who complete pre- and post-assessment from all parents being eligible. Acceptable: 80%	Data assessor report, 2
1.3 What are the procedures that need to be adapted or changed for the later study phases?	Implementation: Existing barriers to participation during sessions and engagement in home practice and other activities	Interviews*	Qualitative interviews and focus groups with parents and facilitators, 2
	Challenges in implementing the programme	Focus group discussions exploring challenges in implementing the programme on a process (e.g., using a collaborative approach and/or explaining concepts such as child led play), and logistical level (e.g., recruitment, session length, location, meals)	Qualitative focus groups with facilitators, 2
2. Among families participating in the programme, are there pre-post improvements on child and parental mental health and behaviour? **Primary Outcome**	Externalizing behaviour problems in children	The parent-report versions for children aged 1½ - 5 and 6 - 18 of the Child Behavior Checklist (CBCL)[34] will be employed. It is a well validated instrument[34] that has been used across different prevention and treatment studies and countries. The externalizing subscale raw score ranges from 0 to 48 (CBCL ½ - 5 version; 24 items) and 0 to 70 (CBCL 6 - 18 version; 35 items) with higher scores indicating more problems.	Parent report, questionnaire, 3
		The Mini International Neuropsychiatric Interview for Children and Adolescents – Parent Version (MINI-KID-P)[35, 36] is a structured interview to evaluate the presence of current psychiatric disorders (based on DSM-5 with corresponding ICD-10-CM codes including child equivalents; using a binary *yes/no* format). The parent-rated version will be employed. The interview is organized in disorder-specific modules and use screening questions for each disorder. The reliability and validity of the screening tool is adequate[37]. It assesses whether or not the criteria for a) Conduct disorder (F91.1, F91.2, F91.8) or b) Oppositional defiant disorder (F91.3) are met (yes/no). The results of the two disorders will be combined to one binary total score with 0 (*no externalizing disorder*) and 1 (*current externalizing disorder*;	Clinician-rated parent report, structured interview, 3

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
		ODD or CD).	
Secondary outcomes	Emotional problems in children	Parent-report on the internalizing subscale of the CBCL (for details see above) will be employed. The raw score ranges from 0 to 62 (CBCL 1 ½ - 5 version; 31 items) and 0 to 64 (CBCL 6 - 18 version; 32 items) with higher scores indicating more problems.	Parent-report, questionnaire, 3
		Using the MINI-KID-P (for details see above), we will assess whether or not the criteria for Separation Anxiety Disorder (F93.0), Social Anxiety Disorder (F40.1) Specific Phobia (F40.2, F93.1), Generalized Anxiety (F41.1, F93.80), or Obsessive-Compulsive Disorder (F42.2) are currently met (*yes/no*). The results will be combined to one binary total score 0 (*no anxiety disorder*) and 1 (*current anxiety disorder*; criteria for at least one of the anxiety disorders met).	Clinician-rated parent report, structured interview, 3
	Psychological distress in parents	The Depression, Anxiety, Stress Scales (DASS)[38] will assess parental psychological distress. It is a 21-item screening tool to measure depression, anxiety, and stress in adults. Parents report on the frequency of symptoms in the previous week using a Likert scale from 0 to 3 (*never* to *always*). The DASS is a widely used measure across parenting studies with good internal consistency and concurrent validity[39]. Total DASS scores range from 0 to 63 with subscales from 0 to 21.	Parent-report, questionnaire, 3
	Parental well-being	The WHO-5 Well-Being Scale (WHO-5)[40] will measure parental psychological well-being. This 5-item screening tool was derived using psychometric analyses from the longer 28-item WHO Well-Being Scale and is widely used in HIC and LMIC. Parents indicate the frequency that they experienced well-being in the past month (e.g., "My daily life has been filled with things that interest me") based on a 6-point Likert scale from 0 to 5 (*at no time* to *all of the time*). Items are added up with scores ranging from 0 to 25. A recent review has demonstrated change sensitivity and good sensitivity and specificity as a screening tool for depression[41]. In order to monitor possible changes in well-being, the percentage score will be used with a score of 0 representing worst possible, a score of 100 representing best possible quality of life.	Parent-report, questionnaire, 3
	Child maltreatment	This construct will be measured using parent report of the ISPCAN Child Abuse Screening Tool-Trial scale[42, 43] an adaptation of a multi-national and consensus-based survey instrument measuring the incidence and prevalence of child abuse and neglect, as well an additional measure of violent acts based on previous sensitivity and specificity analyses. It was	Parent-report, questionnaire, 3

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
		validated in 6 LMIC and 7 languages and measures four types of abuse: physical, emotional and sexual abuse, as well as neglect (sexual abuse is not assessed in this study). The response code was adapted to a scale from 0 to more than 8 times to assess the frequency of a certain behaviour in the past month. This study will assess incidence of child maltreatment by creating dichotomous variables for physical abuse (4 items), verbal abuse (7 items), and neglect (3 items), as well as an overall indication of previous child abuse (0 = *no abuse*; 1 = *previous abuse*). In addition to the original items of the scale, four additional items from the Family Maltreatment Measure are included. Two items will assess discipline strategies related to abuse and two items were added to assess verbal abuse based on past studies evidencing their specificity and sensitivity[44]. We will also assess frequency of overall abuse by summing all the three subscales as well as for each individual subscale.	
	Dysfunctional parenting	The Parenting Scale (PS)[45] will measure dysfunctional discipline practices in parents. Three subscales may be derived (Laxness, Overreactivity, and Verbosity). Each item is rated on a 7-point Likert Scale in which parents are presented with a situation and then are asked to choose between two alternative responses (i.e., situation: "When I say my child can't do something"; response score = 1 *I stick to what I said;* or response score = 7 *I let my child do it anyway*). The factor structure and validity have been extensively researched with acceptable reliability[46]. For computation of the subscale scores as well as the total score, the responses on the items are averaged. Higher scores indicate more dysfunctional parenting.	Parent-report, questionnaire, 3
	Positive parenting and effective discipline	Positive parenting behaviour will be assessed using parent-report of the Parenting of Young Children Scale (PARYC, 21 items)[47]. The PARYC measures the frequency of parent behaviour over the previous month. Items are summed to create a total score as well as scores for the subscales Positive parenting (7 items, e.g., "How often do you play with your child?"), Setting limits (7 items, e.g., "How often do you stick to your rules and not change your mind?") and Proactive parenting (7 items, e.g., "How often do you explain what you want your child to do in clear and simple ways?"). The internal consistency of subscales was good and the validity has been widely researched[47].	Parent-report, questionnaire, 3
Other Pre-specified Outcomes	Intimate partner violence	Intimate partner violence based on adult self-report of perpetration and victimization of intimate partner physical and psychological aggression will	Parent-report, questionnaire, 3

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
		be assessed with a screening instrument, the family maltreatment measure[48] and an adaption of the revised Conflict Tactics Scale (CTS2S)[49]. Assessments will measure the frequency of physical assault, psychological aggression, and physical injury. Answers will be coded on a 5-point Likert scale of 0 to 4, with an additional response for incidences that happened but not in the past month. For the current study, a 9-point Likert scale of 0 to 8 will be used (from *never happened* to *more than 8 times in the past month*), with an additional response for incidences that happened but not in the past month. This measure indicates an overall indication of intimate partner violence on a level of severity (sum of items) and prevalence (dichotomous variable indicating experience of conflict or not) as well as for each subscale. Only severity will be examined.	
	Family functioning	The 12-item short form of the FAD (Family Assessment Device; general functioning)[50] will assess family functioning. The FAD has shown to be a valid instrument for assessing family outcomes in clinical trials and has good internal consistency[50]. For computation of the total score, the responses on each item (ranging from 1 to 4) will be averaged. Thus, the total score will range from 1 to 4 with higher scores indicating more problems in family functioning.	Parent-report, questionnaire, 3
	Parental relationship quality	The 3-item Kansas Marital Satisfaction Scale (KMSS)[51] assesses relationship satisfaction among intimate partners. This scale has been widely used and correlates highly with other measures of relationship satisfaction (e.g., Dyadic Adjustment Scale, Quality of Marriage Index[52]). Items will be rated on a scale from 1 (*extremely dissatisfied*) to 7 (*extremely satisfied*).	Parent-report, questionnaire, 3
	Social community support	Perceived social support will be measured using the emotional support subscale of the Medical Outcome Study Social Support Survey (MOS-SSS, 8 items)[53]. In validation studies this subscale has shown excellent internal consistency[53]. Parents will report on the frequency of how often they receive emotional support (e.g., "Someone you can count on to listen to when you need to talk") on a Likert-like scale of 1 to 5 (from *none* of the time to *all of the time*). An emotional support subscale score will be calculated by averaging the scores for each item.	Parent-report, questionnaire, 3
3. Are the measures and indicators (including potential moderators and RE-AIM items) for the evaluation of	Reach: Recruitment	Number of recruitment strategies employed.	Report form, 4

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
phases 2 and 3 feasible?			
	Implementation: Programme adherence	Dropout rates: percentage of participants who fail to attend at least three consecutive sessions and do not attend any sessions at a later stage.	Implementation monitors, attendance register, 1
		Completion rates: the number of enrolled participants who attend a cut-off threshold of at least 66% of the programme[54].	Implementation monitors, attendance register, 1
	Reach: Representativeness	Comparisons between the study sample and those that were eligible and declined participation (e.g., number of characteristics compared relative to number of differences found on income, age, education etc.).	Parent-report, questionnaires, 1
	Reach and Implementation: Potential barriers to programme participation and engagement	The Obstacles to Engagement Scale (OES)[55] is a 14-item measure including four subscales: Family obstacles (4 items); Relevance of parenting programmes (4 items); Suitability of group-based programmes (4 items); and Barriers due to time commitments (2 items). Internal consistency was above .7 for the first three subscales. Some of the subscales have shown to be predictive for poor parental enrolment and attendance in a parent training[56]. Participants will rate each item on a 4-point sliding Likert scale ranging from *definitely yes* to *definitely no*. Scores for each subscale will be created as well as an overall score by summing totals.	Parent-report, questionnaire, 4
	Effectiveness and Implementation: Participant satisfaction	Parental satisfaction with the PLH programme will be assessed using the Parent Satisfaction Scale[57]. The 40-items measure has four subscales (i.e., whether the programme fulfilled their expectations, acceptability of delivery and teaching methods, acceptability of theoretical parenting techniques, and evaluation of programme facilitators). This scale has been used in other parenting programme studies [e.g., 57] including PLH trials in other countries and will allow comparison of results to those studies.	Parent-report, questionnaire, 2
	Household poverty	The Hunger Scale Questionnaire[58] will ask parents to respond to questions on food shortage and hunger in the household. Parents will respond positively or negatively regarding the occurrence of hunger in the household, whether it occurred during the past 30 days, and if so, whether it occurred more than five times in the past 30 days. The scale produces scores for single occurrence and intensity of hunger. This scale has been used in PLH trials in other countries and will allow comparison of results to those studies.	Parent-report, questionnaire, 3
		The Household Assets modified from UNICEF (2005) Multiple Indicators Cluster Survey (MICS) Household Survey will also assess household poverty with 19 items.	Parent-report, questionnaire, 4

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
	Parent history of child maltreatment	Parental history of maltreatment during childhood in the families of origin will be measured using an adapted version of the International Society for the Prevention of Child Abuse and Neglect Child Abuse Screening Tools Retrospective version (ICAST-R, 5 items)[59]. This scale has been tested in different countries and languages and utilizes parent self-report of experiences during their own childhood (under 18 years old). In this study, incidence of past history of child maltreatment will be scored as dichotomous variables for physical and verbal abuse, as well as an overall score (0 = *no abuse*; 1 = *previous abuse*).	Parent report, questionnaire, 4
	Parental general health	Caregiver general health will be assessed using three items from the Medical Outcomes Study (MOS) Short Form-12 Health Survey (SF-12)[60]. This scale is an adapted version of the MOS SF-34 Health Survey, has demonstrated excellent internal consistency and test-retest reliability and examines physical and mental health[60]. Items include difficulty in normal daily activities, such as cleaning the home, going to work, or carrying a child. Response options are based on a 3-point Likert-like scale (1 = *yes, limited a lot*; 3 = *no, not limited at all*). The third item requires respondents to assess their overall health on a 5-point Likert scale (1 = *excellent*; 5 = *poor*). Three additional items ask respondents whether they or their child have a (physical or mental) disability. If they respond *yes*, they are asked to specify which type of disability.	Parent-report, questionnaire, 3
	Parental alcohol use	The 10-item Alcohol Use Disorders Identification Test (AUDIT)[61] is a screening instrument to detect harmful alcohol use and related problems. It has demonstrated good internal consistency, retest reliability as well as sensitivity and specificity as a screening instrument[62]. Total scores of eight or more are recommended as indicators of hazardous and harmful alcohol use, as well as possible alcohol dependence.	Parent-report, questionnaire, 3
	Neighbourhood safety, social involvement and public services	The Neighbourhood Questionnaire[63] consists of 16 items. It was selected because it allows comparison with the results from another large-scale project (FAST TRACK)[64]. Eleven items are used to assess Neighbourhood safety (5-item subscale), Neighbourhood social involvement (4-item subscale) and Quality of public services (such as police, schools, transportation, 2-item subscale). Responses include different answer formats.	Parent-report, questionnaire, 4
	Attention deficit hyperactivity disorder	The MINI-KID-P (for details see above) will be employed to assess whether or not the criteria for ADHD (F90.0, F90.1, F90.2) are currently met (yes/no). The results will be combined to one binary total score 0 (*no ADHD*) and 1	Clinician-rated parent report, structured interview, 3

Research question	Construct	Indicator/Questionnaire	Informant, data source, assessment point
		(*current ADHD*; criteria met).	

Notes. *Participants are purposively selected from the intervention group with the inclusion criterion of attending at least one intervention session. Selection is based on those with high attendance (> 75%), those with low attendance (< 25%), and those who do not enrol. Assessment point: 1: on-going assessment during the programme implementation, 2: assessed at post intervention, 3: assessed at pre and post intervention, 4: assessed at pre-assessment only.

References

34 Achenbach TM, Rescorla LA. Manual for ASEBA School Age Forms & Profiles. Burlington, VT: University of Vermont Research Center for Children, Youth and Families 2001.

35 Lecrubier Y, Sheehan DV, Weiller E, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry*1997;12(5):224-231 doi:10.1016/S0924-9338(97)83296-8.

36 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(20):22–23.

37 Duncan L, Georgiades K, Wang L, et al. Psychometric evaluation of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *Psychol Assess*2018;30(7):916–928 doi:10.1037/pas0000541 [published Online First: July 2018].

38 Henry JD, Crawford JR. 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 of Clin Psychol*2005;44(2):227–39 doi:10.1348/014466505X29657 [published Online First: June 2005].

39 Antony MM, Bieling PJ, Cox BJ, et al. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychol Assess*1998;10(2):176-181.

40 WHO. Well-being measures in primary health care - the DepCare project health for all. Copenhagen: WHO Regional Office for Europe 1998.

41 Topp CW, Østergaard SD, Søndergaard S, et al. The WHO-5 Well-Being Index: a systematic review of the literature. *Psychother Psychosom*2015;84(3):167–176.

42 Meinck F, Boyes ME, Cluver L, et al. Development and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-TRIAL) among South African adolescents and their primary Caregivers. *Child Abuse Negl*2018;82:45–58 doi:10.1016/j.chiabu.2018.05.022 [published Online First: 31 May 2018].

43 Runyan DK, Dunne MP, Zolotor AJ, et al. The development and piloting of the ISPCAN Child Abuse Screening Tool—Parent version (ICAST-P). *Child Abuse Negl*2009;33(11):826-832 doi:10.1016/j.chiabu.2009.09.006.

44 Slep AMS, Heyman RE, Snarr JD, et al. Practical tools for assessing child maltreatment in clinical practice and public health settings. In: Foran HM, Beach SRH, Slep, AMS, et al. (Eds.). Family violence and family problems: Reliable assessment and the ICD-11. New York: Springer Publishing 2013.

45 Arnold DS, O'leary SG, Wolff LS, et al. The Parenting Scale: a measure of dysfunctional parenting in discipline situations. *Psychol Assess*1993;5(2):137-144 doi:1037/1040-3590.5.2.137.

46 Prinzie P, Onghena P, Hellinckx W. Reexamining the Parenting Scale: Reliability, factor structure, and concurrent validity of a scale for

assessing the discipline practices of mothers and fathers of elementary-school-aged children. *Eur J of Psychol Assess2007;*23(1):24-31 doi:10.1027/1015-5759.23.1.24 [published Online First: 17 January 2007].

47 McEachern A. Parenting Young Children (PARYC): Validation of a self-report parenting measure. *J Child Fam Stud*2011;21(3):498-511 doi:10.1007/s10826-011-9503-y [published Online First: 6 August 2012].

48 Heyman RE, Snarr JD, Slep AMS, et al. Practical tools for assessing partner maltreatment in clinical practice and public health setting. In: Foran HM, Beach SRH, Slep AMS, et al. (Eds.), Family violence and family problems: Reliable assessment and the ICD-11. New York: Springer Publishing 2013.

49 Straus MA, Douglas EM. A Short Form of the Revised Conflict Tactics Scales, and Typologies for Severity and Mutuality. *Violence Vict*2004;19(5):507-520.

50 Byles J, Byrne C, Bolye MH, et al. Ontario Child Health Study: Reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. *Fam Proc*1988;27(1):97-104 doi:10.1111/j.1545-5300.1988.00097.x.

51 Schumm WR, Nichols C, Schectman K, et al. Characteristics of responses to the Kansas Marital Satisfaction Scale by a sample of 84 married mothers. *Psychol Rep*1983;53:567–572.

52 Schumm WR, Paff-Bergen LA, Hatch RC, et al. Concurrent and discriminant validity of the Kansas Marital Satisfaction Scale. *JMF*1986;48(2):381-387.

53 Sherbourne CD, Stewart AL. The MOS social support survey. *Soc Sci Med*1991;32(6),705-714.

54 Webster-Stratton C, Herman KC. Disseminating Incredible Years Series early-intervention programs: Integrating and sustaining services between school and home. *PITS*2010;47(1):36-54 doi:10.1002/Pits.20450 [published Online First: 17 November 2009].

55 Wilson SM, Wildman BG, Ciesla J, et al. Factor structure of the Obstacles to Engagement Scale: Problems, solutions, and hypotheses. *J Child Fam Stud*2014;24(4):891-898 doi: 10.1007/s10826-014-9899-2 [published Online First: April 2014].

56 Dumas JE, Nissley-Tsiopinis J, Moreland AD. From intent to enrollment, attendance, and participation in preventive parenting groups. *J Child Fam Stud*2007;16(1):1–26 doi10.1007/s10826-006-9042-0 [published Online First: 27 April 2006].

57 Martinez CR, Eddy JM. Effects of culturally adapted parent management training on Latino youth behavioral health outcomes. *J Consult Clin Psychol*2005;73(5):841–851 doi:10.1037/0022-006X.73.5.841 [published Online First: 01 October 2005].

58 Labadarios D, Maunder E, Steyn N, et al. National food consumption survey in children aged 1-9 years: South Africa 1999. *Forum Nutr*2003;56:106–109.

59 Dunne MP, Zolotor AJ, Runyan DK, et al. ISPCAN Child Abuse Screening Tools Retrospective version (ICAST-R): Delphi study and field testing in seven countries. *Child Abuse Negl*2009;33(11):815–825 doi:10.1016/j.chiabu.2009.09.005 [published Online First: 22 October 2009].

60 Ware J, Kosinski MJr, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*1996;34(3):220–233.

61 Babor TF, Higgins-Biddle JC, Saunders JB, et al. The Alcohol Use Disorders Identification Test. World Health Organization; 2002.

62 Daeppen JB, Yersin B, Landry U, et al. Reliability and validity of the Alcohol Use Disorders Identification Test (AUDIT) imbedded within a general health risk screening questionnaire: results of a survey in 332 primary care patients. *Alcohol Clin Exp Res*2000;24(5):659–665.

63 Conduct Problems Prevention Research Group (CPPRG). Neighborhood questionnaire. Fast Track Project 1991. www.fasttrackproject.org/techrept/n/nhq/index.php (accessed 10 April 2017).

64 Pinderhughes EE, Nix R, Foster EM, et al. Parenting in context: Impact of neighborhood poverty, residential stability, public services, social networks, and danger on parental behaviors. *J Marriage Fam*2001;63(4):941–95.