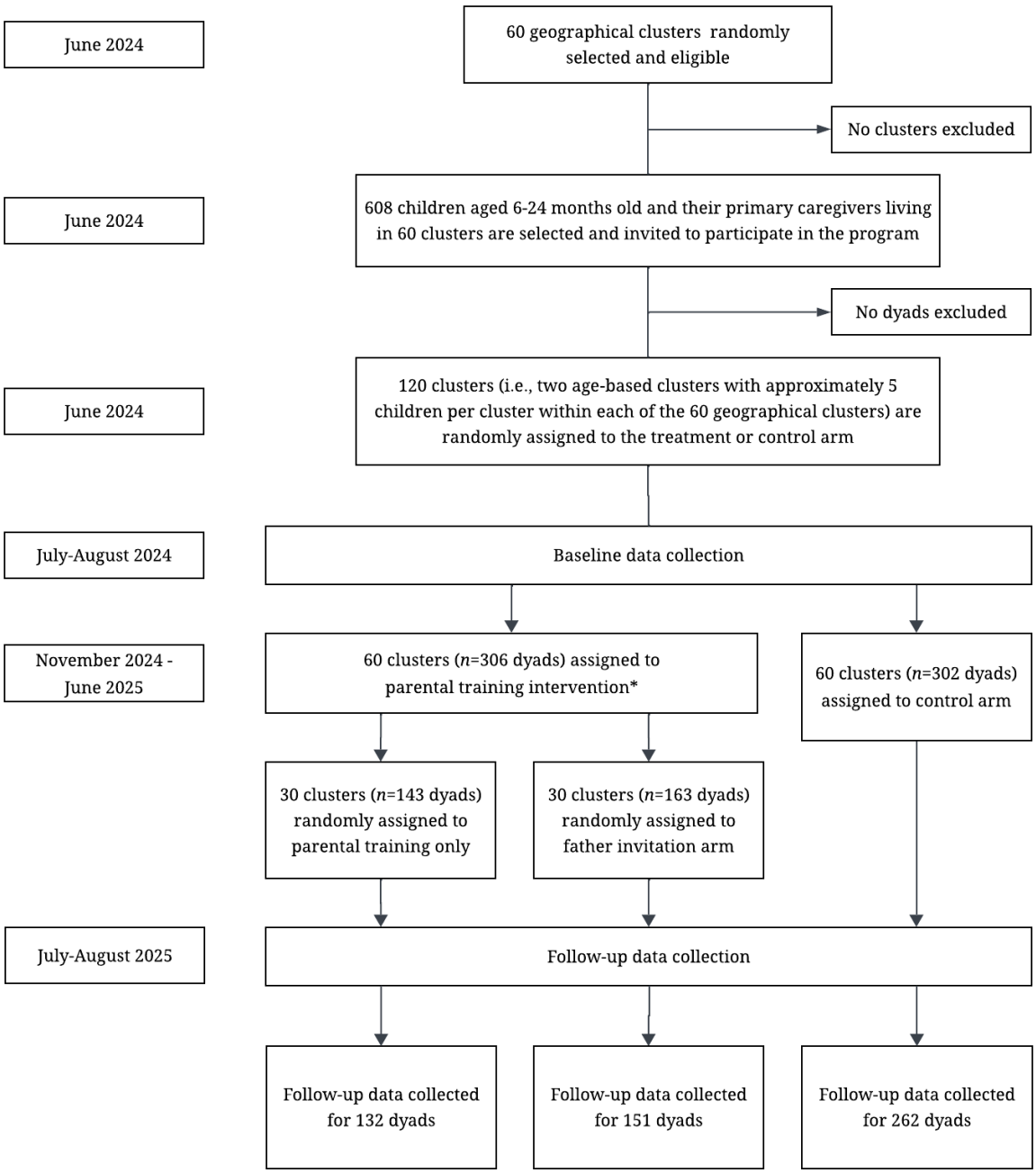
Figure 1: Trial design and participant flow chart



Note: Half of the treatment arm (i.e., 30 clusters in six of the 12 counties, with the six poorer counties being selected) is assigned to receive CCT incentive payments for eight weeks in order to encourage participation in the intervention.

Table 1: Baseline balance table

	Overall	Control (1)	Parenting training only (2)	Parenting training + father involvement (3)	(1) vs. (2) p-value	(1) vs. (3) p-value	(2) vs. (3) p-value
Panel A: Child characteristics							
Child sex (0=female;1=male)	0.528 (0.020)	0.517 (0.033)	0.538 (0.050)	0.540 (0.037)	0.738	0.634	0.993
Child's age (in months)	15.645 (0.242)	15.649 (0.561)	15.902 (0.819)	15.411 (0.804)	0.852	0.838	0.672
Child belongs to ethnic minority (0=no; 1=yes)	0.059 (0.016)	0.054 (0.019)	0.057 (0.025)	0.068 (0.030)	0.986	0.672	0.760
Child was born preterm (0=no; 1=yes)	0.075 (0.011)	0.075 (0.016)	0.071 (0.023)	0.080 (0.023)	0.877	0.862	0.762
Panel B: Household characteristics							
Age of primary caregiver (in years)	37.724 (0.544)	37.268 (0.657)	38.072 (1.179)	38.235 (1.142)	0.549	0.464	0.921
Educational attainment of primary caregiver (in years)	11.296 (0.189	11.311 (0.277)	11.230 (0.305)	11.327 (0.429)	0.840	0.970	0.865
Primary caregiver has adequate health (0=no; 1=yes)	0.913 (0.013)	0.910 (0.020)	0.950 (0.019)	0.889 (0.023)	0.143	0.504	0.040
Household asset factor z-score	0.000 (0.042)	0.008 (0.066)	-0.068 (0.076)	0.045 (0.069)	0.422	0.702	0.278
Panel C: Child development outcomes							
Bayley-III composite cognition scores	101.474 (0.595)	101.113 (0.818)	101.818 (0.988)	101.840 (1.484)	0.611	0.646	0.992
Bayley-III composite language scores	99.191 (0.841)	99.140 (1.109)	100.615 (1.555)	98.037 (1.752)	0.414	0.644	0.263

Bayley-III composite motor scores	104.059 (0.576)	103.990 (0.848)	103.161 (1.179)	104.975 (1.485)	0.640	0.539	0.340
BITSEA competence total score	13.881 (0.134)	13.820 (0.249)	13.979 (0.251)	13.907 (0.313)	0.673	0.839	0.854
BITSEA problem total score	12.015 (0.262)	11.701 (0.363)	13.014 (0.510)	11.716 (0.427)	0.032	0.985	0.070
No. of observations	608	302	143	163			

Notes: Standard errors presented in parentheses. P-values for difference in means across control and the two intervention arms are adjusted for clustering at the township level and stratification at the county level.

Table 2: Description of the outcome measures

Primary outcomes	Measurements	Description
Child development	Bayley-III	A trained enumerator assesses early childhood development - cognitive, (expressive and receptive) language and (fine and gross) motor skills - through behavioural observation at the child's home. Validated and widely used in China.[51–53]
	BITSEA	42 items assessing children's social-emotional development. Validated in China.[54]
Secondary outcomes	**Measurements**	**Description**
Child health	EQ-TIPS	Completed by the primary caregiver, assessing 6 dimensions of HRQoL: movement, play, pain, social interaction, communication and eating. Validated in China.[55]
Child screen time	Self-made items	7 items assessing the amount of time children spent on electronic devices and video programmes, adapted from Zhao *et al.*[56]
Stimulating caregiving practices	FCI	21 items assessing caregiver investment in stimulating parenting practices. Commonly used in China.[13]
Caregiver beliefs	KPCS	15 items assessing perceived parental self-efficacy of parents of infants aged 0-12 months. Validated Chinese version.[57]
	PSOC	17 items assessing parenting efficacy and satisfaction, divided into an 8-item efficacy subscale and a 9-item satisfaction subscale. Validated in China.[58]
	PACOTIS	23 items assessing parental perceptions and behavioural tendencies towards their infants
	CGSS	27 items on five scales (i.e., Toys and Activities Stereotyped for Boys, Toys and Activities Stereotyped for Girls, Education for Marriage and Family, Disapproval of Other Gender Characteristics and Helping at Home) assessing caregivers' attitudes about child

		gender socialization, as adapted by Bai *et al*.[59]
Caregiver mental health symptoms	DASS-21	21 items assessing depression, anxiety and stress symptoms. Widely used in China.[42,60]
	PSS	18 items assessing parental stress. Validated Chinese version.[61]
Caregivers' perceived social support	MSPSS	12 items assessing perceived social support from family, friends and significant others. Validated in China.[62]

Bayley- III, Bayley Scales of Infant and Toddler Development- III; BITSEA, Brief Infant- Toddler Social and Emotional Assessment; EQ-TIPS, EuroQol Toddler and Infant Populations**;** FCI, Family Care Indicators; KPCS, Karitane Parenting Confidence Scale; PSOC, Parenting Sense of Competence Scale; PACOTIS, Parenting Cognitions and Conduct Toward the Infant Scale; CGSS, Child Gender Socialization Scale; DASS-21, Depression Anxiety Stress Scale; PSS, Parental Stress Scale; MSPSS, Multidimensional Scale of Perceived Social Support.

Table 3: Adverse events

Category	Anticipated / Unanticipated	Description of adverse event	Number of participants affected (n)
Serious (life-threatening)	Unanticipated	Severe accidental injury or medical emergency during parenting sessions.	TBD
Other (non-life-threatening)	Anticipated	Minor injuries during play activities (e.g., falls or bruises), temporary discomfort, fatigue or distress among children during sessions or assessments.	TBD
Other (non-life-threatening)	Unanticipated	Family tensions or interpersonal conflicts related to increased father involvement.	TBD