BMJ Global Health

Early childhood development and parental training interventions in rural China: a systematic review and meta-analysis

Dorien Emmers ,[1,2] Qi Jiang,[2] Hao Xue,[2] Yue Zhang,[3] Yunting Zhang,[4] Yingxue Zhao,[5] Bin Liu,[6] Sarah-Eve Dill,[2] Yiwei Qian,[7] Nele Warrinnier,[1,8] Hannah Johnstone,[2] Jianhua Cai,[9] Xiaoli Wang,[10] Lei Wang,[11] Renfu Luo,[12] Guirong Li,[13] Jiajia Xu,[13] Ming Liu,[14] Yaqing Huang,[14] Wenjie Shan,[4] Zhihui Li,[4,15] Yu Zhang,[16] Sean Sylvia,[17] Yue Ma,[2] Alexis Medina,[2] Scott Rozelle[2]

To cite: Emmers D, Jiang Q, Xue H, *et al*. Early childhood development and parental training interventions in rural China: a systematic review and meta-analysis. *BMJ Global Health* 2021;**6**:e005578. doi:10.1136/bmjgh-2021-005578

Handling editor Seye Abimbola

DE and QJ are joint first authors.

Received 3 March 2021
Accepted 19 June 2021

For numbered affiliations see end of article.

Correspondence to
Dr Yue Zhang;
zhangyue0416@163.com

ABSTRACT

Introduction Inadequate care during early childhood can lead to long-term deficits in skills. Parenting programmes that encourage investment in young children are a promising tool for improving early development outcomes and long-term opportunities in low-income and middle-income regions, such as rural China.

Methods We conducted a systematic review and a meta-analysis to investigate the prevalence of early developmental delays and stimulating parenting practices as well as the effect of parental training programmes on child development outcomes in rural China. We obtained data in English from EconPapers, PubMed, PsycARTICLES, Cochrane Library, Web of Science and Scopus (Elsevier) and in Chinese from China National Knowledge Infrastructure, Wanfang Data and VIP Information. We conducted frequentist meta-analyses of aggregate data and estimated random-effects meta-regressions. Certainty of evidence was rated according to the Grading of Recommendations Assessment, Development and Evaluation approach.

Results We identified 19 observational studies on the prevalence of developmental delays and stimulating parenting practices for children under 5 years of age (n=19 762) and ten studies on the impact of parental training programmes on early child development (n=13 766). Children's risk of cognitive, language and social-emotional delays in the rural study sites (covering 14 provinces mostly in Central and Western China) was 45%, 46%, and 36%, respectively. Parental training programmes had a positive impact on child cognition, language and social-emotional development.

Conclusion There is evidence to suggest that early developmental delay and the absence of stimulating parenting practices (ie, reading, storytelling and singing with children) may be prevalent across rural, low-income and middle-income regions in Central and Western China. Results support the effectiveness of parental training programmes to improve early development by encouraging parental engagement.

Trial registration number This study was registered with PROSPERO (CRD42020218852).

Key questions

What is already known?

- Parenting programmes that encourage caregiver psychosocial stimulation investment in young children are a promising tool for improving development outcomes and long-term opportunities in life in low-income and middle-income regions, such as rural China.
- The prevalence of developmental delay, its associations with stimulating parenting practices (ie, reading, storytelling and singing with children), and impacts of parenting training programmes in poor, rural communities of China have not been well documented.
- Previous global reviews of child development and related interventions reported that developmental delays in China's rural population exist and can be improved with parental training programmes. However, no previous study has sought to quantify the overall impacts of parental training interventions.

INTRODUCTION

Investment in early childhood development (ECD) has lasting returns for both individuals and nations.[1–6] Across all low-income and middle-income countries (LMICs), however, it is estimated that up to 43% of children below the age of 5 are at risk for ECD delays.[7] Although China is one of the most rapidly developing nations in the world, it is still considered an LMIC,[8] and research has found major gaps in economic development and human capital between urban and rural areas of China.[9 10] Although China as a whole is classified by the World Bank as an upper-middle-income country, more than 70% of young children are born and raised in rural communities with living standards on par with

Key questions

What are the new findings?

- We conducted a systematic review and meta-analysis on early developmental delay, parenting practices and impacts of parental training programmes in rural China.
- We find that between 36% and 46% of young children below the age of 5 in the rural study sites which are nearly all in low-income and middle-income regions in Central and Western China are developmentally delayed. This high prevalence of delay coincides with evidence that three in four caregivers do not engage their young children in interactive reading, storytelling or singing activities.
- Parental training programmes that focus on child psychosocial stimulation can have positive impacts on child cognition, language and social-emotional scores (of 0.26 SD (95% CI: 0.18 to 0.35 SD), 0.17 SD (95% CI: 0.06 to 0.28 SD) and 0.14 SD (95% CI: 0.03 to 0.24 SD), respectively). One of the underlying mechanisms of the success of parental training programmes appears to be increasing caregiver engagement in stimulating parenting practices and increasing the parenting knowledge of caregivers (0.39 SD (95% CI: 0.24 to 0.54 SD) and 0.20 SD (95% CI: 0.11 to 0.28 SD), respectively).

What do the new findings imply?

- Early cognitive, language and social-emotional delay may be prevalent in rural communities in Central and Western China.
- Early developmental delays are partly due to infrequent engagement of caregivers in stimulating parenting practices such as reading or storytelling with children.
- Parenting interventions can improve child development outcomes. Future randomised trials are needed to inform the design and implementation of programmes at a larger scale.

lower-middle-income economies.[11 12] Given this large share of children who are growing up in rural communities, the ability of these children to acquire skills may have important implications for the nation's transition into a high-income, high-wage, skilled-labour economy.

Research on ECD in rural China has grown in recent years. In this emerging body of work, there are three main types of studies: (1) those that measure the prevalence of developmental delays[13–15]; (2) empirical work that identifies the sources of delays, including research that documents the share of caregivers who invest in stimulating parenting and a quality home environment[16 17] and (3) randomised controlled trial (RCT)-based evaluations of interventions designed to increase parental knowledge and engagement as a way to improve child development (Qian *et al*, 2020, unpublished manuscript).[18] Despite the high quality of these empirical studies, all have been regional in focus. There have been no large-scale efforts to examine ECD and parental investment across the large number of communities in dispersed geographical areas that characterise rural China. Such research is needed to determine the overall nature of ECD and parenting practices across rural communities in China and to offer policy directions for future investments in rural ECD.

The goal of this study is to conduct a meta-analysis of all empirical studies focused on ECD in rural and migrant communities in China. We aim to achieve three specific objectives: first, to document the prevalence of developmental delays among young children (below the age of 5) across rural communities in China; second, to identify the prevalence of stimulating parenting practices among rural families; and third, to assess the effectiveness of interventions that aim to increase engagement in stimulating parenting practices, increase parenting knowledge and lower risks of developmental delay. Overall, we hope to provide a holistic snapshot of ECD challenges in rural China and to systematically consider the evidence in regard to effective solutions to those challenges.

METHODS

Search strategy and selection criteria

To meet our study objectives, we conducted systematic searches of the literature in three areas: first, studies that present empirical findings on levels of cognitive, language and social-emotional development in rural China; second, studies that present empirical findings on stimulating parenting practices in rural China; and third, studies that present empirical impacts of interventions designed to improve ECD by improving parental investments (behaviours) in rural China. In conducting systematic reviews, we adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines. We searched six academic databases for studies published in English: EconPapers, PubMed, PsycARTICLES, Cochrane Library, Web of Science and Scopus (Elsevier). In addition, we searched three databases for studies published in Mandarin: China National Knowledge Infrastructure, Wanfang Data and VIP Information. The details of each systematic review are described below.

For our systematic search of empirical findings on the prevalence of ECD delays and parenting practices in rural China, we followed a two-step protocol. In the first step, we searched for empirical primary data studies in rural China published over the past 20 years (15 November 2000–15 November 2020). Our keywords included the terms: (child* OR infant*) AND "rural China" AND "early childhood" (see online supplemental appendix 1 for the full search strategy). To search empirical evidence of developmental delays, we added: AND ("development* delay" OR "cognit* delay" OR "language delay" OR "ling* delay" OR "mot* delay" OR "emotion* delay"). For our search of evidence on caregiving practices, we added: AND (parenting OR caregiving). In the next step, two independent screeners reviewed the titles and abstracts of the papers to check whether the studies satisfied our inclusion criteria. We used four inclusion criteria: (1) the study is an empirical primary data study; (2) the study concerns developmental delays and parent–child interactive caregiving practices for healthy children under age 5 (not severely malnourished; prematurely born; or suffering from a severe disease, mental trauma or disability); (3) the study is situated in rural China and (4) the study contains at least one outcome measure of

infant cognition, language or social-emotional delay or the caregiver's engagement in reading, storytelling or singing activities with the child. We restrict our attention to studies that use standardised instruments, such as the Bayley Scales of Infant Development (BSID) and the Ages and Stages Questionnaires (ASQ), for assessment of developmental delay and the Family Care Indicators (FCI) survey to evaluate the prevalence of stimulating parenting practices. In this paper, we do not include studies that assess early motor delay, as only motor development of the rural Chinese population has been shown to be normal (with an approximate 15% delay) and not a target of parenting intervention programmes.[19 20]

For our third systematic search of the literature, we aimed to gather all empirical evidence on the impacts of interventions designed to improve ECD by improving parental investments (behaviours) in rural and migrant areas of China. We searched for impact evaluations of ECD experiments in rural China published over the past 20 years (15 November 2000–15 November 2020). Our keywords included the search: (intervention* OR trial* OR experiment* OR RCT*) AND (stimulation OR nutrition) (see online supplemental appendix 1 for the full search strategy). Note that we included the search terms (stimulation OR nutrition) because we aimed to include multiple types of parenting training programmes (ie, programmes that provide parents with training on psychosocial stimulation and/or nutrition), but we did not aim to search for nutrition supplementation or micronutrient fortification programmes. In the next step, we screened the titles and abstracts of the papers to check whether the studies satisfied our inclusion criteria. We used five inclusion criteria: (1) the study is an empirical primary data study; (2) the study concerns a RCT that evaluates a parental training programme that involves children under age 5 (who are not severely malnourished; prematurely born; or suffering from a severe disease, mental trauma or disability) and their primary caregiver; (3) the study is situated and conducted in rural China; (4) the study contains at least one outcome measure of children's cognition, language or social-emotional development and (5) the study reports the means or SDs of development outcomes (to facilitate comparison across studies).

In addition to these searches, we consulted reference lists of comprehensive reviews and contacted experts on the organising committee of The 1000 Day Initiative for information and leads on any unpublished studies and data. The 1000 Day Initiative is a coalition of experts who work on ECD issues within China, with members of the organising committee who belong to top Chinese research institutions, including Peking University, Tsinghua University, Shanghai Jiao Tong University, and the National Center for Women and Child Health within the Chinese Center for Disease Control and Prevention. Finally, if data (eg, the prevalence of delay) were missing from any of the studies we found, we contacted the principal investigator of the study for additional information.

Data analysis

We first examine the prevalence of ECD delays reported in studies across rural communities in China. We define developmental delay as a cognitive, language or social-emotional development score of 1 or more SDs below the mean of a reference population whose developmental trajectory is expected to be normal (ie, a child population in developed regions who was not prematurely born, severely malnourished, or severely diseased; see online supplemental appendix 2 for more detail). This is in line with the guidelines of the BSID (the gold standard of ECD measurement), which conventionally defines children with a score of more than 1 SD below the normative sample mean as mildly delayed in their development.[21] Using 1 SD below the normative mean as a cut-off to capture all severities of developmental impairment is also in line with definitions from the Global Research on Developmental Disabilities Collaborators and the American Association on Intellectual and Developmental Disabilities.[22] Many academic studies have used 1 SD below the normative mean as their cut-off as well.[23 24] We report and compare the shares of each study sample who is delayed in cognitive, language and social-emotional development. We conduct a meta-analysis of the prevalence of delay by pooling prevalence of delays across studies using a DerSimonian and Laird random-effects model.

In addition to presenting risks of delay among our sampled children, we also investigate parental engagement in stimulating caregiver practices in rural China. We use indicators of the FCI survey, a validated survey instrument developed by UNICEF to evaluate the quality of the home environment (see table 1A).[25 26] We report on the share of parents in each study who engaged in interactive reading, storytelling or singing in the days prior to the survey. These three indicators of the FCI survey are selected because they are most commonly reported in the literature on cognitively stimulating parenting practices in rural China. In addition, we compute aggregate shares of rural caregivers in China who engage in stimulating parenting practices. We conduct a meta-analysis of the prevalence of certain parenting practices by pooling the prevalence of parenting practices across studies using a DerSimonian and Laird random-effects model.

Third, for each of the impact evaluation studies identified by our systematic search, we present the impact on children's cognitive, language and social-emotional development. To facilitate the comparison of treatment impacts across studies, we again conduct a frequentist meta-analysis of the estimated treatment impacts expressed in standardised mean differences, using a DerSimonian and Laird random-effects model. Of the studies that yield significant impacts on one or more indicators of child development, we then identified features common to all or most of the interventions to draw policy lessons for implementers and future researchers.

We use the risk of bias tool for prevalence studies developed by Hoy *et al* to assess the risk of bias for prevalence

Table 1 Summary of included studies

(A) Studies on developmental delays and stimulating parenting practices

	Authors (year)	Year of data collection	Province	n	Child age (months)	Measure: Developmental delays	Measure: Parenting practices
1	Jin *et al*[62] (2007)	2003	Anhui	100	0–24	Gesell development schedules	No
2	Ma *et al*[63] (2008)	2005–2006	Guangdong	749	4–8	BSID-I	No
3	Yang *et al*[64] (2019)	2006–2008	Shaanxi	657	24	BSID-II	No
4	Yue *et al*[65] (2017)	2014	Shaanxi	1442	18–30	BSID-I	FCI*
5	Zhou *et al*[66] (2019)	2013	Shanxi, Guizhou	2953	0–35	ASQ-3	MICS/FCI†
6	Zhang *et al*[14] (2018)	2013	Shanxi, Guizhou	2514	6–35	ASQ-3	No
7	Wei *et al*[67] (2018)	2013	Shanxi, Guizhou	2664	3–35	ASQ:SE	No
8	Luo *et al*[68] (2019) Emmers *et al* (2020), unpublished manuscript	2015	Yunnan, Hebei	215 234	6–18	BSID-III, ASQ:SE	FCI*
9	Wang *et al*[17] (2020)	2016	Shaanxi	1809	6–24	BSID-III, ASQ:SE	FCI
10	Qian *et al* (2020), unpublished manuscript	2016	Shaanxi	1068	6–18	BSID-III	FCI*
11	Li *et al* (2018), unpublished manuscript	2016	Henan	273	6–36	ASQ-3	FCI*
12	Zhou *et al*[69] (2019)	2016–2017	Jiangxi, Ningxia, Qinghai, Xinjiang	1755	1–59	ASQ-3	No
13	Wang *et al*[19] (2019)‡	2017	Beijing, Henan, Shaanxi	81 271 273	6–30	BSID-III	FCI
14	Zhong *et al*[70] (2019)	2017	Guizhou	444	1–23	BSID-III	No
15	Bai *et al* (2020), unpublished manuscript	2018	Shaanxi	260	6–36	ASQ-3, ASQ:SE	FCI
16	Tan *et al*[71] (2020)	2018	Jiangxi, Guizhou, Sichuan, Hebei, Henan	847	0–36	ASQ:SE	MICS/FCI†
17	Wang *et al* (2020), unpublished manuscript	2019	Yunnan	1014	6–24	BSID-III	FCI
18	Emmers and Wang (2020), unpublished manuscript	2019	Shaanxi	139	6–36	BSID-III, ASQ-3	FCI
Total			14	19 762			

(B) RCTs that evaluate ECD interventions in China

	Authors (year)	Year of baseline data collection	Study design	Province	n	Age of child at baseline (months)	Curriculum	Activities	Location	Frequency	Duration
1	Jin *et al*[62] (2007)	2003	RCT	Anhui	100	0–24	S	O	H	6-monthly	6 months

Continued

Table 1 Continued

(B) RCTs that evaluate ECD interventions in China

	Authors (year)	Year of baseline data collection	Study design	Province	n	Age of child at baseline (months)	Curriculum	Activities	Location	Frequency	Duration
2a	Sylvia *et al*[54] (2020)	2014	Cluster-RCT	Shaanxi	513	18–30	S+N	O	H	Weekly	6 months
2b	Wang *et al* (2020), unpublished manuscript	2014	Cluster-RCT	Shaanxi	513	18–30	S+N	O	H	Weekly	6 months¶
3	Heckman *et al*[72] (2020)	2015	Cluster-RCT	Gansu	1567	12–46	S	O	H	Weekly	2 years
4a	Luo *et al*[18] (2019)	2015	Cluster-RCT	Yunnan, Hebei	449	6–18	S+N	O	H	Bi-weekly	1 year
4b	Emmers *et al* (2020), unpublished manuscript	2015	Cluster-RCT	Yunnan, Hebei	449	6–18	S+N	O	H	Bi-weekly	2 years
5a	Zhong *et al* (2020), unpublished manuscript	2016	Cluster-RCT	Shaanxi	1720	6–24	S	O+G+P	C	Weekly one-on-one training sessions, unlimited access to centres and group sessions§	1 year
5b	Qian *et al* (2020), unpublished manuscript	2016	Cluster-RCT	Shaanxi	880	6–18	S	O+G+P	C	Weekly one-on-one training sessions, unlimited access to centres and group sessions§	2 years
6	Li *et al* (2018), unpublished manuscript	2016	Cluster-RCT	Henan	1775	6–48	S	O+G+P	C+H	Periodic one-on-one training sessions, unlimited access to centres and group sessions§	10 months
7	Bai *et al* (2020), unpublished manuscript	2018	Cluster-RCT	Shaanxi	956	6–36	S	O+G+P	C+H	Weekly one-on-one training sessions, unlimited access to centres and group sessions§	9 months
Total				6	13 766						

*Modified FCI assessed parenting practices by the primary caregiver in the day before the survey.
†MICS/FCI assessed parenting practices by the child's mother/father in the three days before the survey.
‡Wang *et al* (2019) use a pooled sample, including both original survey data (not previously reported) and survey data from previously reported studies, also included in this review. For our analysis, we include only the original data reported in Wang *et al* (2019). The remaining observations are reported in Luo *et al* (2019), Emmers *et al* (2020) and Zhong *et al* (2017).
§ECD centres were installed at centrally located places in intervention communities. Caregivers could decide how often they wanted to frequent parenting centres and/or participate in group reading or play activities at the centres.
¶Wang *et al* (2020) evaluates the persistence of treatment effects 2.5 years after program completion.
ASQ, Ages and Stages Questionnaires; BSID, Bayley Scales of Infant Development; C, ECD centre; ECD, early childhood development; FCI, family care indicators; G, group sessions; H, home; N, nutrition; O, one-on-one sessions; P, play area; RCT, randomised controlled trial; S, psychosocial stimulation.

studies of developmental delay and parenting practices.[27] We use the Effective Public Health Practice Project quality assessment tool to assess the risk of bias for RCTs.[28] Two reviewers independently rated the risk of bias of each study. Any disagreement was resolved by consensus with a third member of the review team. We then synthesised data in tabular formats. Further, we graded the overall certainty of evidence on the effectiveness of parental training programmes using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Finally, we assess heterogeneity in risk of delay and treatment impacts across studies by programme curricula and delivery mode, measurement tool and geographical location. We use frequentist meta-analysis to assess heterogeneity in prevalence rates of delay and estimated treatment impacts expressed in standardised mean differences based on a DerSimonian and Laird random-effects model. The statistical analysis was conducted with Stata version 16.1.

Role of the funding source

There was no funding source for this study.

RESULTS

We identified 19, 12 and 10 papers on the prevalence of developmental delay, the prevalence of stimulating parenting practices, and the impact of parental training on child development outcomes, respectively (see figure 1A–C). Table 1 provides an overview of the included prevalence studies (table 1A) and evaluations of RCTs (table 1B). The risk of bias was generally low to moderate after considering the observational design of the included studies (see online supplemental tables 3.1,3.2).

Table 1A shows that data for all studies included in our meta-analysis of prevalence of delay and stimulating parenting practices were collected within the past 20 years. Of the studies, 13 were published within the past 15 years, and 6 are still unpublished. They evaluate 19 different samples from low-income and middle-income rural communities in 14 provinces mostly in Central and Western China: Anhui, Beijing (a rural migrant community), Guangdong, Guizhou, Hebei, Henan, Jiangxi, Ningxia, Qinghai, Shaanxi, Shanxi, Sichuan, Xinjiang and Yunnan. Hence, when we refer to 'study sites' in this study, we mean study sites in low-income and middle-income rural regions or migrant communities in Central and Western China. All 19 762 children were below age 5 at the time of data collection. Online supplemental table 7.1 lists each study, the province in which it took place, the implementing organisation, and any involvement of authors of this review.

Table 1B provides an overview of the 10 studies included in our meta-analyses of the impacts of parental training programmes on child development and parenting practices. Baseline data were collected within the past 20 years. Four of the studies were published within the past 15 years, and six are still unpublished. They evaluate nine different RCTs implemented in low-income and middle-income rural communities in six provinces mostly in Central and Western China: Anhui, Gansu, Hebei, Henan, Shaanxi and Yunnan. All 13 766 children who were involved in one of these studies were below 4 years of age at baseline (most of the children were below 30 months of age at baseline). In each of these RCTs, parental training on psychosocial stimulation (and, in some cases, nutrition) was delivered during one-on-one sessions over a period of 6 months up to 2 years. The frequency of the one-on-one training

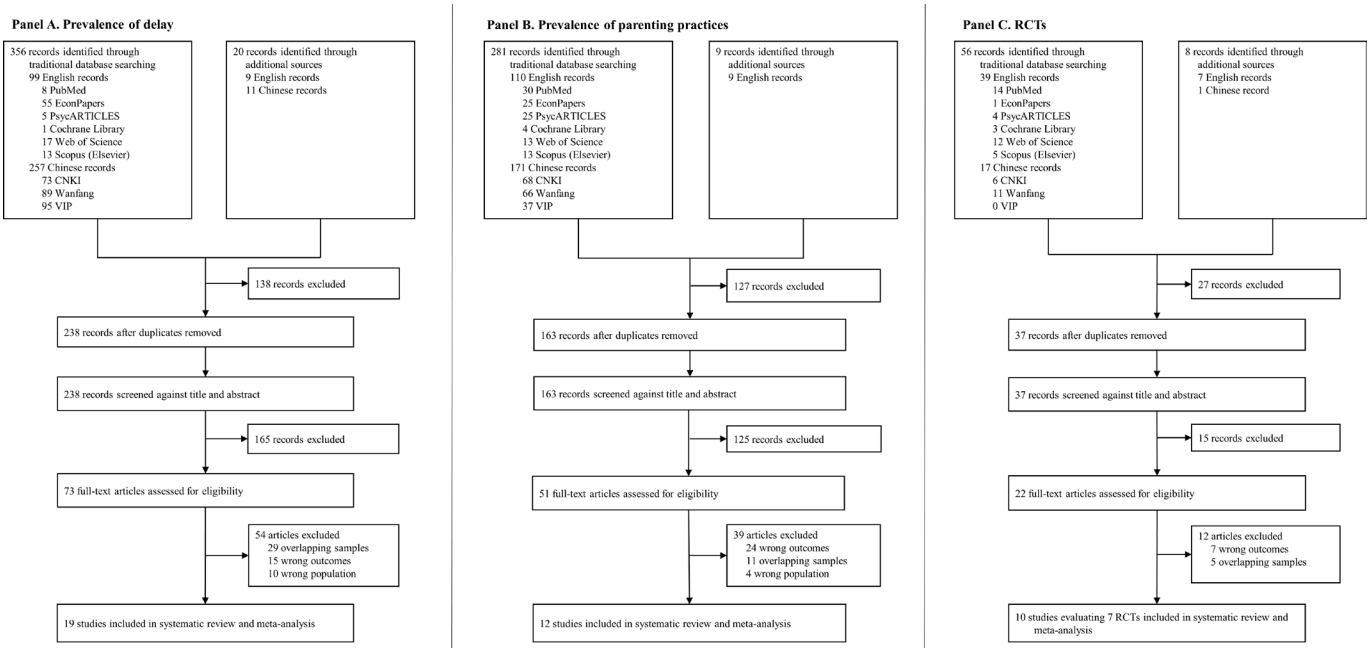


Figure 1 Study selection. CNKI, China National Knowledge Infrastructure; RCT, randomised controlled trial.

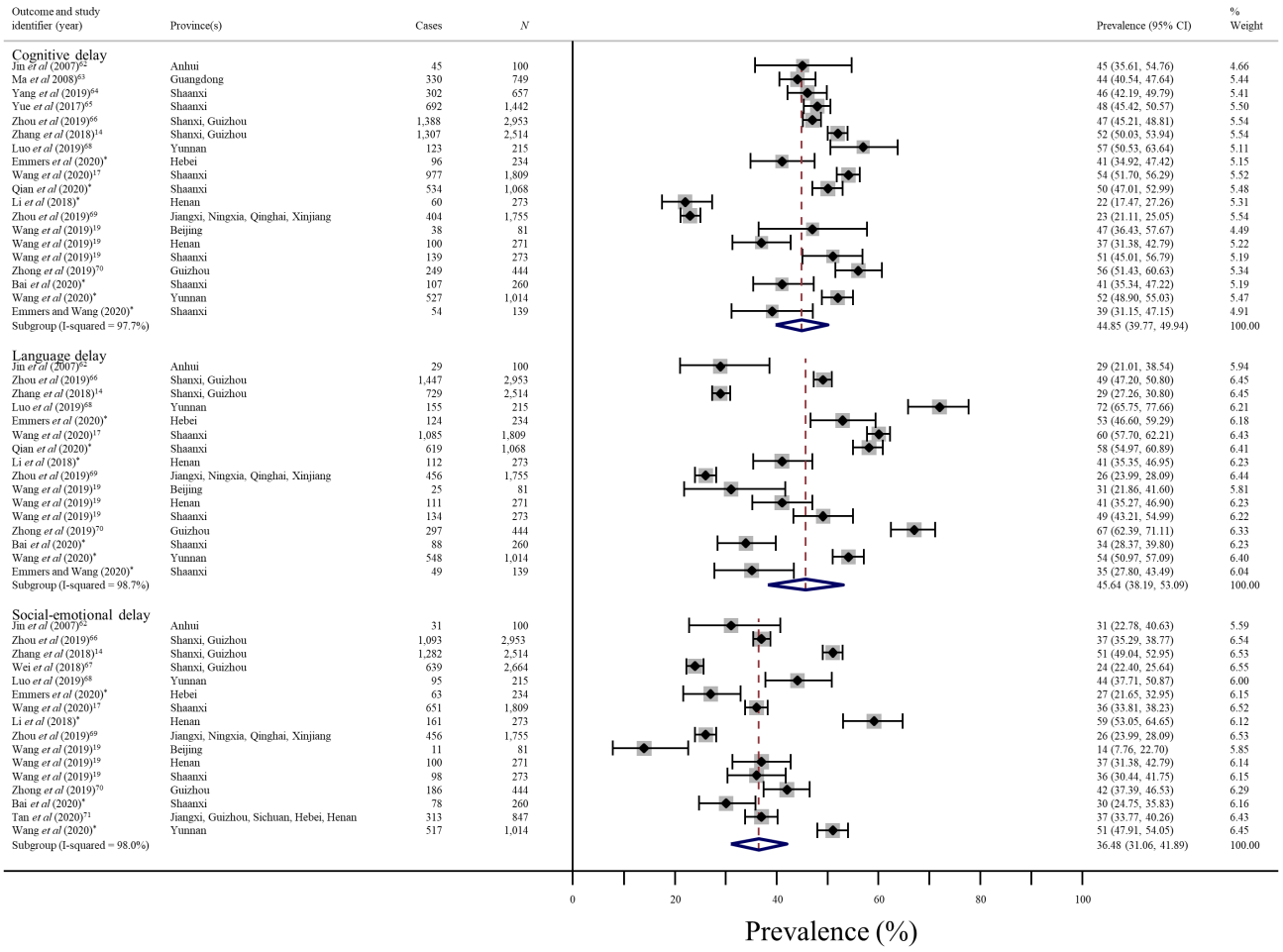


Figure 2 Pooled prevalence of ECD delay. * Unpublished manuscript.

sessions varied from weekly to once every 6 months. One-on-one sessions were delivered by parenting trainers at home or in ECD centres that were installed at centrally located venues. If an ECD centre was installed, then intervention households received unlimited access to the open play areas and group training sessions organised at the centres (Bai *et al*, 2020; Zhong *et al*, unpublished manuscripts). All studies with outcome measures of cognition, language or social-emotional development or parenting practices or parenting knowledge are included in the respective meta-analysis.

Figure 2 presents estimates of ECD delays among infants and toddlers in the pooled sample. The share of infants and toddlers with cognitive delays ranges from 22% to 57% (see figure 2, top panel). The weighted average across all studies shows that, of the 19 762 infants and toddlers in the pooled sample, 44.8% (95% CI: 39.8% to 49.9%) were at risk of cognitive delay. The average risk of language delay in each study ranges from 26% to 72%, and the weighted average of the reported studies shows a language delay risk of 45.6% (95% CI: 38.2% to 53.1%) (see middle panel). Risk of social-emotional delays were slightly lower, ranging from 14% to 59%, and the weighted average shows that 36.5% (95% CI: 31.1% to 41.9%) of samples display social-emotional delays (see figure 2, bottom panel).

Figure 3 shows the prevalence of stimulating parenting practices among caregivers across the set of pooled studies. Between 2% and 76% of caregivers had read a book to their child over the past 3 days, with only 23.3% (95% CI: 9.4% to 37.2%) of caregivers, on average, having read a book to their child across the studies (see figure 3, top panel). The share of caregivers who sang songs with their children in the past 3 days ranged between 19% and 85%, with a weighted average of 44.8% (95% CI: 29.5% to 60.1%) across all datasets (see figure 3, middle panel). Between 9% and 76% of caregivers told stories to their child in the past 3 days, and the weighted average of all datasets shows that only 25.2% (95% CI: 9.5% to 40.9%) of caregivers had told a story to their child in the past 3 days (see figure 3, bottom panel).

Figure 4 portrays the intervention effects of the reviewed parental training programmes on child cognitive development and parenting practices (reading, storytelling and singing with children).

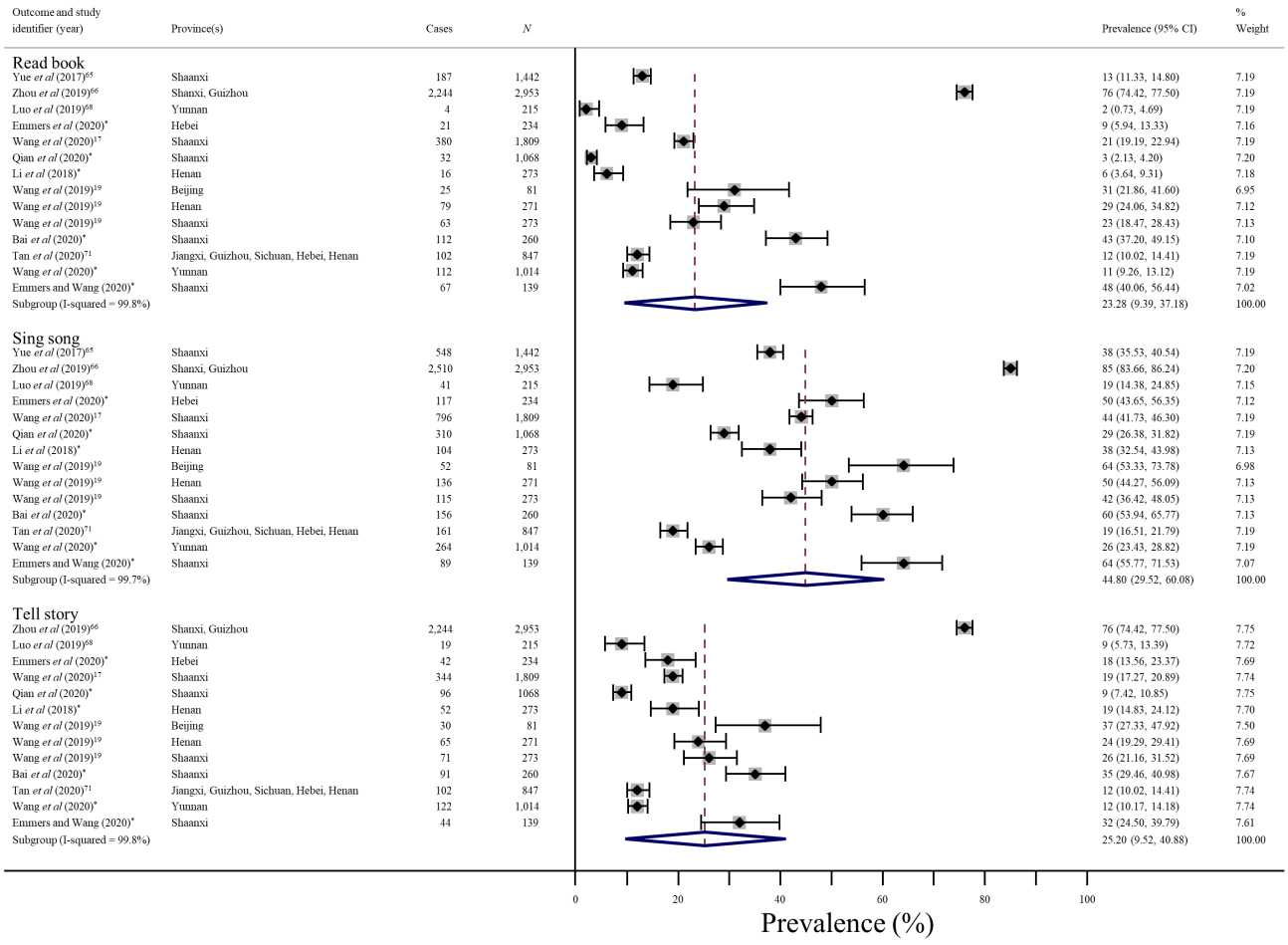


Figure 3 Pooled engagement in interactive caregiver-child activities. * Unpublished manuscript.

All of the 10 RCT evaluations include an outcome measure of cognitive development scores (see top panel of figure 4), and all find significant short-run improvements in cognition. The size of the impact on cognition ranges from 0.11 SD to 0.75 SD The mean standardised effect size of child cognition is 0.26 SD (95% CI: 0.18 to 0.35 SD). Further, the bottom panel of figure 4 shows that seven studies reported impacts of parenting training interventions on a summary measure of parental engagement in stimulating parenting practices, with a mean effect size of 0.39 SD (95% CI: 0.24 to 0.54 SD).

Treatment impacts on child language and social-emotional development and parental knowledge and beliefs are summarised in online supplemental figures 5.1 and 5.2. Impacts on child language and social-emotional development are each measured in nine of the 10 intervention studies. For both outcome measures, impacts are less pronounced than those for child cognition. Three studies had a significant impact on language development, and two had a significant impact on social-emotional development. The mean standardised effect size is 0.17 SD (95% CI: 0.06 to 0.28 SD) for language development and 0.14 SD (95% CI: 0.03 to 0.24 SD) for social-emotional development. Online supplemental figure 5.2 shows that six of the intervention studies reported intervention impacts on parenting knowledge and beliefs, with a mean effect size of 0.20 SD (0.11–0.28 SD).

Based on a GRADE summary analysis, we conclude with high certainty that parental training programmes can improve cognitive outcomes of children under age 5 in rural Chinese communities (see first row of online supplemental table 4.1). Further, such training programmes can, with moderate certainty, raise language and social-emotional development of young children (see rows 2 and 3 of online supplemental table 4.1). Finally, we conclude with moderate certainty that the parenting programmes led to increases in stimulating parenting practices and parenting knowledge (see rows 4 and 5 of online supplemental table 4.1). This evidence indicates that, after receiving training for improving the skill development of their young children, parents are likely to both learn and to practice what they learn.

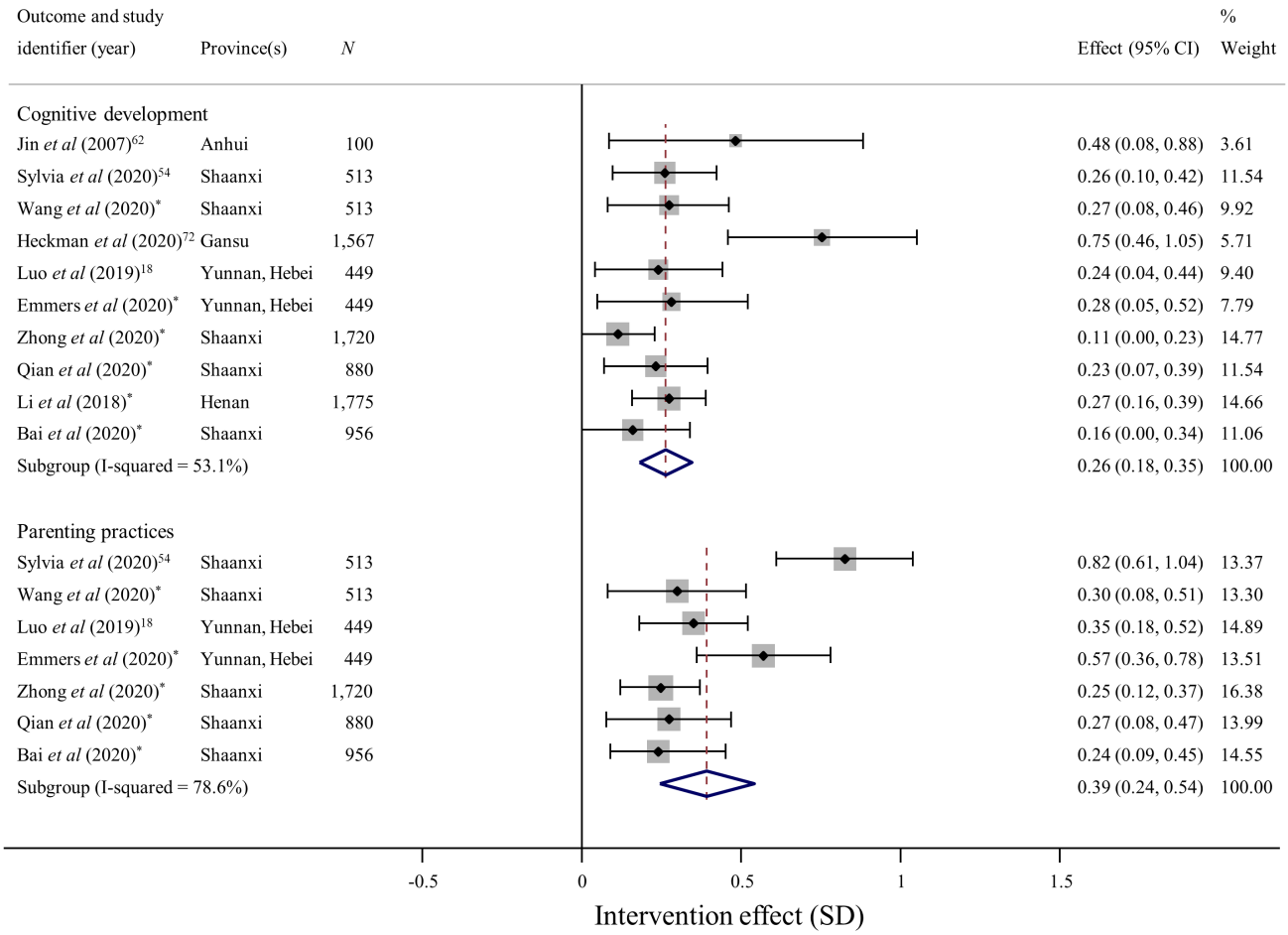


Figure 4 Intervention effects on child cognitive development and parenting practices. * Unpublished manuscript.

Note that one (unpublished) study by Wang *et al*[17] included in our meta-analysis, evaluates the persistence of treatment effects 2.5 years after programme completion. This study provides important evidence in support of persistent treatment impacts on child cognition, parental investment and parental beliefs and knowledge in the longer run.

We also conduct an analysis of heterogeneity in prevalence and child development delays and estimated treatment impacts by programme curriculum, delivery mode, measurement tool and geographical location. We find that centre-based and home-based programmes both have a positive and significant impact on child cognitive development (see online supplemental figure 6.4). The average impact of centre-based programmes, however, is smaller than the average impact of home-based programmes (ie, 0.19 SD (95% CI: 0.11 to 0.27 SD) and 0.34 SD (95% CI: 0.21 to 0.48 SD), p value of between-group heterogeneity <0.05). The remainder of the results of the heterogeneity analysis are presented and discussed in detail in online supplemental appendix 6.

DISCUSSION

In this study, we aggregate all empirical studies of ECD in rural China, providing a holistic snapshot of ECD challenges in low-income and middle-income rural communities in Central and Western China, as well as the effectiveness of interventions designed to address these challenges. A major finding is evidence that suggests that early developmental delays may be prevalent across rural communities oin Central and Western China. The studies spanned 14 provinces, covering all of China's major geographical regions. They included samples from villages, townships, and migrant communities and focused on healthy children under 5 years old. A major strength of this study is that most of the included studies evaluated population-based random samples, which affirms the representativeness of the study samples. All of the included studies identified rates of developmental delays that were higher than what would be expected from a healthy population. The average prevalence of cognitive delay across all studies, based on a random-effects meta-analysis, was 45%. This is comparable to that observed in other LMICs (42%).[29]

Because delays at this age have been linked to lifelong impacts on school performance, educational attainment, job market outcomes and even criminal activity, these high risks of delay should be viewed as a key area for intervention to promote long-term individual and social well-being.

A second major finding is that rural parents infrequently engage in stimulating parenting practices, such as reading, storytelling and singing with their children. On average, 23% of families read with their children on a regular basis, and 25% tell stories to their children. Frequent engagement in singing is somewhat higher, at 45%. These prevalence levels are consistent with those of Walker *et al*[30] and somewhat lower than those of Nonoyama-Tarumi and Ota,[31] two studies conducted in developing settings around the world. The low prevalence of parent-child engagement can be seen as a reflection of the economic inequality that demarcates rural and urban areas.[32] The human development literature is unanimous in establishing that households with higher incomes and in more developed regions on average invest more time and resources in their children.[33]

There are several factors that may limit caregivers in the rural study sites from frequently engaging in stimulating parenting practices, including financial stress, time constraints, knowledge constraints and mental health problems of the caregiver. A handful of quantitative and qualitative studies in rural China have found knowledge constraints to be the underlying primary factor for the low prevalence of stimulating parenting practices in rural China. Although rural households typically have fewer economic resources than urban households, two studies found that the majority of rural households have sufficient economic resources to invest in their children's development as well as a desire to do so.[16 34] Similarly, a recent study found that many rural caregivers reported having sufficient leisure time that could be used for interactive play with their children.[34] In contrast, there is evidence that rural caregivers do not have salient knowledge of how to effectively engage in stimulating parenting practices and often lack reliable sources of information about parenting.[16 34] This may be compounded by mental health issues among rural caregivers in China, which several studies have found to be prevalent and linked to reduced ECD outcomes.[35–38]

In light of these problems, 10 studies have completed rigorous evaluations of ECD interventions among the 0–5 age group. These interventions are designed to improve ECD primarily by targeting caregiver knowledge and skills through one-on-one parental training. We find with high certainty that these interventions improve early cognitive development (see online supplemental table 4.1). We find with moderate certainty that they improve early language and social-emotional development as well as increase parenting knowledge and engagement in stimulating parenting practices. These findings are consistent with findings from other low-resource settings globally.[20 39–48] These impacts suggest that increases in parental investment in stimulating parenting practices, such as reading, storytelling and singing, and changes in parenting knowledge are important mechanisms behind the impacts on child development.

Several key lessons emerge from the evaluation studies included in our meta-analysis. First, we find that one-on-one parental training is effective for improving early cognitive development in rural China. Each of the interventions included in our analysis contained some version of one-on-one training (see table 1B), and all found positive and significant impacts on children's cognitive development (see top panel of figure 4). Six different curricula were used; delivery modes included in-home and centre-based parental training sessions; and programmes were variously implemented by local village women, volunteers, doctors and family planning officials. The main commonality among the interventions was the use of a one-on-one training component. We did not identify any evaluations of group-based interventions in rural China, another delivery format that has been used in ECD interventions internationally.[49–52] The results of our meta-analysis suggest, however, that, among the many different approaches to crafting an effective ECD intervention in rural China, one-on-one parental training appears to be a common element.

A second major lesson learnt from our meta-analysis of ECD interventions is that, although centre-based and home-based interventions have positive impacts on ECD, home-based interventions seem to have a larger impact on child cognition (0.34 SD vs 0.19 SD; see online supplemental figure 6.4). This can be partly explained by the fact that the most vulnerable children, who would benefit the most from this type of intervention, are more likely to miss out on a centre-based programme than on a home-based programme.[53] To understand the reason, we contrast findings from the unpublished study of Zhong *et al*, a centre-based intervention, and Sylvia *et al*,[54] a home-based intervention. The two studies were conducted by the same research group; the interventions used one-on-one parental training, following the same curriculum, and both were implemented in the same region of the country. The main difference was the location of the training: One was delivered in the home, and the other was delivered in a parenting centre. In the home-based intervention, the highest impacts are observed among children with the lowest levels of parental investment at baseline. In contrast, in the centre-based intervention, children with the lowest levels of parental investment self-select out of the intervention due to hurdles such as high opportunity cost of parents' time to visit the centre or limited access to reliable and effective transportation. The findings of these two studies are consistent with the research finding that participation

in centre-based programmes is widely unequal in developing settings.[31 55]

We acknowledge limitations to our work. First, we define developmental delay as a cognitive, language or social-emotional development score of 1 or more SDs below the mean of a reference population whose developmental trajectory is expected to be normal (see online supplemental appendix 2 for more detailed information). Although many other academic studies also use 1 SD below the normative mean as their cut-off,[23 24] others may use different cut-offs, so caution should be used when comparing our results with those from other settings. Second, the reviewed studies use various measurement tools to assess the prevalence of ECD delay. The BSID is a comprehensive diagnosis tool for developmental progress, whereas the ASQ-3 test is a short screening tool for the detection of children at risk of delay. Studies have found that the ASQ-3 may be a less accurate measurement tool for the detection of developmental delay, although evidence on whether the ASQ-3 test overestimates or underestimates the prevalence of developmental delay as compared with BSID assessments is mixed.[56–58] Our measurement-based heterogeneity analysis finds that the point estimates of the prevalence of cognition and language delay are slightly lower in studies that use the ASQ-3 as compared with studies that use the BSID (37.0% vs 48.2% and 35.8% vs 52.7%; see online supplemental figures 6.7 and 6.8, respectively), but we do not detect a difference in the observed prevalence of social-emotional delay when using ASQ-3, ASQ:SE, or BSID (40.5% vs 32.3% vs 36.1%; online supplemental figure 6.9). Note that, regardless of the measurement tool used, the prevalence of delay is high. Further, based on the results presented in online supplemental appendix figures 6.10–6.12, we find no systematic differences in estimated treatment impacts on early cognitive, language and social-emotional development across studies that used the ASQ-3 or the BSID.

We also acknowledge the inclusion of several unpublished papers and reports in our analyses. Because these studies have not yet undergone peer review, there may be unintentional errors or biases in their data, although it should be noted that we found low levels of bias in all of our formal bias checks. As argued by Dwan *et al*,[59] empirical research shows that results in published studies are more likely to be positive and statistically significant than are results in unpublished studies. Moreover, publishing studies with insignificant impacts is more difficult and may take longer, which may lead to a 'pipeline bias.' This type of publication bias is a threat to the validity of systematic reviews. In line with routine practice in meta-analyses, we use a visual analysis of a funnel plot, the trim-and-fill method, and Egger's test of small-study effects to test for publication bias.[59–61] We find no evidence of publication bias when we combine the published with the unpublished studies (see online supplemental appendix 7 for a detailed discussion). Hence, we concluded that the benefits of including unpublished studies outweigh the limitations.

In conclusion, the reviewed evidence shows consistently that early developmental delay and the absence of stimulating parenting practices (ie, reading, storytelling and singing with children) may be prevalent across rural communities that are similar to populations in the study sites in Central and Western China. Our meta-analysis, however, has uncovered several gaps in the literature with regard to the potential role that intervention programmes can play. First, although we find that parenting training programmes have large and consistent impacts on children's cognitive development, we also find that they have smaller and less-consistent impacts on social-emotional and language development. Therefore, future in-depth evaluations of potential impacts on these development areas, including studies that examine the mechanisms behind the observed impacts on these dimensions of development, could provide useful information to guide future intervention designs. Second, it remains unclear to what extent intervention programmes are able to reach the most vulnerable families, who are likely to benefit the most from this type of intervention programme. For example, could a clinic-based setting take advantage of the existing infrastructure and leverage existing social trust to host an effective and inclusive intervention? How can practitioners help to stem attrition from ECD interventions? Limited global evidence suggests that the most marginalised populations are more likely to miss out on clinic-based and centre-based interventions due to limiting factors, such as a remote living location or time constraints.[53 55] Further research on inclusiveness, participation constraints and optimal programme targeting can shed light on how we can provide more children, including the most disadvantaged ones, with a fair start in life.

Author affiliations
[1]LICOS Centre for Institutions and Economic Performance, KU Leuven, Leuven, Belgium
[2]Stanford Center on China's Economy and Institutions, Freeman Spogli Institute for International Studies & Stanford Institute for Economic Policy Research, Stanford University, Stanford, California, USA
[3]National Center for Women and Children Health, Chinese Center for Disease Control and Prevention, Beijing, China
[4]Child Health Advocacy Institute, Department of Developmental and Behavioral Pediatrics, National Children's Medical Center, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[5]The Thousand Day Initiative, Beijing, China
[6]Xinhe Foundation, Beijing, China
[7]Department of Economics, University of Southern California, Los Angeles, California, USA
[8]School of Economics and Finance, Queen Mary University of London, London, UK
[9]Administrators of Training Center of the National Health Commission of the PRC, Beijing, China
[10]School of Public Health, Peking University Health Science Centre, Beijing, China
[11]International Business School, Shaanxi Normal University, Xi'an, Shaanxi, China
[12]School of Advanced Agricultural Sciences, Peking University, Beijing, China
[13]International Center for Action Research on Education, Henan University School of Education, Kaifeng, Henan, China
[14]Save the Children International China Program, Beijing, China

[15]Vanke School of Public Health, Tsinghua University, Beijing, China
[16]Hupan Modou Foundation, Hangzhou, China
[17]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Acknowledgements We would also like to acknowledge the effort of several of the groups that were part of the research efforts, especially those that helped in designing, supporting and/or implementing the interventions, including UNICEF, OneSky, China Development Research Foundation, Save the Children, and others.

Contributors DE and QJ contributed equally to this study. DE conceived the study, collected the data, verified the underlying data, conducted the statistical analysis, interpreted the data and drafted the manuscript. QJ conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. HX contributed to the statistical analysis, verified the underlying data, interpreted the data, conducted the bias assessment, and provided critical edits to the manuscript. YueZ contributed to the statistical analysis, verified the underlying data, interpreted the data, and provided critical edits to the manuscript. YTZ contributed to the statistical analysis, verified the underlying data, interpreted the data, and provided critical edits to the manuscript. YXZ reviewed and provided critical edits to the manuscript. BL reviewed and provided critical edits to the manuscript. S-ED collected the data, conducted the bias assessment, interpreted the data and helped to draft the manuscript. YQ conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. NW conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. HJ conducted the bias assessment, interpreted the data and provided critical edits to the manuscript. JC provided critical edits to the manuscript. XW conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. LW conducted the statistical analysis, interpreted the data and provided critical edits to the manuscript. RL reviewed and provided critical edits to the manuscript. GL conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. JX conducted the statistical analysis and interpreted the data. ML conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. YH conducted the statistical analysis, interpreted the data and provided critical edits to the manuscript. WS conducted the statistical analysis, interpreted the data and provided critical edits to the manuscript. ZL conducted the statistical analysis, interpreted the data and provided critical edits to the manuscript. YuZ conceived the study, interpreted the data, and provided critical edits to the manuscript. SS conceived the study, interpreted the data and provided critical edits to the manuscript. YM conducted the statistical analysis, interpreted the data, and provided critical edits to the manuscript. AM interpreted the data, conducted the bias assessment and drafted the manuscript. SR conceived the study, collected the data, interpreted the data and provided critical edits to the manuscript.

Funding DE acknowledges support from KU Leuven, Long-term structural funding-Methusalem funding by the Flemish Government and by the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO), and the Fonds de la Recherche Scientifique-FNRS under EOS Project No. G0G4318N (EOS ID 30784531).

Competing interests We have listed the institutions or organisations that were in charge of programme implementation and the involvement of coauthors of this manuscript in reviewed intervention studies in online supplemental appendix table 7.1.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on request. The Stata code and data permissible for sharing will be made available on request from DE (dorien.emmers@kuleuven.be).

ORCID iD
Dorien Emmers http://orcid.org/0000-0001-7750-9196

REFERENCES

1 Grantham-McGregor S, Cheung YB, Cueto S, *et al*. Developmental potential in the first 5 years for children in developing countries. *Lancet* 2007;369:60–70.
2 Knudsen EI, Heckman JJ, Cameron JL, *et al*. Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proc Natl Acad Sci U S A* 2006;103:10155–62.
3 Attanasio OP. The determinants of human capital formation during the early years of life: theory, measurement, and policies. *J Eur Econ Assoc* 2015;13:949–97.
4 Heckman JJ. Skill formation and the economics of investing in disadvantaged children. *Science* 2006;312:1900–2.
5 Heckman JJ, Moon SH, Pinto R, *et al*. The rate of return to the High/Scope Perry preschool program. *J Public Econ* 2010;94:114–28.
6 Almond D, Currie J. Human capital development before age five. In: Ashenfelter O, Card D, eds. *Handbook of labor economics*. Amsterdam: Elsevier, 2011: 1315–486.
7 Lu C, Black MM, Richter LM. Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level. *Lancet Glob Health* 2016;4:e916–22.
8 Glawe L, Wagner H. China in the middle-income trap? *China Econ Rev* 2020;60:101264.
9 Wang L, Li M, Abbey C, *et al*. Human capital and the middle income trap: how many of China's youth are going to high school? *Dev Econ* 2018;56:82–103.
10 Bai Y, Zhang S, Wang L, *et al*. Past successes and future challenges in rural China's human capital. *J Contemp China* 2019;28:883–98.
11 National Bureau of Statistics of the People's Republic of China. National statistical yearbook, 2010. Available: http://www.stats.gov.cn/tjsj/ndsj/2010/indexeh.htm [Accessed 01 Dec 2020].
12 National Bureau of Statistics of the People's Republic of China. National statistical yearbook, 2018. Available: http://www.stats.gov.cn/tjsj/ndsj/2018/indexeh.htm [Accessed 01 Dec 2020].
13 Wei QW, Zhang JX, Scherpbier RW, *et al*. High prevalence of developmental delay among children under three years of age in poverty-stricken areas of China. *Public Health* 2015;129:1610–7.
14 Zhang J, Guo S, Li Y, *et al*. Factors influencing developmental delay among young children in poor rural China: a latent variable approach. *BMJ Open* 2018;8:e021628.
15 Zhou H, Ye R, Sylvia S, *et al*. "At three years of age, we can see the future": Cognitive skills and the life cycle of rural Chinese children. *Demogr Res* 2020;43:169–82.
16 Yue A, Shi Y, Luo R, *et al*. Stimulation and early child development in China: caregiving at arm's length. *J Dev Behav Pediatr* 2019;40:458–67.
17 Wang B, Luo X, Yue A, *et al*. Family environment in rural China and the link with early childhood development. *Early Child Dev Care* 2020.
18 Luo R, Emmers D, Warrinnier N, *et al*. Using community health workers to deliver a scalable integrated parenting program in rural China: a cluster-randomized controlled trial. *Soc Sci Med* 2019;239:112545.
19 Wang L, Liang W, Zhang S, *et al*. Are infant/toddler developmental delays a problem across rural China? *J Comp Econ* 2019;47:458–69.
20 Singla DR, Kumbakumba E, Aboud FE. Effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural Uganda: a community-based, cluster randomised trial. *Lancet Glob Health* 2015;3:e458–69.
21 Caesar R, Boyd RN, Colditz P, *et al*. Early prediction of typical outcome and mild developmental delay for prioritisation of service delivery for very preterm and very low birthweight infants: a study protocol. *BMJ Open* 2016;6:e010726.
22 Global Research on Developmental Disabilities Collaborators. Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet Glob Health* 2018;6:e1100–21.
23 Cromwell EA, Dube Q, Cole SR, *et al*. Validity of US norms for the Bayley scales of infant Development-III in Malawian children. *Eur J Paediatr Neurol* 2014;18:223–30.

24 Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatr Res* 2014;75:670–4.
25 Frongillo E, Sywulka S, Kariger P. *UNICEF psychosocial care indicators project: final report to UNICEF*. Ithaca: Cornell University Division of Nutritional Sciences, 2003.
26 Hamadani JD, Tofail F, Hilaly A, *et al*. Use of family care indicators and their relationship with child development in Bangladesh. *J Health Popul Nutr* 2010;28:23–33.
27 Hoy D, Brooks P, Woolf A, *et al*. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol* 2012;65:934–9.
28 Effective Public Health Practice Project. Quality assessment tool for quantitative studies, 2010. Available: http://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf [Accessed 30 Nov 2020].
29 Gil JD, Ewerling F, Ferreira LZ, *et al*. Early childhood suspected developmental delay in 63 low- and middle-income countries: large within- and between-country inequalities documented using National health surveys. *J Glob Health* 2020;10:010427.
30 Walker SP, Wachs TD, Gardner JM, *et al*. Child development: risk factors for adverse outcomes in developing countries. *Lancet* 2007;369:145–57.
31 Nonoyama-Tarumi Y, Ota Y. *Early childhood development in developing countries: pre-primary education, parenting, and health care. background paper education for all global monitoring report*. 2011, 2010.
32 Zhang J. A survey on income inequality in China. *J Econ Lit* 2021.
33 Francesconi M, Heckman JJ. Child development and parental investment: introduction. *Econ J* 2016;126:F1–27.
34 Li R, Rose N, Zheng YM, *et al*. Early childhood reading in rural China and obstacles to caregiver investment in young children: a mixed-methods analysis. *Int J Environ Res Public Health* 2021;18:1457.
35 Yue A, Gao J, Yang M, *et al*. Caregiver depression and early child development: a mixed-methods study from rural China. *Front Psychol* 2018;9:2500.
36 Zhang S, Dang R, Yang N, *et al*. Effect of caregiver's mental health on early childhood development across different rural communities in China. *Int J Environ Res Public Health* 2018;15:2341.
37 Zhang S, Wang L, Xian Y, *et al*. Mental health issues among caregivers of young children in rural China: prevalence, risk factors, and links to child developmental outcomes. *Int J Environ Res Public Health* 2021;18:197.
38 Zhong J, Wang T, He Y, *et al*. Interrelationships of caregiver mental health, parenting practices, and child development in rural China. *Child Youth Serv Rev* 2021;121:105855.
39 Aboud FE, Akhter S. A cluster-randomized evaluation of a responsive stimulation and feeding intervention in Bangladesh. *Pediatrics* 2011;127:e1191–7.
40 Aboud FE. Evaluation of an early childhood parenting programme in rural Bangladesh. *J Health Popul Nutr* 2007;25:3–13.
41 Andrew A, Attanasio O, Augsburg B, *et al*. Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial. *J Child Psychol Psychiatry* 2020;61:644–52.
42 Attanasio OP, Fernández C, Fitzsimons EOA, *et al*. Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. *BMJ* 2014;349:g5785–12.
43 Attanasio O, Baker-Henningham H, Bernal R. Early stimulation and nutrition: the impacts of a scalable intervention. National Bureau of economic research (NBER) working paper series, 2018. Available: https://www.nber.org/papers/w25059 [Accessed 02 Dec 2020].
44 Chang SM, Grantham-McGregor SM, Powell CA, *et al*. Integrating a parenting intervention with routine primary health care: a cluster randomized trial. *Pediatrics* 2015;136:272–80.
45 Fernald LCH, Kagawa RMC, Knauer HA, *et al*. Promoting child development through group-based parent support within a cash transfer program: experimental effects on children's outcomes. *Dev Psychol* 2017;53:222–36.
46 Rao N, Sun J, Chen EE. Effectiveness of early childhood interventions in promoting cognitive development in developing countries: a systematic review and meta-analysis. *Hong Kong J Paediatr* 2017;22:14–25.
47 Worku BN, Abessa TG, Wondafrash M, *et al*. Effects of home-based play-assisted stimulation on developmental performances of children living in extreme poverty: a randomized single-blind controlled trial. *BMC Pediatr* 2018;18:1–11.
48 Yousafzai AK, Rasheed MA, Rizvi A, *et al*. Effect of integrated responsive stimulation and nutrition interventions in the lady health worker programme in Pakistan on child development, growth, and health outcomes: a cluster-randomised factorial effectiveness trial. *Lancet* 2014;384:1282–93.
49 Mehrin SF, Hamadani JD, Salveen N-E, *et al*. Adapting an evidence-based, early childhood parenting programme for integration into government primary health care services in rural Bangladesh. *Front Public Health* 2020;8:1052.
50 Walker S, Powell C, Chang S. Delivering parenting interventions through health services in the Caribbean. *IZA Discussion Paper* 2015;12106.
51 Grantham-McGregor S, Adya A, Attanasio O, *et al*. Group sessions or home visits for early childhood development in India: a cluster RCT. *Pediatrics* 2020;146:e2020002725.
52 Encouraging early childhood stimulation from parents and caregivers to improve child development, 2020. Abdul Latif Jameel poverty action lab. Available: J-PAL Policy Insights
53 Bendini M. Parents or centers: how should governments prioritize early investments in children? 2017. Available: https://blogs.worldbank.org/developmenttalk [Accessed June 8, 2018].
54 Sylvia S, Warrinnier N, Luo R. From quantity to quality: delivering a home-based parenting intervention through China's family planning cadres. *Econ J* 2020.
55 Hamadani JD, Mehrin SF, Tofail F, *et al*. Integrating an early childhood development programme into Bangladeshi primary health-care services: an open-label, cluster-randomised controlled trial. *Lancet Glob Health* 2019;7:e366–75.
56 Schonhaut L, Armijo I, Schönstedt M, *et al*. Validity of the ages and stages questionnaires in term and preterm infants. *Pediatrics* 2013;131:e1468–74.
57 Yue A, Jiang Q, Wang B, *et al*. Concurrent validity of the ages and stages questionnaire and the Bayley scales of infant development III in China. *PLoS One* 2019;14:e0221675.
58 Steenis LJP, Verhoeven M, Hessen DJ, *et al*. Parental and professional assessment of early child development: the ASQ-3 and the Bayley-III-NL. *Early Hum Dev* 2015;91:217–25.
59 Dwan K, Gamble C, Williamson PR, *et al*. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* 2013;8:e66844.
60 Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics* 2018;74:785–94.
61 Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455–63.
62 Jin X, Sun Y, Jiang F, *et al*. "Care for Development" intervention in rural China: a prospective follow-up study. *J Dev Behav Pediatr* 2007;28:213–8.
63 Ma L, Chi L, Su Y. Investigation of development and growth status of 0–18 months children in Guangdong (in Chinese). *Matern Child Health J* 2008;13:1844–7.
64 Yang X, Yin Z, Cheng Y, *et al*. Features and associated factors of the behavioral development of 24-month-old children in rural China: follow-up evaluation of a randomized controlled trial. *Sci Rep* 2018;8:13977.
65 Yue A, Shi Y, Luo R, *et al*. China's invisible crisis: cognitive delays among rural toddlers and the absence of modern parenting. *China J* 2017;78:50–80.
66 Zhou S, Zhao C, Huang X, *et al*. The effect of a community-based, integrated and nurturing care intervention on early childhood development in rural China. *Public Health* 2019;167:125–35.
67 Wei Q, Zhang C, Zhang J, *et al*. Caregiver's depressive symptoms and young children's socioemotional development delays: a cross-sectional study in poor rural areas of China. *Infant Ment Health J* 2018;39:209–19.
68 Luo R, Jia F, Yue A, *et al*. Passive parenting and its association with early child development. *Early Child Dev Care* 2019;189:1709–23.
69 Zhou H, Ding Y, Yang Y, *et al*. Effects on developmental outcomes after cesarean birth versus vaginal birth in Chinese children aged 1-59 months: a cross-sectional community-based survey. *PeerJ* 2019;7:e7902.
70 Zhong J, Gao J, Liu C, *et al*. Quantity–quality trade-off and early childhood development in rural family: evidence from China's Guizhou province. *Int J Environ Res Public Health* 2019;16:1307.
71 Tan C, Zhao C, Dou Y, *et al*. Caregivers' depressive symptoms and social–emotional development of left-behind children under 3 years old in poor rural China: the mediating role of home environment. *Child Youth Serv Rev* 2020;116:105109.
72 Heckman JJ, Liu B, Lu M. Treatment effects and the measurement of skills in a prototypical home visiting program. National Bureau of economics research (NBER) working paper series, 2020. Available: https://www.nber.org/papers/w27356 [Accessed 15 Nov 2020].

Appendix

Appendix 1: Examples of Search Strings by Language, Topic, and Database

Search strings in English:

1. Delays: 99 in total
    a. PubMed
        i. Search conducted on November 15, 2020
        ii. ((((“development* delay” OR “cognit* delay” OR “language delay” OR “ling* delay” OR “mot* delay” OR “emotion* delay”) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)
        iii. Number of search results: 8
    b. EconPapers
        i. Search conducted on November 15, 2020
        ii. ((((“development* delay” OR “cognit* delay” OR “language delay” OR “ling* delay” OR “mot* delay” OR “emotion* delay”) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)
        iii. Number of search results: 55
    c. PsycARTICLES
        i. Search conducted on November 15, 2020
        ii. ((((“development* delay” OR “cognit* delay” OR “language delay” OR “ling* delay” OR “mot* delay” OR “emotion* delay”) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)
        iii. Number of search results: 5
    d. Cochrane Library
        i. Search conducted on November 15, 2020
        ii. TITLE-ABS-KEY(((((“development* delay” OR “cognit* delay” OR “language delay” OR “ling* delay” OR “mot* delay” OR “emotion* delay”) AND (child* OR infant*)) AND “rural China”) AND “early childhood”))
        iii. Number of search results: 1
    e. Web of Science
        i. Search conducted on November 15, 2020
        ii. TOPIC: ((((((“development* delay” OR “cognit* delay” OR “language delay” OR “ling* delay” OR “mot* delay” OR “emotion* delay”) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)))
        iii. Number of search results: 17
    f. Scopus (Elsevier)
        i. Search conducted on November 15, 2020
        ii. TITLE-ABS-KEY(((((“development* delay” OR “cognit* delay” OR “language delay” OR “ling* delay” OR “mot* delay” OR “emotion* delay”) AND (child* OR infant*)) AND “rural China”) AND “early childhood”))
        iii. Number of search results: 13
2. Parenting: 110 in total
    a. PubMed
        i. Search conducted on November 15, 2020
        ii. ((((parenting OR caregiving) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)
        iii. Number of search results: 30
    b. EconPapers
        i. Search conducted on November 15, 2020
        ii. ((((parenting OR caregiving) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)
        iii. Number of search results: 25
    c. PsycARTICLES
        i. Search conducted on November 15, 2020

Emmers D, *et al*. *BMJ Global Health* 2021; 6:e005578. doi: 10.1136/bmjgh-2021-005578

ii. ((((parenting OR caregiving) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 25

d. Cochrane Library
i. Search conducted on November 15, 2020
ii. TITLE-ABS-KEY(((((parenting OR caregiving) AND (child* OR infant*)) AND “rural China”) AND “early childhood”))
iii. Number of search results: 4

e. Web of Science
i. Search conducted on November 15, 2020
ii. TOPIC: ((((((parenting OR caregiving) AND (child* OR infant*)) AND “rural China”) AND “early childhood”)))
iii. Number of search results: 13

f. Scopus (Elsevier)
i. Search conducted on November 15, 2020
ii. TITLE-ABS-KEY(((((parenting OR caregiving) AND (child* OR infant*)) AND “rural China”) AND “early childhood”))
iii. Number of search results: 13

3. Parenting interventions: 39 in total

a. PubMed
i. Search conducted on November 15, 2020
ii. (((((intervention* OR trial* OR RCT* OR experiment*) AND (child* OR infant*)) AND (stimulation OR nutrition)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 14

b. EconPapers
i. Search conducted on November 15, 2020
ii. (((((intervention* OR trial* OR RCT* OR experiment*) AND (child* OR infant*)) AND (stimulation OR nutrition)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 1

c. PsycARTICLES
i. Search conducted on November 15, 2020
ii. (((((intervention* OR trial* OR RCT* OR experiment*) AND (child* OR infant*)) AND (stimulation OR nutrition)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 4

d. Cochrane Library
i. Search conducted on November 15, 2020
ii. (((((intervention* OR trial* OR RCT* OR experiment*) AND (child* OR infant*)) AND (stimulation OR nutrition)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 3

e. Web of Science
i. Search conducted on November 15, 2020
ii. (((((intervention* OR trial* OR RCT* OR experiment*) AND (child* OR infant*)) AND (stimulation OR nutrition)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 12

f. Scopus (Elsevier)
i. Search conducted on November 15, 2020
ii. (((((intervention* OR trial* OR RCT* OR experiment*) AND (child* OR infant*)) AND (stimulation OR nutrition)) AND “rural China”) AND “early childhood”)
iii. Number of search results: 5

Search strings in Chinese:

4. Delays: 257 in total

a. 中国知网 （CNKI）
i. 高级检索：学术期刊+ 学位论文 + 特色期刊
ii. 表达式

1. 主题：发育 + 认知 + 语言 + 运动 + 情感
2. AND 主题：迟缓 + 缓慢 + 滞后
3. AND 主题：婴儿 + 婴幼儿
4. AND 主题：农村

iii. 结果：共 73 篇

b. 万方数据 （Wanfang）

i. 专业检索：期刊论文+ 学位论文

ii. 来源数据库： 万方

iii. 表达式：主题:(发育 + 认知 + 语言 + 运动 + 情感)*主题:(迟缓 + 缓慢 + 滞后 + 障碍)*主题:(婴儿 + 婴幼儿)*主题:(农村)

iv. 结果：89 条

c. 维普（VIP）

i. 检索式检索

ii. 表达式：R=(发育 OR 认知 OR 语言 OR 运动 OR 情感) AND R=(迟缓 OR 缓慢 OR 滞后) AND R=(婴儿 OR 婴幼儿) AND R=(农村)

iii. 结果：95 条

5. Parenting: 171 in total

a. 中国知网 （CNKI）

i. 高级检索：学术期刊 + 学位论文 + 特色期刊

ii. 表达式

1. 主题：养育 + 照顾 + 照料
2. AND 主题：婴儿 + 婴幼儿
3. AND 主题：农村

iii. 结果：共 68 篇

b. 万方数据 （Wanfang）

i. 专业检索：期刊论文 + 学位论文

ii. 来源数据库： 万方

iii. 表达式：主题:(养育 + 照顾 + 照料)*主题:(婴儿 + 婴幼儿)*主题:(农村)

iv. 结果：66 条

c. 维普（VIP）

i. 检索式检索

ii. 表达式：R=(养育 OR 照顾 OR 照料) AND R=(婴儿 OR 婴幼儿) AND R=(农村)

iii. 结果：37 条

6. Parenting interventions: 17 in total

a. 中国知网 （CNKI）

i. 高级检索：学术期刊 + 学位论文 + 特色期刊

ii. 表达式

1. 主题：干预 + 试验 + 随机对照试验 + 实验
2. AND 主题：婴儿 + 婴幼儿
3. AND 主题：养育 + 照顾 + 照料
4. AND 主题：农村

iii. 结果：6 条

b. 万方数据 （Wanfang）

i. 专业检索：期刊论文 + 学位论文

ii. 来源数据库： 万方

iii. 表达式：主题:(干预 + 试验 + 随机对照试验 + 实验)*主题:(婴儿 + 婴幼儿)*主题:(养育 + 照顾 + 照料)*主题:(农村)

iv. 结果：11 条

c. 维普（VIP）

i. 检索式检索

ii. 表达式：R=(干预 OR 试验 OR 随机对照试验 OR 实验) AND R=(婴儿 OR 婴幼儿) AND R=(养育 OR 照顾 OR 照料) AND R=(农村)

iii. 结果：0 条

Appendix 2: Prevalence of Developmental Delay

1. Data Extraction and Processing

Prevalence was extracted and processed in three ways:

1. If the reviewed study reported the prevalence of developmental delay, defined as a developmental score of more than 1 SD below the healthy population mean, we extracted the percentage of delay reported in the study; or

2. If the reviewed study reported the percentage of delay, but this percentage was defined based on a different cutoff score, we imposed the assumptions of a normal distribution to derive the percentage of children with a developmental score of more than 1 SD below the healthy population mean. For example, if the study defined developmental delay as a score of at least 2 SD below the mean of a healthy populations, we derived the percentage of children with a score of at least 1 SD below this mean, assuming that the developmental scores of the sample population follow a normal distribution; or

3. If the mean and SD of development scores were reported instead of the percentage of delay, we calculated the percentage of delay based on published norms of a representative reference population in combination with the assumption that the developmental scores of the sample population follow a normal distribution.

2. Measurement of Prevalence of Developmental Delay Based on BSID-III Scores

2.1. Normative sample

The ASQ-3, ASQ:SE, BSID-I, and BSID-II have been validated and standardized based on an urban Chinese normative sample.[25–29] No normative Chinese sample for the BSID-III scores, however, has been established. Therefore, the raw score conversion charts included in the original US BSID-III administration manual for each age category are used to compute BSID-III composite scores in the reviewed studies.[30]

Hua et al.[31] assessed BSID-III (composite and raw) cognition scores for over 1,500 urban Chinese children. As a robustness check, we can alternatively use the distribution of this urban Chinese sample to norm the BSID-III cognition scores. Hua et al. report the mean (SD) of BSID-III cognition scores for seven separate age groups. Disregarding the limited number of age groups, which limits the accuracy with which we can compute China-specific age-adjusted normed scores, Appendix Table 2.1 shows that using China-specific norms to define cognitive delay does not lead to a systematic increase or decline in reported prevalences of delay. Note that, in absence of a study that establishes the mean (SD) of BSID-III language and motor composite scores for a large-scale urban Chinese sample, we are unable to conduct the same robustness check for BSID-III language and motor delays.

2.2. Cutoff scores used to define developmental delay

Note that the BSID was designed to have a mean composite score of 100 with a SD of 15 for a healthy population.[30] Empirical evidence from population-based random samples globally finds higher means and lower SDs for the cognition, language, and motor scales of the third edition of the BSID (BSID-III).[25,32,33] The literature shows that, when using the traditional cutoff score of 85, the prevalence of moderate developmental delay is underestimated.[32,34] In absence of a normative Chinese sample, BSID-III cutoff scores are defined based on the means (SDs) of BSID-III composite scores reported in the international literature (i.e., mean (SD) is expected to be 105 (9.6), 109 (12.3), and 100 (15) for the cognitive, language, and social-emotional scales, respectively.[30,32,35,36] Finally, note that the recent study of Hua et al.[31] reported a mean (SD) of 102 (6.9) for BSID-III composite cognition scores of urban Chinese children. When we use this cutoff score instead of the cutoff score extracted from the global literature, the cutoff score remains nearly identical to cutoff scores reported in the reviewed international literature (i.e., 95.4 vs. 95.1).

Emmers D, *et al. BMJ Global Health* 2021; 6:e005578. doi: 10.1136/bmjgh-2021-005578

Appendix Table 2.1: Prevalence of Cognitive Delay based on US vs. China BSID-III Norms

No.	Author(s) (Year)	Child ages (months)	Mean (SD) of BSID-III composite cognition scores (based on US norms)	Mean (SD) of BSID-III raw cognition scores	Prevalence of cognitive delay	
					US norms[a]	China norms[b]
1	Luo et al. (2019a)[8]	6–18	95.7 (12.9)	38.6 (8.6)	57%	66%
2	Emmers et al. (2020)[9]	6–18	100.5 (11.9)	43.5 (7.7)	41%	43%
3	Wang et al. (2020a)[10]	6–24	95.9 (12.6)	45.5 (11.2)	54%	38%
4	Qian et al. (2020)[11]	6–18	96.7 (13.4)	40.9 (10.3)	50%	55%
5	Wang et al. (2019)[14]	6–30	96.5 (13.3)	46.2 (11.6)	47%	36%
6	Zhong et al. (2019)[15]	1–23	94.5 (16.2)	43.6 (11.2)	56%	45%
7	Wang et al. (2020b)[18]	6–24	96.8 (13.1)	45.9 (11.2)	52%	37%
8	Emmers & Wang (2020)[19]	6–36	101.4 (14.9)	55.6 (14.2)	39%	47%

Note. a. US norms are defined based on the normative US sample.[30] If mean (SD) of the BSID-III raw cognition scores was not reported in the paper, we derived the raw scores using the conversion charts included in the original BSID-III administration manual for each age category.[30] If the sample covered multiple age categories, the median sample age was used. b. China norms are defined based on the distribution of Bayley-III raw cognition scores reported in Hua et al.[31] by age category. If the sample covered multiple age categories, the median sample age was used.

Appendix 3: Risk of Bias Assessments

Appendix Table 3.1: Risk of Bias Assessment for Prevalence Studies (Hoy et al., 2012)

Study details		External validity* (No = high risk; Yes = low risk)					Internal validity* (No = high risk; Yes = low risk)					Overall
Author(s)	**Year**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11 (Overall risk of bias)**
Jin et al.	2007[1]	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	Moderate
Ma et al.	2008[2]	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	Low
Yang et al.	2019[3]	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	Low
Yue et al.	2017[4]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Low
Zhou et al.	2019a[5]	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Low
Zhang et al.	2018[6]	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Low
Wei et al.	2018[7]	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Low
Luo et al.	2019a[8]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low
Emmers et al.	2020[9]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Low
Wang et al.	2020a[10]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Low
Qian et al.	2020[11]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Low
Li et al.	2018[12]	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Low
Zhou et al.	2019b[13]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Low
Wang et al.	2019[14]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Low
Zhong et al.	2019[15]	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	Low
Bai et al.	2020[16]	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Low
Tan et al.	2020[17]	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Low
Wang et al.	2020b[18]	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	Low
Emmers & Wang	2020[19]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Low

Note. *Items of Hoy et al.'s (2012) risk of bias tool:
1. Was the study's target population a close representation of the rural population in relation to relevant variables, e.g., age, sex, occupation?
2. Was the sampling frame a true or close representation of the target population?
3. Was some form of random selection used to select the sample, OR was a census undertaken?
4. Was the likelihood of non-response bias minimal?
5. Were data collected directly from the subjects (as opposed to a proxy)?
6. Was an acceptable case definition used in the study?
7. Was the study instrument that measured the parameter of interest (e.g., prevalence of low back pain) shown to have reliability and validity (if necessary)?
8. Was the same mode of data collection used for all subjects?
9. Was the length of the shortest prevalence period for the parameter of interest appropriate?
10. Were the numerator(s) and denominator(s) for the parameter of interest appropriate?

Appendix Table 3.2: Risk of Bias Assessment of Intervention Studies (using EPHPP tool)

Author(s)	Year	A. Selection bias rating	B. Study design rating	C. Confounders rating	D. Blinding rating	E. Data collection methods rating	F. Withdrawals and dropout rating	Global rating
Jin et al.	2007[1]	3	3	3	1	3	3	3
Sylvia et al.	2020[20]	3	3	3	2	3	3	3
Wang et al.	2020c[21]	3	3	3	2	3	3	3
Heckman et al.	2020[22]	3	3	3	2	3	3	3
Luo et al.	2019b[23]	3	3	3	3	3	3	3
Emmers et al.	2020[9]	3	3	3	3	3	2	3
Zhong et al.	2020[24]	3	3	3	2	3	2	3
Qian et al.	2020[11]	3	3	3	2	3	2	3
Li et al.	2018[12]	2	3	3	2	3	1	2
Bai et al.	2020[16]	3	3	3	3	3	3	3

Note. The Effective Public Health Practice Project (EPHPP) quality assessment tool for quantitative studies is a standardized tool to assess the quality of intervention studies. The tool evaluates study quality based on the domains of selection bias, study design, confounders, blinding, data collection methods, withdrawals, and dropouts. The tool can be used to evaluate a variety of intervention studies, including RCTs and case-control studies. According to the grading guidelines, the study obtains 1, 2, and 3 points, respectively, if the study is rates as weak, moderate, or strong for a specific domain. Domain scores are averaged to compute the global rating score. The maximum global rating score is 3. Based on the global score, studies are assigned a global quality rating of weak (1–1.50), moderate (1.51–2.50), or strong (2.51–3).

Appendix 4: GRADE Summary of Findings

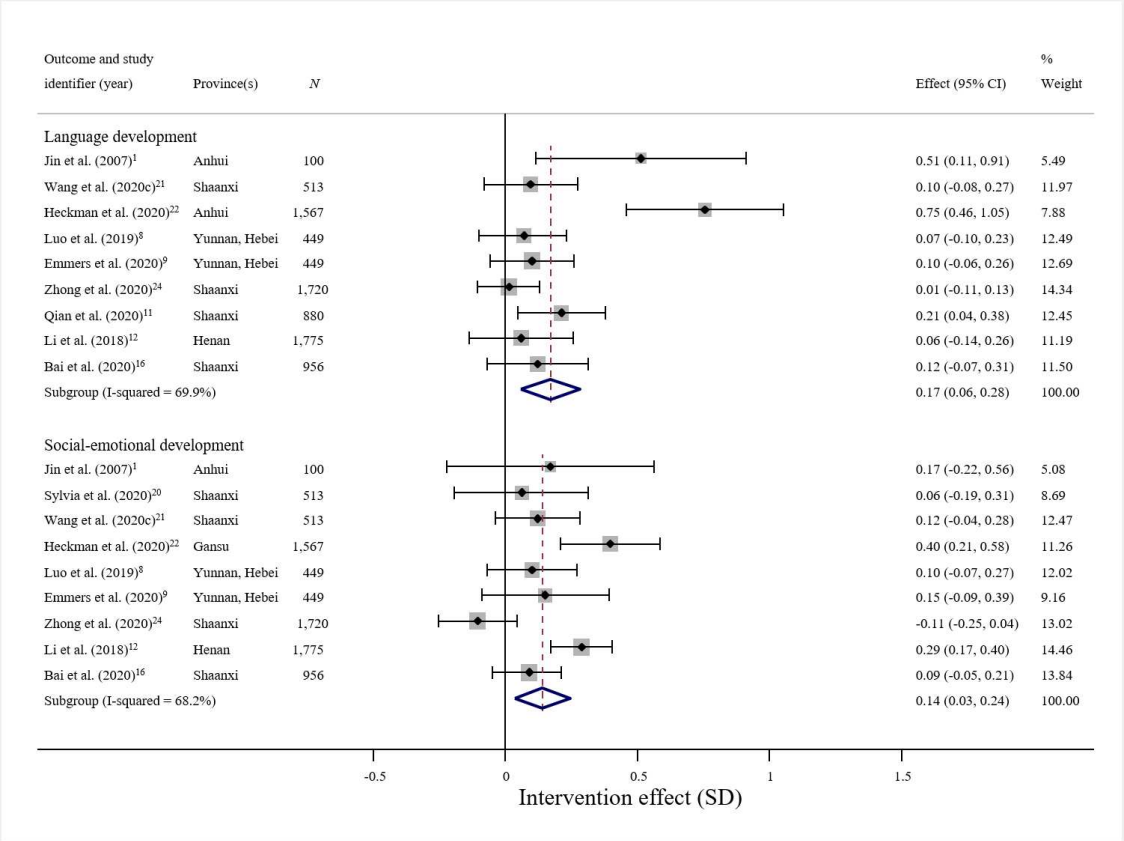
Appendix Table 4.1: GRADE Summary of Findings on Effects of Parental Training Programs on ECD Outcomes and Parenting Practices and Beliefs

	Certainty assessment							Summary of findings		
	Participants (no. of studies)	**Study design**	**Risk of bias**	**Inc.**	**Ind.**	**Imp.**	**Pub. bias**	**Impact**	**Indicative effect sizes[a], SMD (95% CI)**	**Certainty**
Primary outcomes										
Cognitive development	7,960 (10)	RCT	No	S(-1)	No	No	No	Positive	0.26 (0.18 to 0.35)	We have high certainty that parental training programs improve early cognition outcomes of 0- to 5-year-olds in rural areas of China[b]
Language development	7,960 (9)	RCT	No	S(-1)	No	No	No	Positive	0.17 (0.06 to 0.28)	We have moderate certainty that parental training programs improve early language outcomes of 0- to 5-year-olds in rural areas of China[c]
Social-emotional development	7,080 (9)	RCT	No	S(-1)	No	No	No	Positive	0.14 (0.03 to 0.24)	We have moderate certainty that parental training programs improve early social-emotional outcomes of 0- to 5-year-olds in rural areas of China[*d]
Secondary outcomes										
Parenting practices	4,518 (7)	RCT	No	S(-1)	S(-1)	No	No	Positive	0.39 (0.24 to 0.54)	We have moderate certainty that parental training programs improve early cognition outcomes of 0- to 5-year-olds in rural areas of China[e]
Parenting knowledge and beliefs	3,562 (6)	RCT	No	S(-1)	S(-1)	No	No	Positive	0.20 (0.11 to 0.28)	We have moderate certainty that parental training programs improve early cognition outcomes of 0- to 5-year-olds in rural areas of China[f]

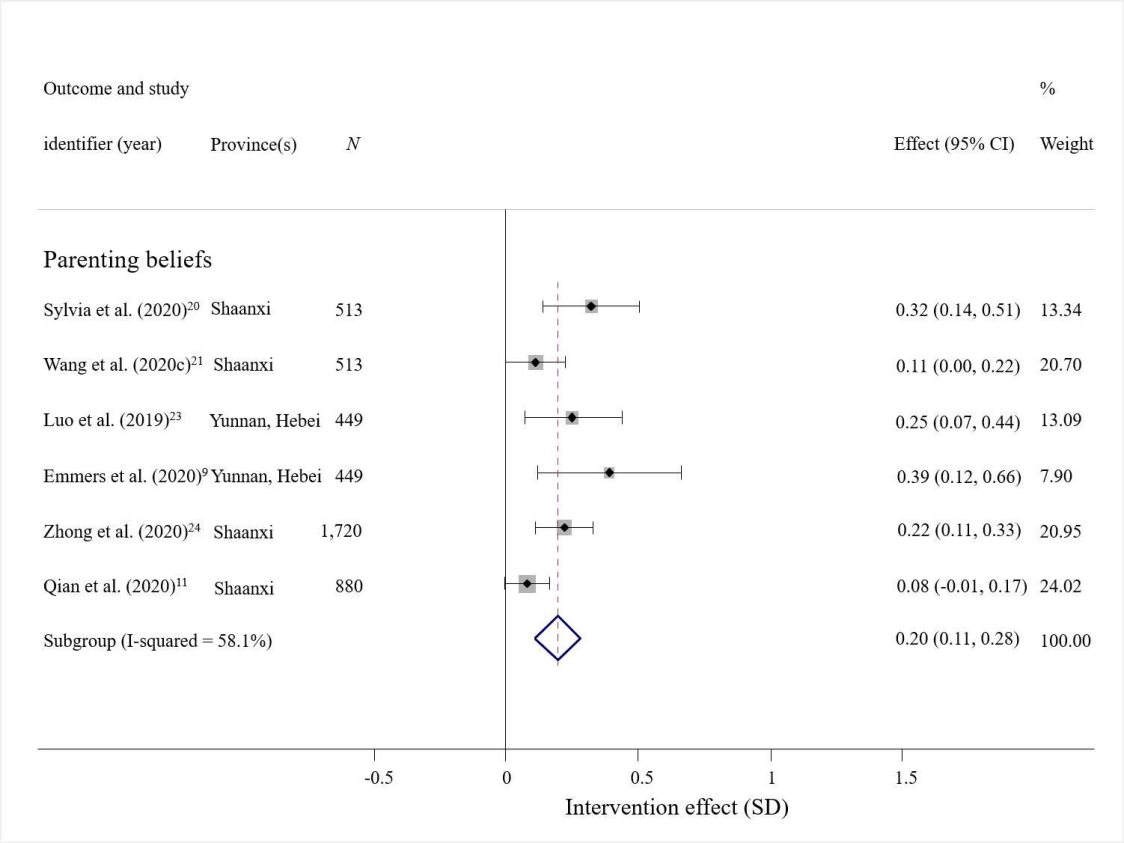
Note. GRADE: Grading of Recommendations Assessment, Development, and Evaluation; ECD: Early Childhood Development; SMD: standardized mean difference; S(-1): serious risk, downgrade by one point. Inc.: Inconsistency; Ind.: Indirectness; Imp,: Imprecision.
a. From meta-analysis; b. Rating down for inconsistency in outcome measures of parenting practices, up for dose-response estimates; c. Rating down for inconsistency in outcome measures of parenting practices; d. Rating down for inconsistency in outcome measures of parenting practices; e. Rating down for inconsistency and indirectness in outcome measures of parenting practices, up for dose-response estimation; f. Rating down for inconsistency and indirectness in outcome measures of parenting beliefs, up for dose-response estimation.

Appendix 5: Intervention Effects on Outcomes of Child Development and Parental Beliefs

Appendix Figure 5.1: Intervention Effects on Early Language and Social-Emotional Development



Appendix Figure 5.2: Intervention Effects on Parental Beliefs



Appendix 6: Heterogeneity in Intervention Effects

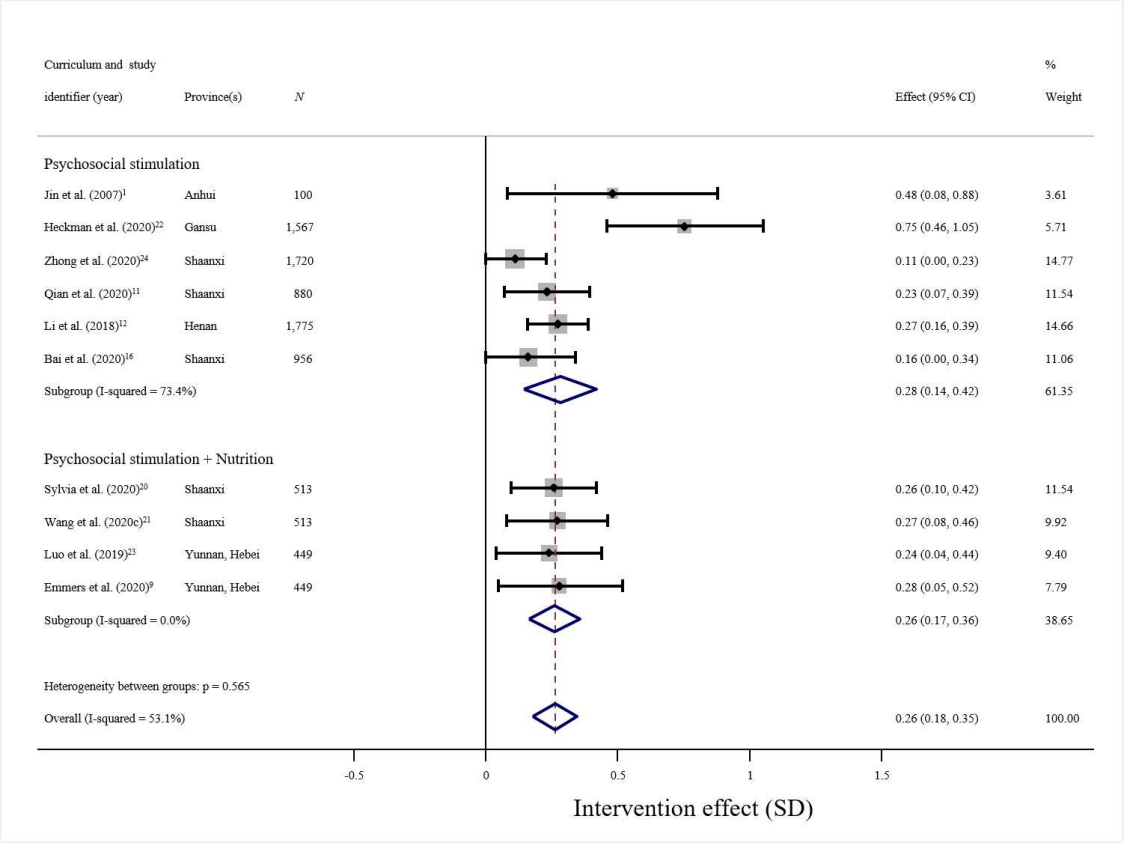
We conduct a heterogeneity analysis of treatment impacts by program curricula and delivery mode, measurement tool, and geographic location. Figures 6.1–6.18 in this Appendix document show the results of the heterogeneity testing. Figures 6.1–6.3 and Figures 6.4–6.6 present the heterogeneity in treatment impacts across studies with different program curricula and delivery modes, respectively. First, we find no evidence of systematic differences in treatment impacts across single-component psychosocial stimulation programs and integrated psychosocial stimulation and nutrition programs (*p*-values of between-group heterogeneity > 0.10; see Figures 6.1–6.3). Second, Figures 6.4–6.6 show that center-based programs have, on average, an impact of 0.19 SD [95% CI: 0.11–0.27 SD], 0.09 SD [95% CI: -0.005–0.18 SD], and 0.09 SD [95% CI: -0.13–0.32 SD on early cognitive, language, and social-emotional development, respectively. In contrast, home-based programs have, on average, an impact of 0.34 SD [95% CI: 0.21–0.48 SD], 0.27 SD [95% CI: 0.05–0.48 SD], and 0.17 SD [95% CI: 0.06–0.28 SD on child cognitive, language, and social-emotional development, respectively. We observe that treatment impact heterogeneity between center-based and home-based programs is significant at the 5 percent level for cognitive development but not for language or social-emotional development.

Figures 6.7–6.12 present heterogeneity in the prevalence of delay and the size of estimated treatment impacts on child development outcomes by type of ECD measurement tool. The results of our measurement-based heterogeneity analysis indicate that the point estimates of the prevalence of cognition and language delay are lower in studies in which the ASQ-3 test is used than in studies in which the BSID is administered (37.0% [95% CI: 24.0–50.1%] vs. 48.2% [95% CI: 45.3–51.1%] and 35.8% [95% CI: 25.2–46.3%] vs. 52.1% [95% CI: 47.1–58.3%]; see Figures 6.7 and 6.8 in Appendix 6, respectively). We do not detect a meaningful difference in the observed prevalence of social-emotional delay when using ASQ-3 tests, ASQ:SE tests, or the BSID (40.5% [95% CI: 29.7–51.3%] vs. 32.3% [95% CI: 23.1–41.4%] vs. 36.1% [95% CI: 27.4–44.8%], see Figure 6.9). Further, Figures 6.10–6.12 show that the parenting interventions have, on average, an impact of 0.20 SD [95% C:I 0.12–0.27 SD], 0.09 SD [95% CI: 0.00–0.17 SD], and 0.03 SD [95% CI: -0.09–0.16 SD] on BSID cognition, language, and social emotional scores, respectively. Note that only one of the reviewed impact evaluations used ASQ-3 scores to measure ECD outcomes.

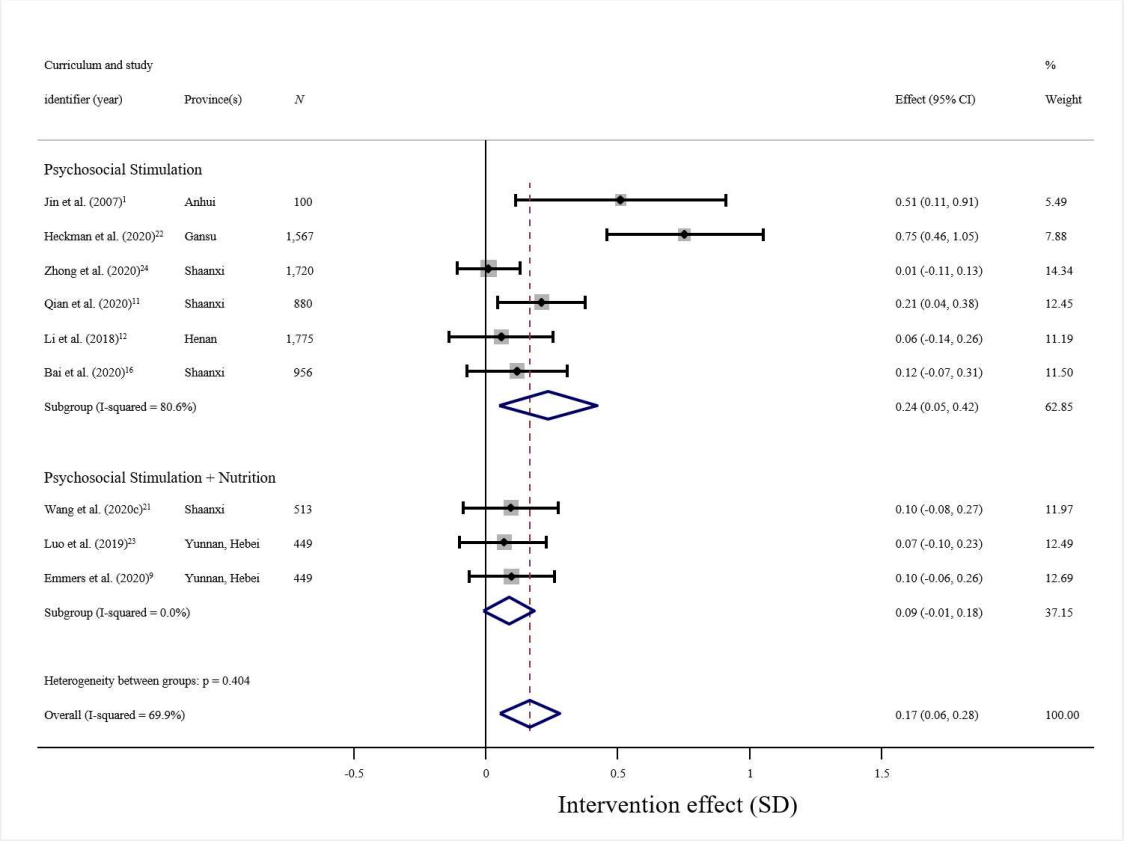
Finally, Figures 6.13–6.18 show the heterogeneity in the prevalence of delay and the size of estimated treatment impacts on child development outcomes by geographic location of the study. The reviewed studies span 14 provinces across Eastern, Central, and Western China. By grouping studies by geographic location, we find that the risk of early cognitive and language delay is higher in Western China than in Eastern and Central China (see Figures 6.13 and 6.14), but treatment impacts on child cognitive and language development do not differ significantly across these regions (*p*-values of between-group heterogeneity > 0.10; see Figures 6.16 and 6.17). Evidence from two studies conducted in Eastern China indicates that the prevalence of social-emotional delay and treatment impacts on social-emotional child outcomes in Eastern China are lower than in Central and Western China (see Figures 6.15 and 6.18). The evidence in Figures 6.15 and 6.18 suggests that the risk of social-emotional delay and the potential role for parenting interventions may be smaller in Eastern China. We acknowledge, however, that we are unable to draw conclusions with regard to geographic variation in our results due to data limitations of this study.

1. Heterogeneity by Program Curriculum

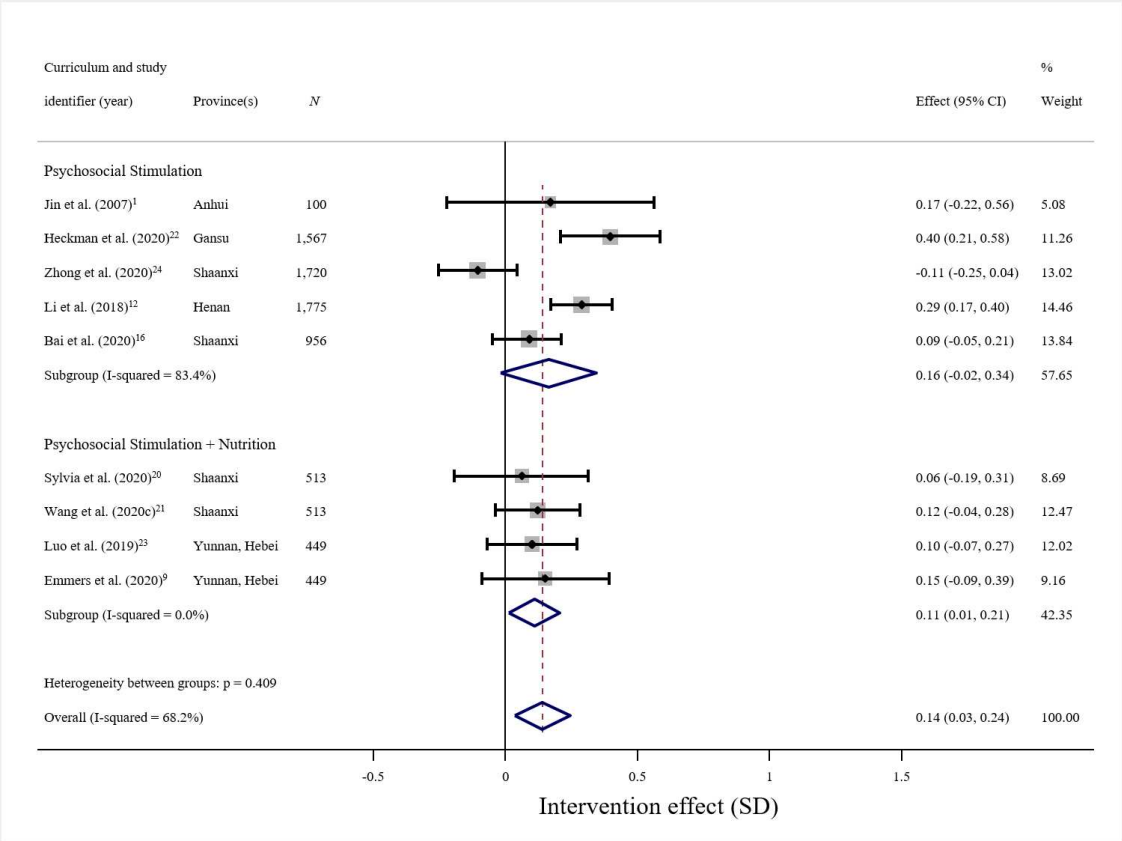
Appendix Figure 6.1: Pooled Intervention Effects on Early Cognitive Development by Program Curriculum



Appendix Figure 6.2: Pooled Intervention Effects on Early Language Development by Program Curriculum

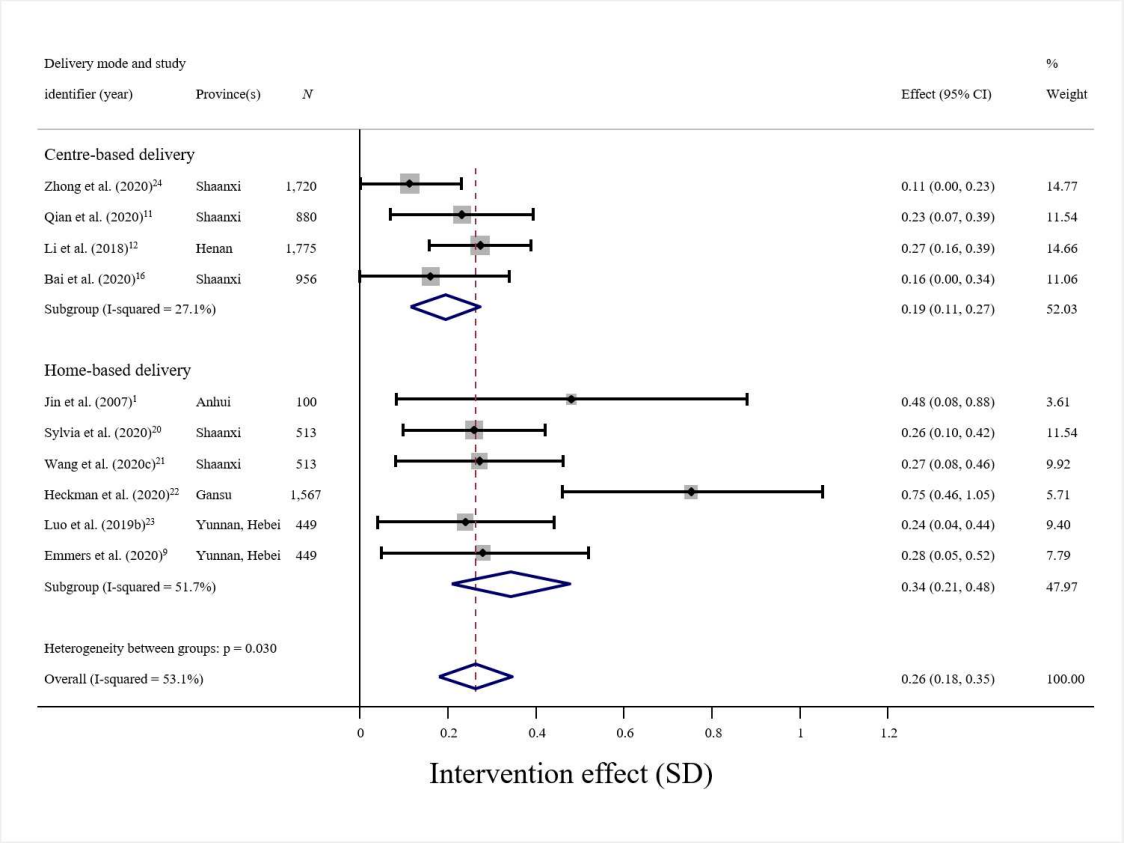


Appendix Figure 6.3: Pooled Intervention Effects on Early Social-Emotional Development by Program Curriculum

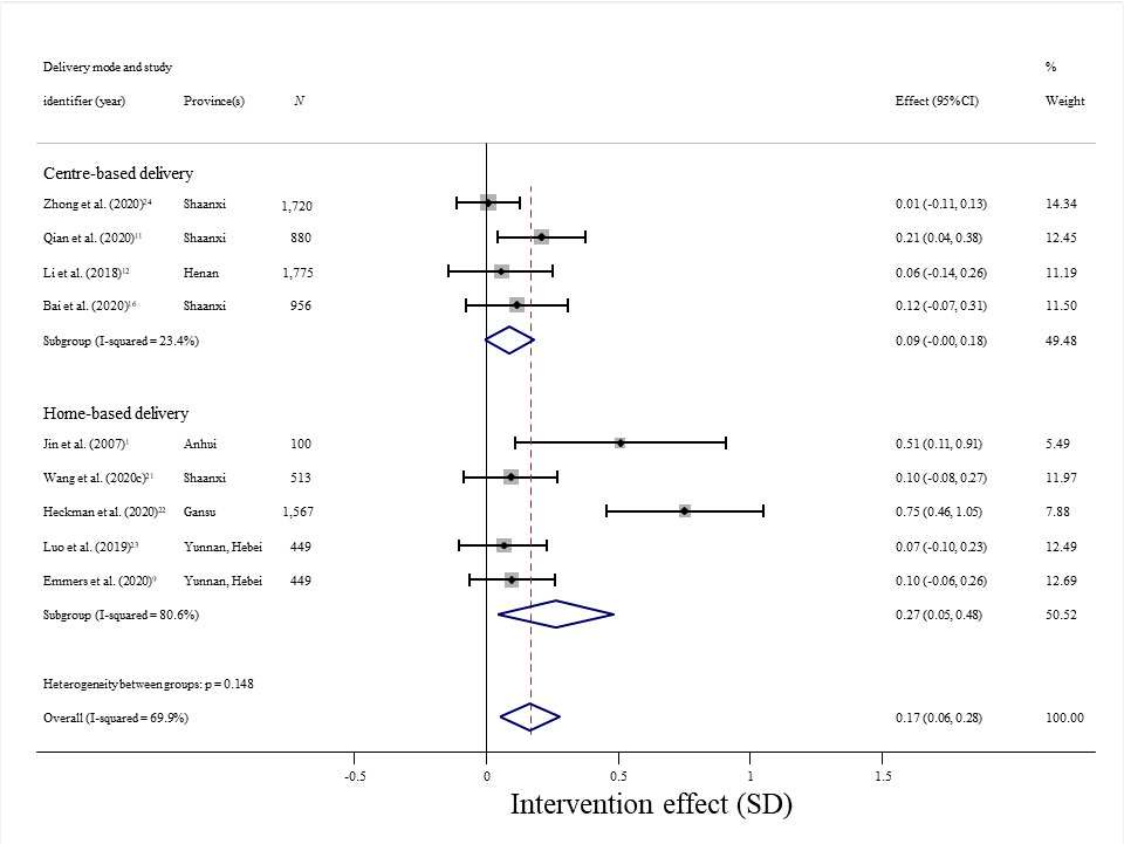


2. Heterogeneity by Delivery Mode

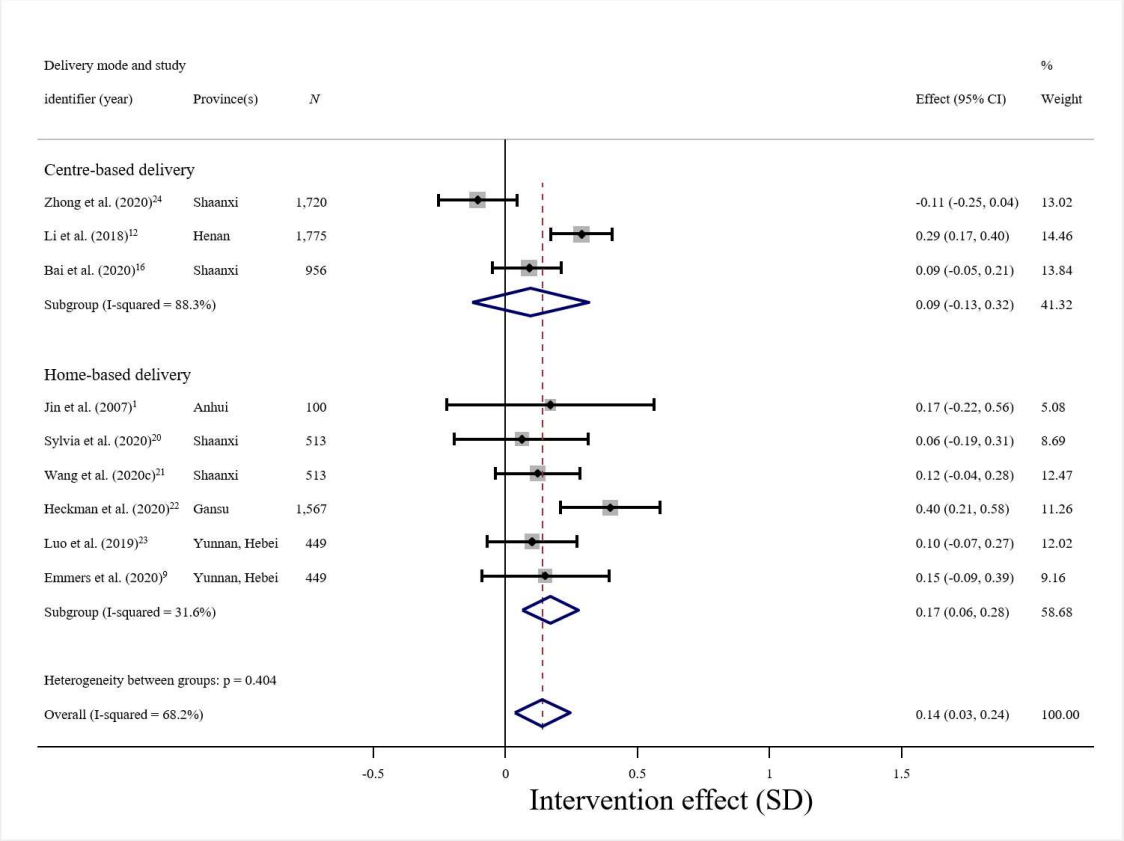
Appendix Figure 6.4: Pooled Intervention Effects on Early Cognitive Development by Delivery Mode



Appendix Figure 6.5: Pooled Intervention Effects on Early Language Development by Delivery Mode

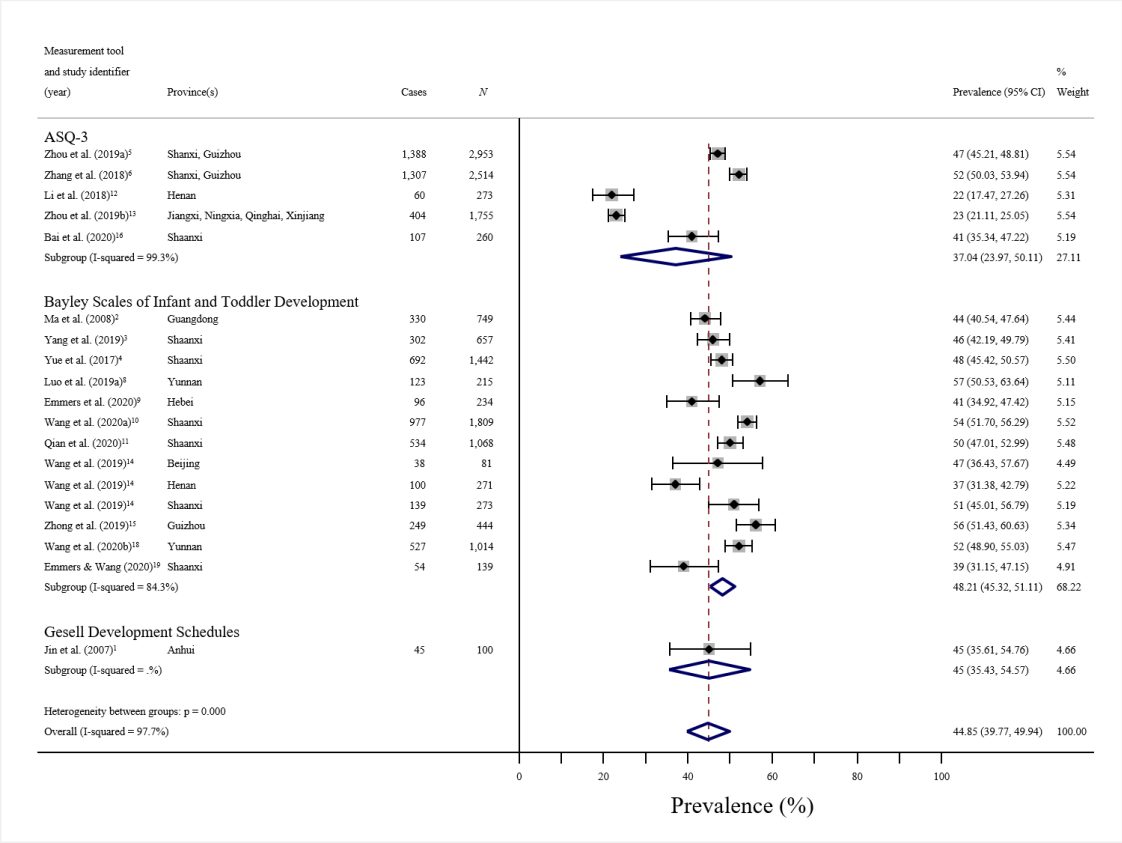


Appendix Figure 6.6: Pooled Intervention Effects on Early Social-Emotional Development by Delivery Mode

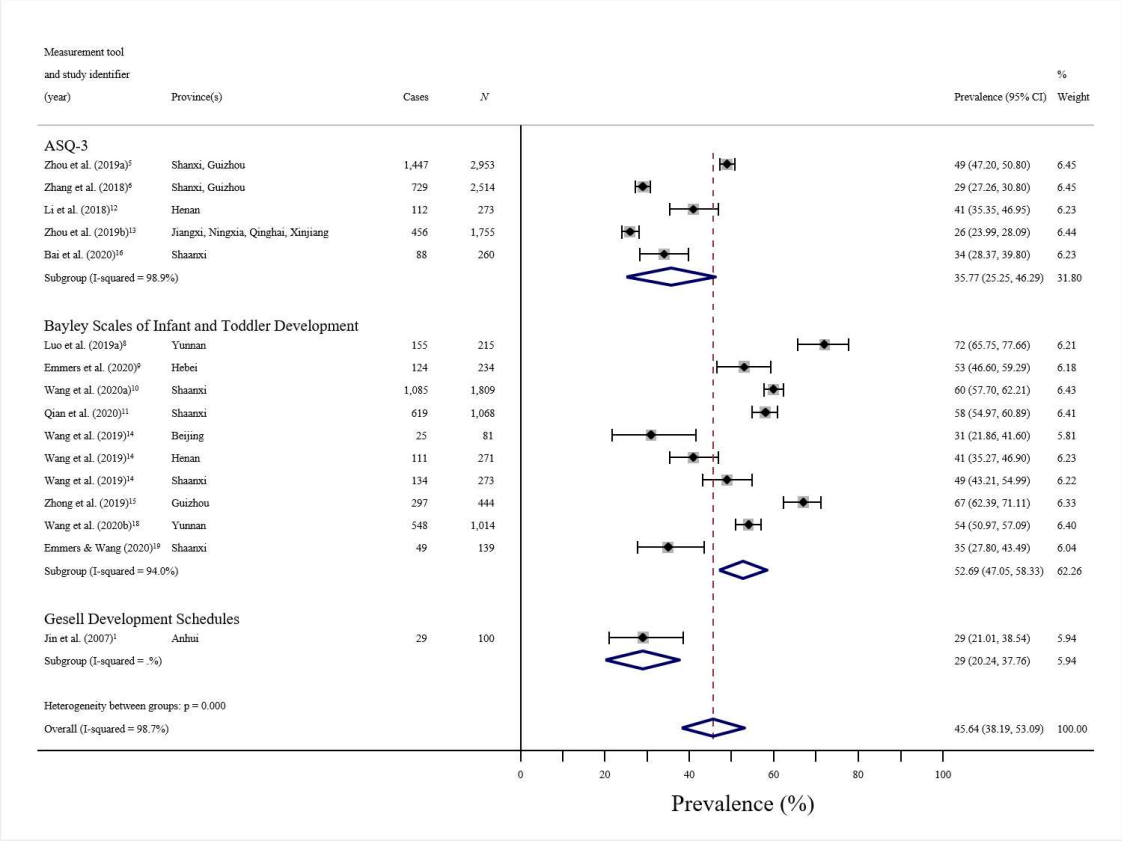


3. Heterogeneity by Measurement Tool

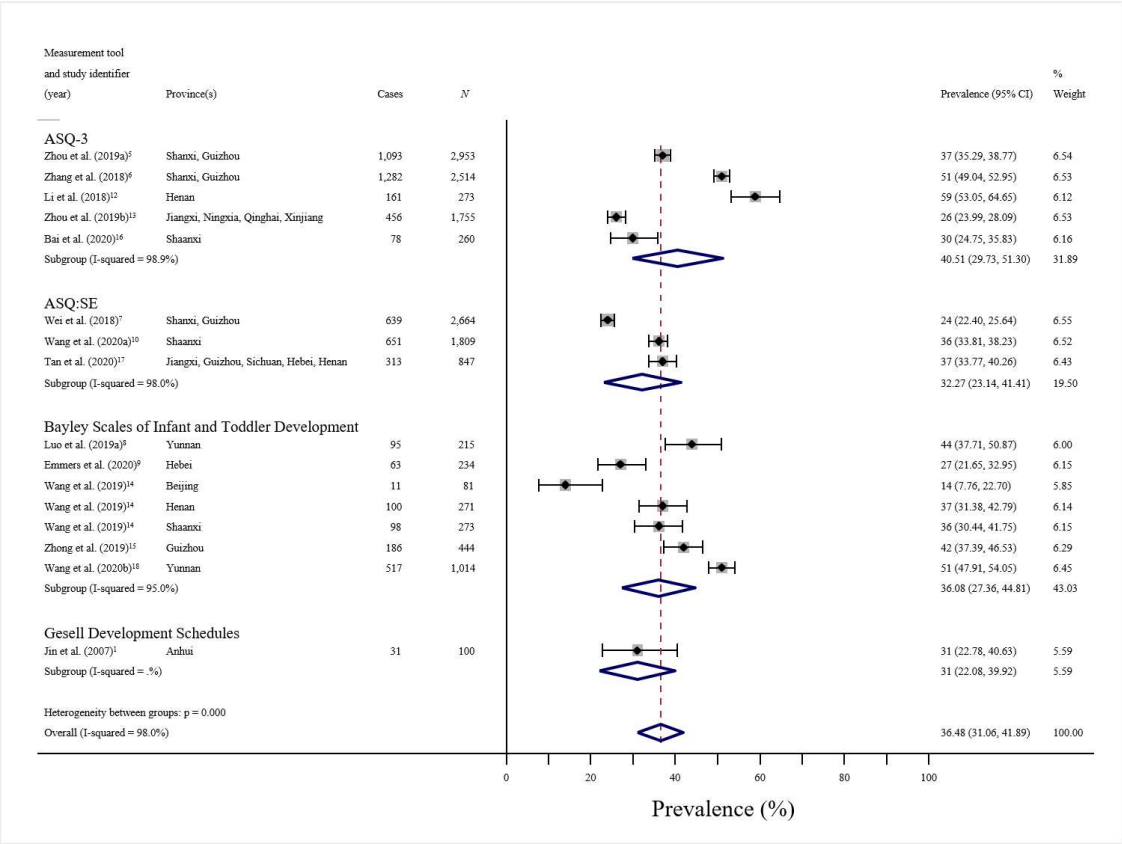
Appendix Figure 6.7: Pooled Prevalence of Early Cognitive Delay by Measurement Tool



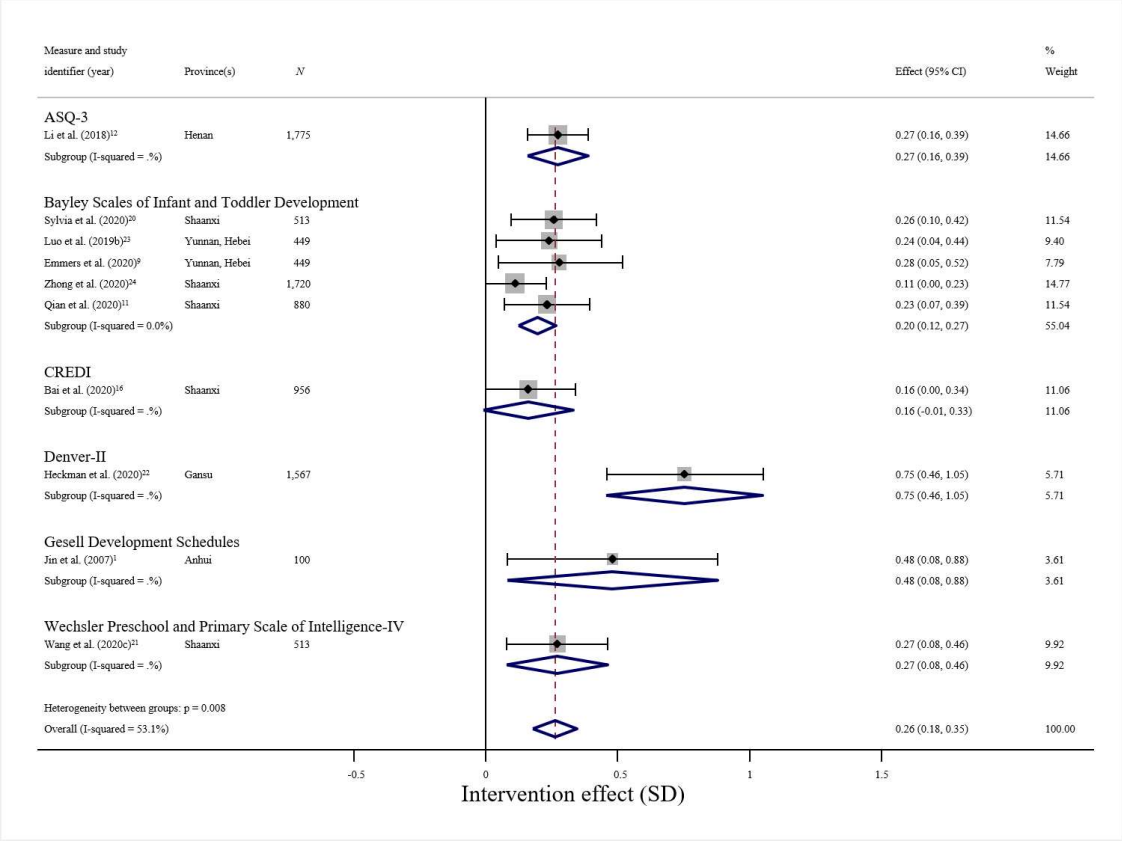
Appendix Figure 6.8: Pooled Prevalence of Early Language Delay by Measurement Tool



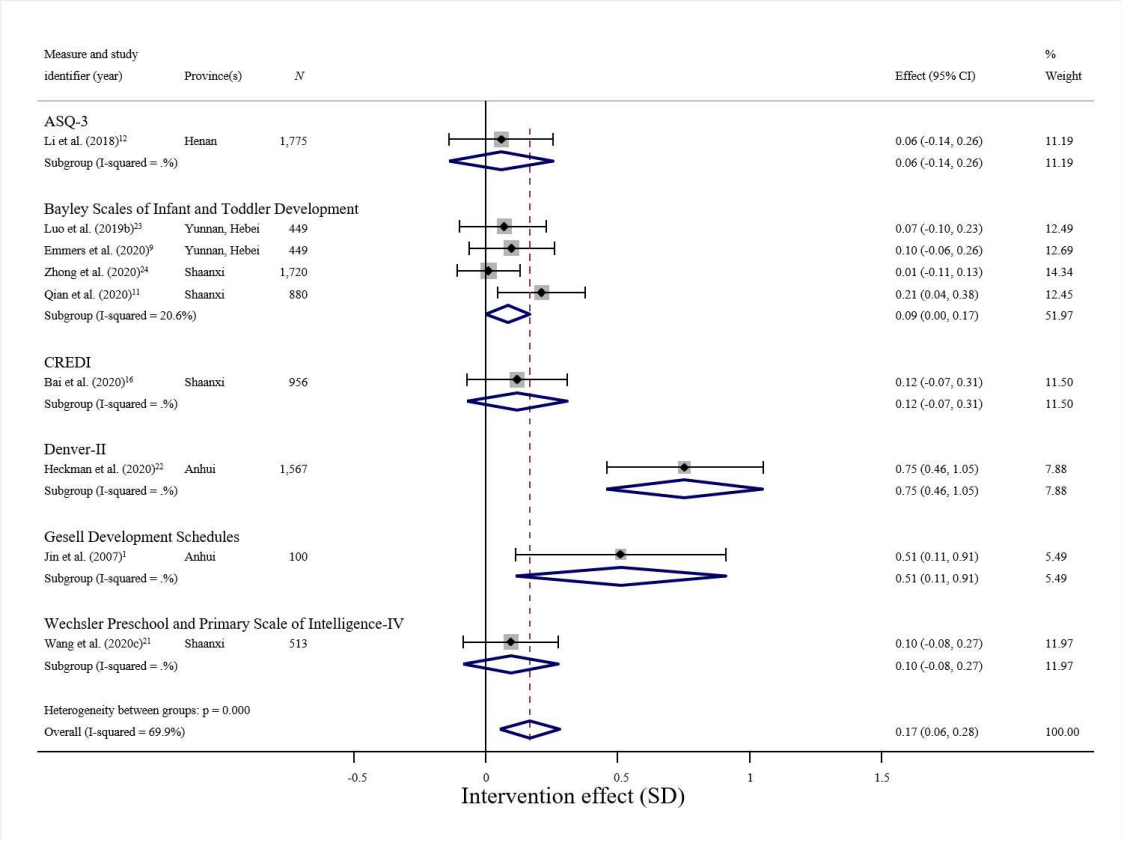
Appendix Figure 6.9: Pooled Prevalence of Early Social-Emotional Delay by Measurement Tool



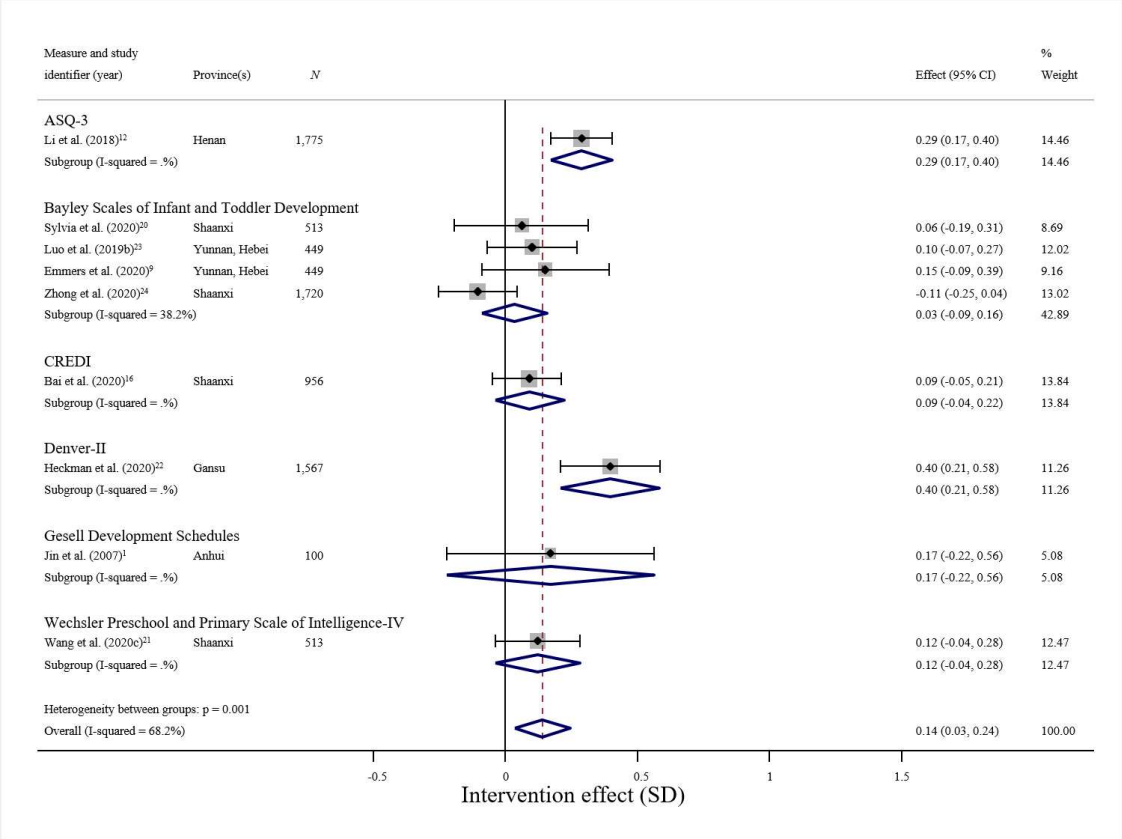
Appendix Figure 6.10: Pooled Intervention Effects on Early Cognitive Development by Measurement Tool



Appendix Figure 6.11: Pooled Intervention Effects on Early Language Development by Measurement Tool

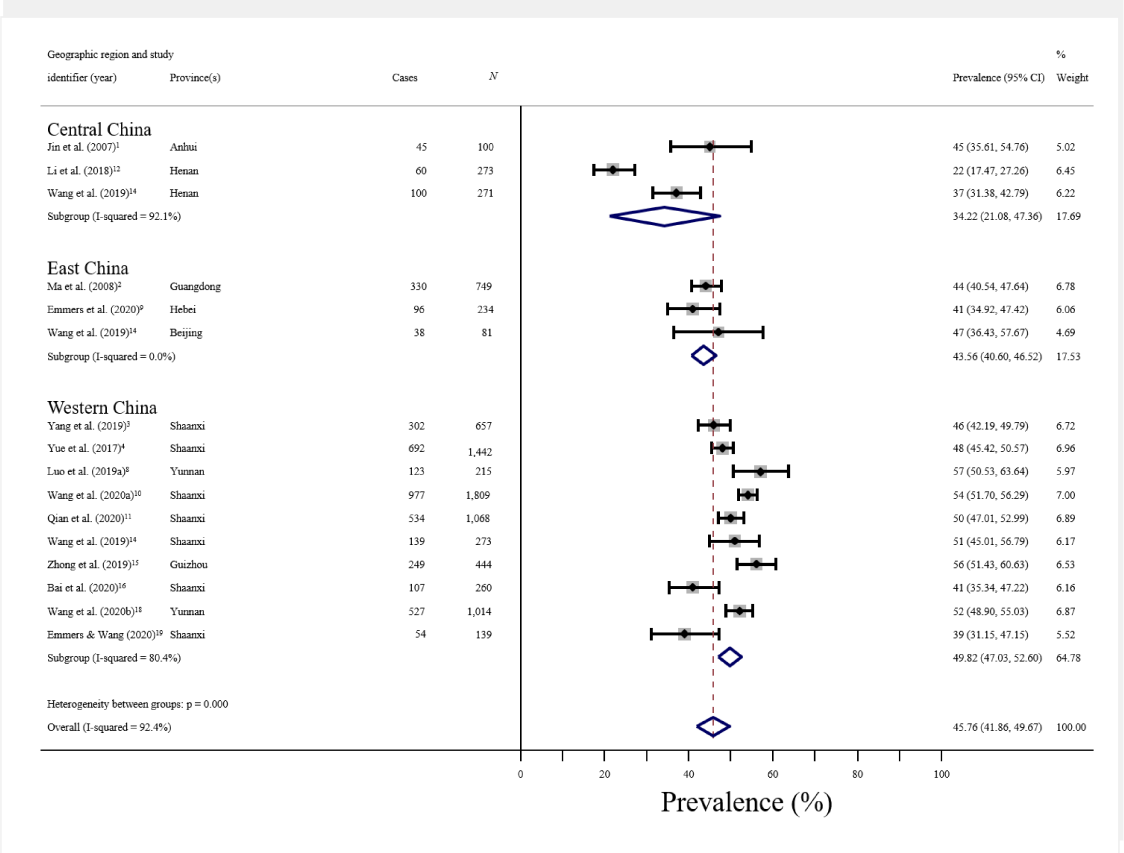


Appendix Figure 6.12: Pooled Intervention Effects on Early Social-Emotional Development by Measurement Tool

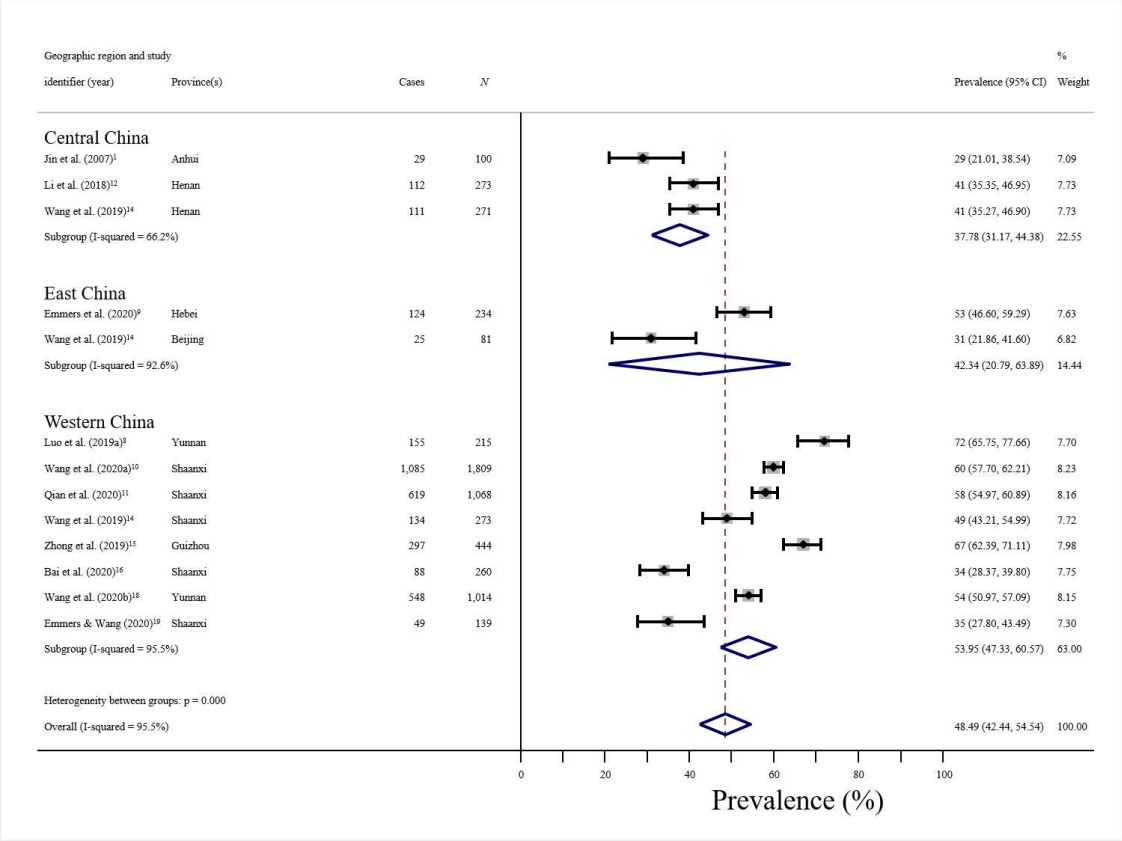


4. Heterogeneity by Geographic Region

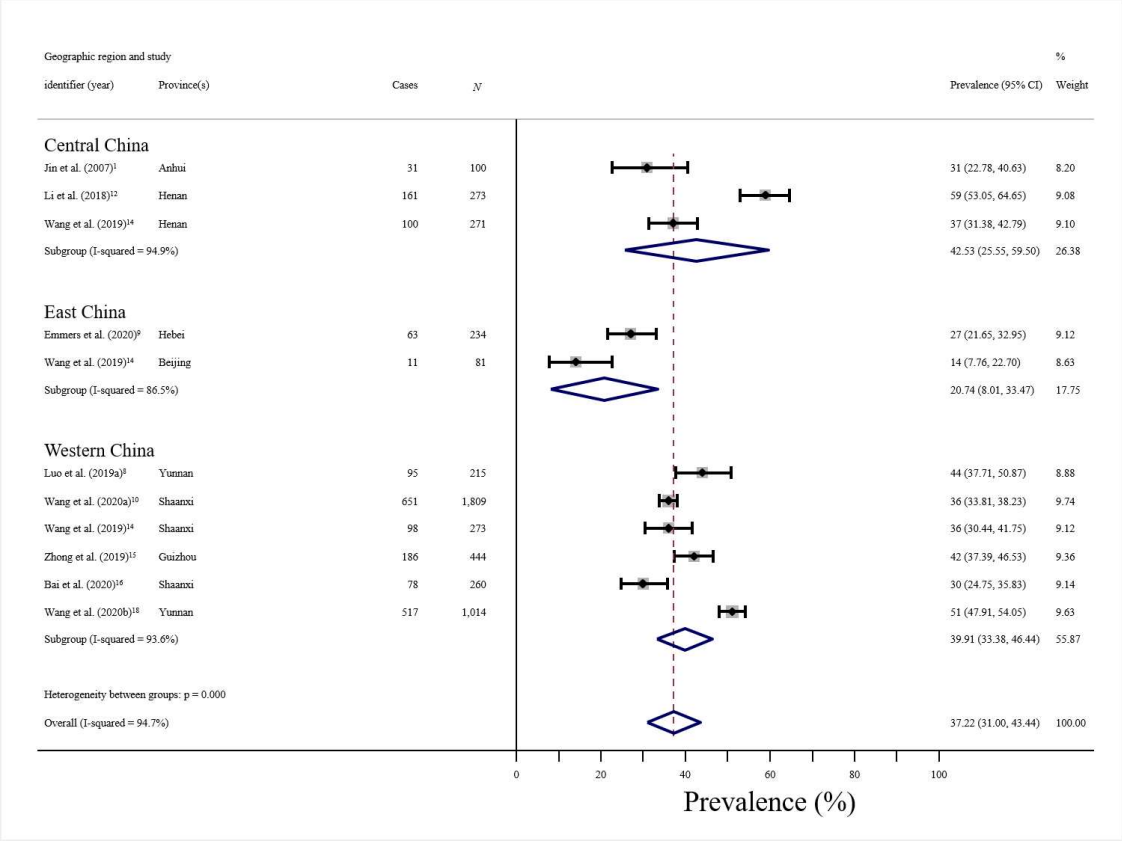
Appendix Figure 6.13: Pooled Prevalence of Early Cognitive Delay by Geographic Region



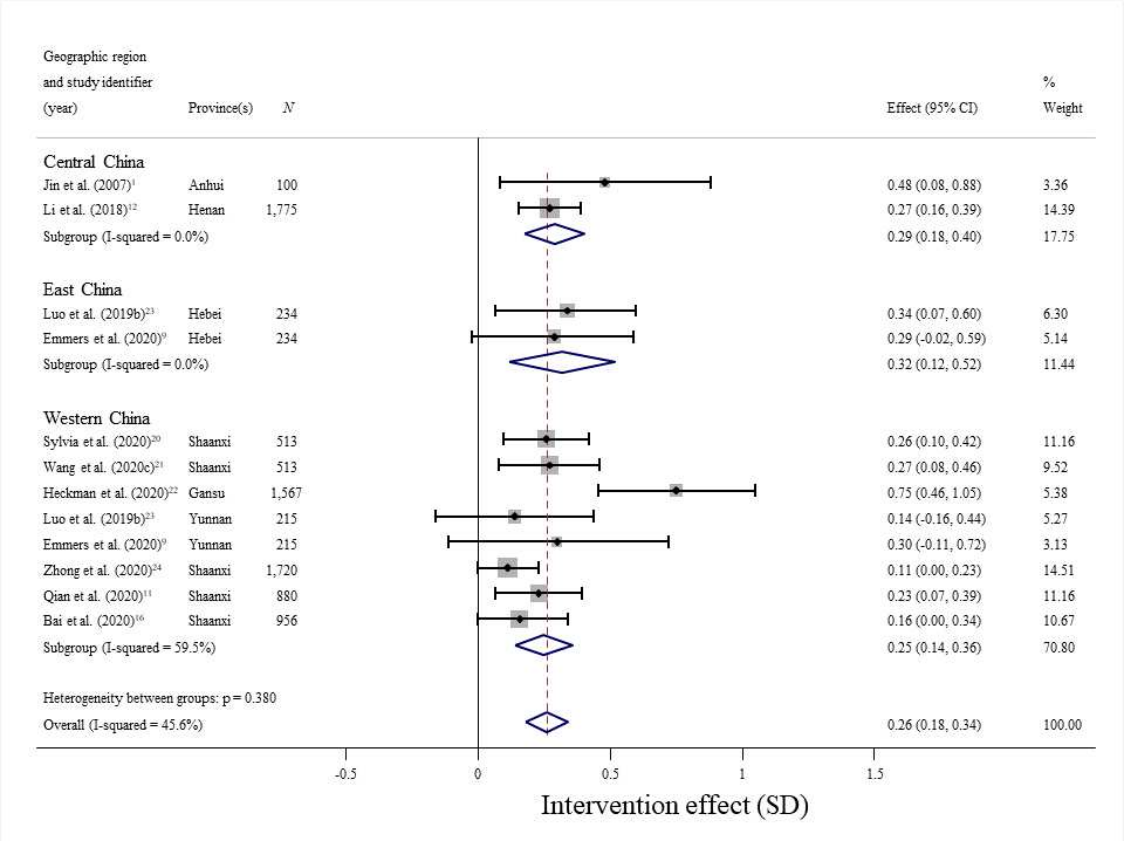
Appendix Figure 6.14: Pooled Prevalence of Early Language Delay by Geographic Region



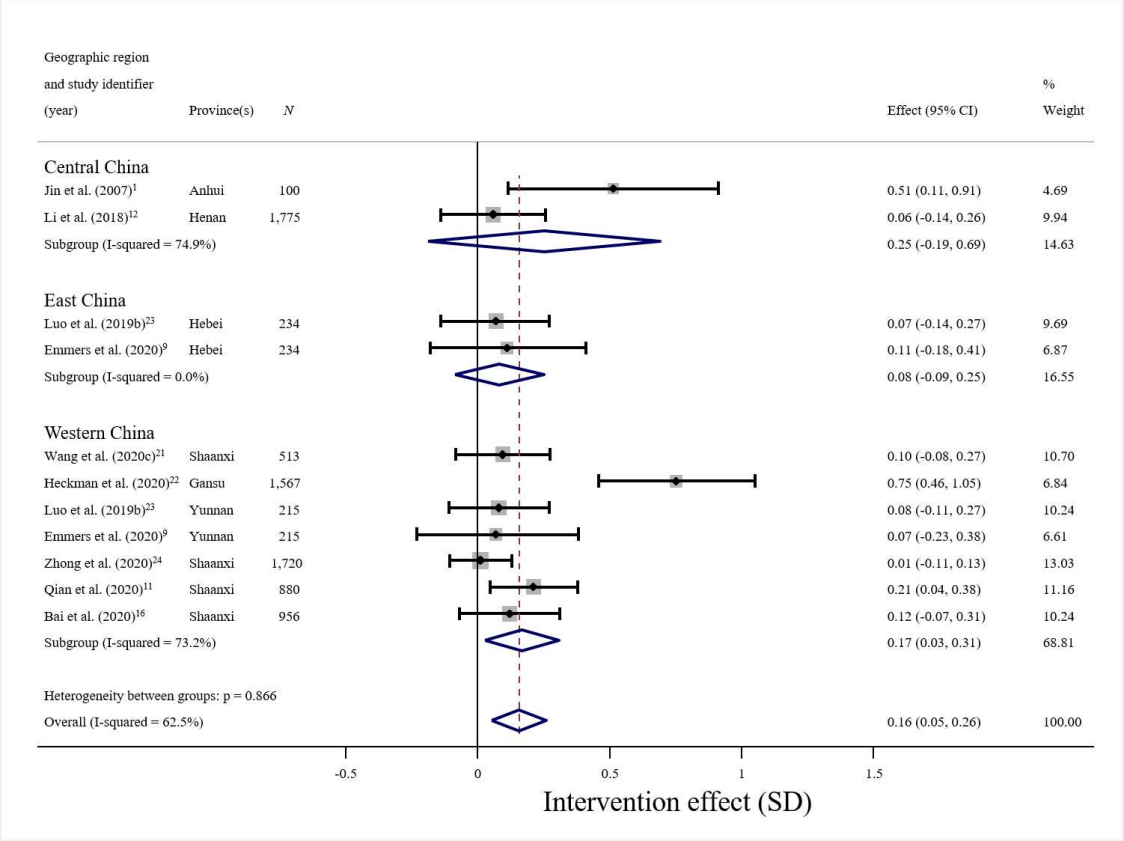
Appendix Figure 6.15: Pooled Prevalence of Early Social-Emotional Delay by Geographic Region



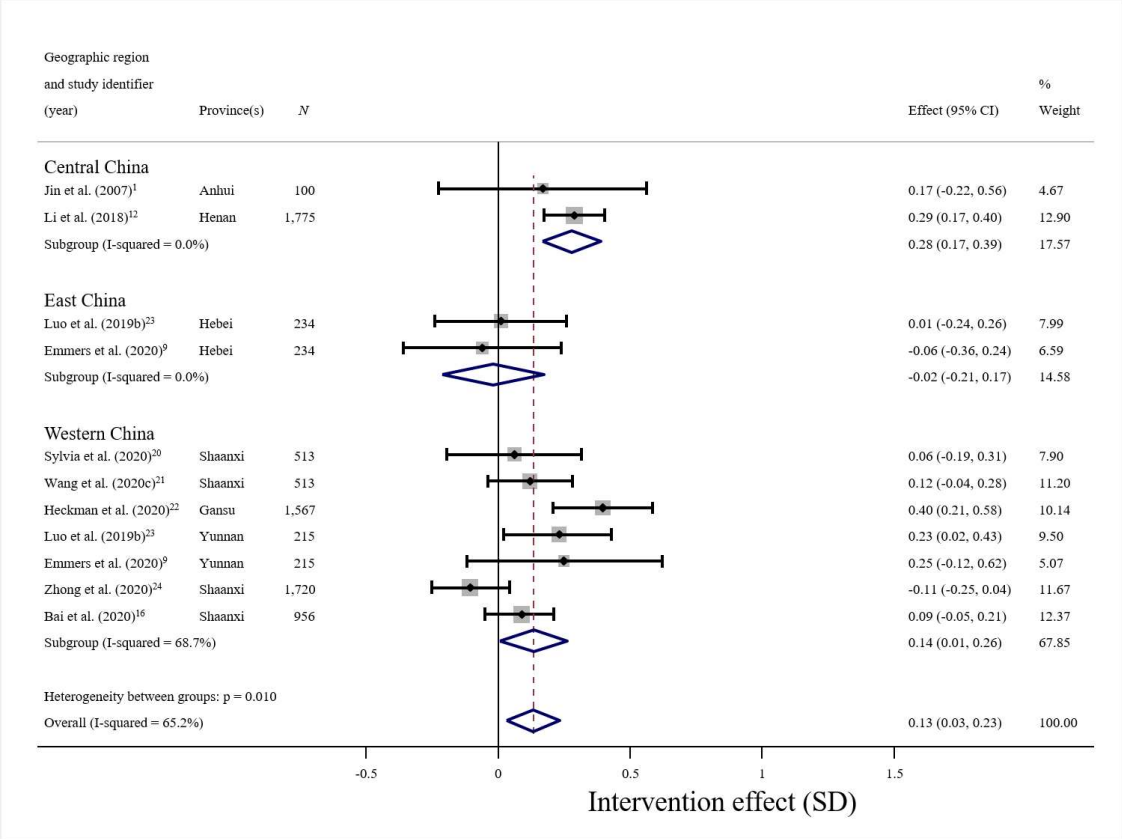
Appendix Figure 6.16: Pooled Intervention Effects on Early Cognitive Development by Geographic Region



Appendix Figure 6.17: Pooled Intervention Effects on Early Language Development by Geographic Region



Appendix Figure 6.18: Pooled Intervention Effects on Early Social-Emotional Development by Geographic Region



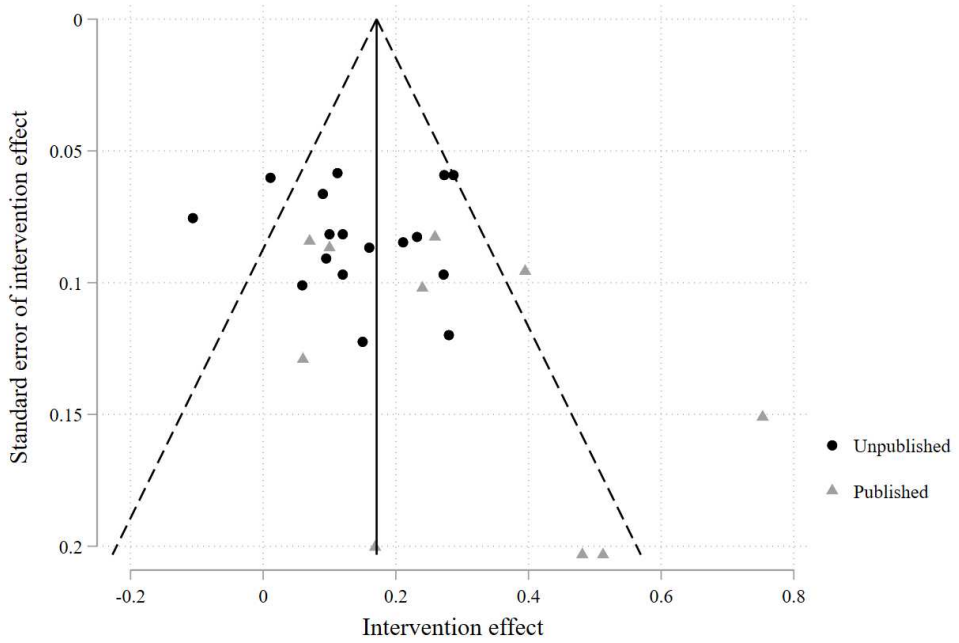
Appendix 7: Unpublished studies

1. Publication bias

We include several unpublished papers and reports in our analyses. Because these studies have not yet undergone peer review, there may be unintentional errors or biases in their data, although it should be noted that we found low levels of bias in all of our formal bias checks. As argued by Dwan et al.[37] empirical research shows that results in published studies are more likely to be positive and statistically significant than are results in unpublished studies. Moreover, publishing studies with insignificant impacts is more difficult and may take longer, which may lead to a "pipeline bias." This type of publication bias is a threat to the validity of systematic reviews.

A common method to test for publication bias is to analyze a funnel plot visually. A funnel plot usually presents effect sizes plotted against their standard errors. Appendix Figure 7.1 shows the funnel plot of the intervention effects in the published and unpublished studies that are included in the meta-analysis of this study. In the presence of publication bias, the funnel plot is expected to be skewed. The funnel plot of the published studies looks more skewed than does the funnel plot of the unpublished studies. The visual examination of asymmetry of funnel plots, however, can be subjective.[38] Therefore, we additionally use the trim-and-fill method and Egger's test to examine the presence of publication bias. First, the trim-and-fill method computes the number of studies that may be missing, and an adjusted intervention effect is derived by "trimming" smaller studies that cause funnel plot asymmetry and imputing studies that may be missing.[39] The trim-and-fill test indicates that no studies are missing from the reviewed literature and that the reported intervention effects are appropriate. Second, Egger's test regresses effect sizes on their precision.[38] In the absence of publication bias, the regression intercept is expected to be zero. It has been shown in the literature that Egger's test is more likely to overestimate the risk of publication bias for studies that assess the impacts on continuous outcome measures.[40] Even though the reviewed studies evaluate impacts on continuous ECD scores, the null hypothesis of the Egger test of no small-study effects is not rejected at the 10% level when we control for between-study heterogeneity induced by differences in ECD measurement tools ($p = 0.12$).

Appendix Figure 7.1: Funnel plot with pseudo 95% confidence limits



2. Actor(s) in charge of program implementation and role of co-authors

We provide information on the actors who were involved in the implementation of the included (published and unpublished) intervention studies and the role of co-authors in these studies.

Appendix Table 7.1: Actor(s) in Charge of Program Implementation and Role of Co-authors

Author(s)	Year	Province(s)	Actor(s) in charge of implementation	Role of co-authors
Jin et al.	2007[1]	Anhui	Research team that consists of members from the Department of Pediatric Development and Behavior of Shanghai Children's Medical Centre, Macfarlane Burnet Institute for Medical Research and Public Health, and School of Medicine of Shanghai Jiaotong University	No involvement
Sylvia et al.	2020[20]	Shaanxi	Family Planning Commission of the People's Republic of China	NW, SS, AM, RL, SR were involved in program evaluation
Wang et al.	2020c[21]	Shaanxi	Family Planning Commission of the People's Republic of China	LW, NW, YQ, SS, SR were involved in program evaluation
Heckman et al.	2020[22]	Gansu	State Council of the People's Republic of China: China Development Research Foundation (CDRF)	No involvement
Luo et al.	2019b[23]	Yunnan, Hebei	Save the Children (Humanitarian Aid Organization)	LR, DE, NW, SS, SR were involved in program evaluation
Emmers et al.	2020[9]	Yunnan, Hebei	Save the Children (Humanitarian Aid Organization)	DE, LR, SS, SR were involved in program evaluation
Zhong et al.	2020[24]	Shaanxi	Local villages and township doctors	RL, SS, SD, AM, SR were involved in program evaluation
Qian et al.	2020[11]	Shaanxi	Local villages and township doctors	QY, JQ, MA, RS were involved in program evaluation
Li et al.	2018[12]	Henan	One Sky (Foundation)	No involvement
Bai et al.	2020[16]	Shaanxi	Hupan Modou Foundation and National Health Commission of the People's Republic of China	DE was involved in program evaluation

Appendix Table 7.2: Summary of Included Studies and their Citations

Panel A. Reviewed studies included in systematic review or meta-analyses

No.	Author(s)	Year of data collection	Province(s)	Citation
1.	Jin et al. (2007)	2003	Anhui	Jin X, Sun Y, Jiang F, et al. "Care for Development" intervention in rural China: a prospective follow-up study. *J Dev Behav Pediatr* 2007;28:213–18.
2.	Ma et al. (2008)	2005–2006	Guangdong	Ma L, Chi L, Su Y, et al. Investigation of development and growth status of 0–18 months children in Guangdong (in Chinese). *Maternal and Child Health Care of China* 2008;13:1844–47.
3.	Yang et al. (2019)	2006–2008	Shaanxi	Yang X, Yin Z, Cheng Y, et al. Features and associated factors of the behavioral development of 24-month-old children in rural China: follow-up evaluation of a randomized controlled trial. *Sci Rep* 2018;8(1):13977.
4.	Yue et al. (2017)	2014	Shaanxi	Yue A, Shi Y, Luo R, et al. China's invisible crisis: cognitive delays among rural toddlers and the absence of modern parenting. *China J* 2017;78:50–80.
5.	Zhou et al. (2019a)	2013	Shanxi, Guizhou	Zhou S, Zhao C, Huang X, et al. The effect of a community-based, integrated and nuturing care intervention on early childhood development in rural China. *Public Health* 2019;167:125–35.
6.	Zhang et al. (2018)	2013	Shanxi, Guizhou	Zhang J, Guo S, Li Y, et al. Factors influencing developmental delay among young children in poor rural China: a latent variable approach. *BMJ Open* 2018;8:e021628.
7.	Wei et al. (2018)	2013	Shanxi, Guizhou	Wei Q, Zhang C, Zhang J, et al. Caregiver's depressive symptoms and young children's socioemotional development delays: a cross-sectional study in poor rural areas of China. *Infant Ment Health J* 2018;39:209–19.
8.	Luo et al. (2019a)	2015	Yunnan, Hebei	Luo R, Jia F, Yue A, et al. Passive parenting and its association with early child development. *Early Child Dev Care* 2019;189:1709–23.
9.	Emmers et al. (2020)	2015	Yunnan, Hebei	Emmers D, Luo R, Rozelle S, et al. Dynamics of early capability formation: evidence from a cluster-randomized parenting experiment in rural China. REAP Working Paper. 2020.
10.	Wang et al. (2020a)	2016	Shaanxi	Wang B, Luo X, Yue A, et al. Family environment in rural China and the link with early childhood development. *Early Child Dev Care* 2020; published online June 29. DOI:10.1080/03004430.2020.1784890.
11.	Qian et al. (2020)	2016	Shaanxi	Qian Y, Jiang Q, Medina A, et al. The impact of group-based early childhood intervention: evidence from below 1-year-old infants in rural China. REAP Working Paper. 2020.
12.	Li et al. (2018)	2016	Henan	Li G, Li Y. An evaluation of an early childhood development intervention in rural communities in Henan province. Henan University Department of Education Working Paper. 2018.
13.	Zhou et al. (2019b)	2016–2017	Jiangxi, Ningxia, Qinghai, Xinjiang	Zhou H, Ding Y, Yang Y, et al. Effects on developmental outcomes after cesarean birth versus vaginal birth in Chinese children aged 1–59 months: a cross-sectional community-based survey. *PeerJ* 2019;7:e7902.
14.	Wang et al. (2019)	2017	Beijing, Henan, Shaanxi	Wang L, Liang W, Zhang S, et al. Are infant/toddler developmental delays a problem across rural China?. *J Comp Econ* 2019;47:458–69.
15.	Zhong et al. (2019)	2017	Guizhou	Zhong J, Gao J, Liu C, Huang J, Luo R. Quantity-quality trade-off and early childhood development in rural family: evidence from China's Guizhou province. *Int J Environ Res Pub He* 2019;16:1307.
16.	Bai et al. (2020)	2018	Shaanxi	Bai Y, Emmers D, Li Y, et al. Effects of a government-led parenting programme on early child development outcomes in rural China: a cluster-randomised controlled trial. Catholic University Working Paper. 2020.
17.	Tan et al. (2020)	2018	Jiangxi, Guizhou, Sichuan, Hebei, Henan	Tan C, Zhao C, Dou Y, et al. Caregivers' depressive symptoms and social–emotional development of left-behind children under 3 years old in poor rural China: the mediating role of home environment. *Child Youth Serv Rev* 2020;116:105109.
18.	Wang et al. (2020b)	2019	Yunnan	Wang L. Rural early childhood development delays and parenting practices in yunnan province: data from baseline study sponsored by Save-the-Children. Shaanxi Normal University Working Paper. 2020.

19.	Emmers & Wang (2020)	2019	Shaanxi	Emmers D, Wang L, Sylvia S, et al. Rural-urban differences in the intergenerational transmission of cognitive capabilities in China. REAP Working Paper. 2020.
20.	Sylvia et al. (2020)	2014	Shaanxi	Sylvia S, Warrinnier N, Luo R, et al. From quantity to quality: delivering a home-based parenting intervention through China's family planning cadres. *Econ J* 2020; published online September 14. DOI:10.1093/ej/ueaa114.
21.	Wang et al. (2020c)	2014	Shaanxi	Wang L, Warrinnier N, Qian Y, et al. Persistence and fade-out: the impact of a short-term early childhood intervention at school entry. REAP Working Paper. 2020.
22.	Heckman et al. (2020)	2015	Gansu	Heckman JJ, Liu B, Lu M, et al. Treatment effects and the measurement of skills in a prototypical home visiting program. National Bureau of Economics Research (NBER) Working Paper Series. 2020. https://www.nber.org/papers/w27356 (accessed November 15, 2020).
23.	Luo et al. (2019b)	2015	Yunnan, Hebei	Luo R, Emmers D, Warrinnier N, et al. Using community health workers to deliver a scalable integrated parenting program in rural China: a cluster-randomized controlled trial. *Soc Sci Med* 2019;239:112545.
24.	Zhong et al. (2020)	2016	Shaanxi	Zhong J, Luo R, Sylvia S, et al. Passive versus active service delivery: comparing the effects of two parenting interventions on early cognitive development in rural China. REAP Working Paper. 2020.

Panel B. Other references

25. Huang H, Tao S, Zhang Y, et al. Standardization of Bayley Scales of Infant Development in Shanghai. *Chin J Child Health* 1993;1(3):158–160.
26. Yi S, Luo X, Yang Z, et al. The revising of Bayley Scales of Infant Development (BSID) in China. *Chin J Clin Psychol* 1993;1:71–75.
27. Xu M, Liu XH, Zhou XH, et al. Study on norms of Bayley Scales of Infant Development at the country in Shaanxi Province. *Chinese Journal of Child Health Care* 2009;17:125–127.
28. Bian X, Yao G, Squires J, et al. Translation and use of parent-completed developmental screening test in Shanghai. *J Early Child Res* 2012;10(2):162–175.
29. Bian X, Xie H, Squires J, et al. Adapting a parent-completed, socioemotional questionnaire in china: the Ages & Stages Questionnaires: Social-Emotional. *Infant Ment Health J* 2017;38(2);258–266.
30. Bayley N. Bayley Scales of Infant and Toddler Development–Third Edition. 2006, San Antonio, TX: Harcourt Assessment.
31. Hua J, Li Y, Ye K, et al. The reliability and validity of Bayley-III cognitive scale in China's male and female children. *Early Hum Dev* 2019;129:71–78.
32. Lowe JR, Erickson SJ, Schrader R, et al. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing? *Acta paediatr* 2012:101(2):e55–e58.
33. Månsson J, Stjernqvist K, Serenius F, et al. Agreement between Bayley-III measurements and WISC-IV measurements in typically developing children. *J Psychoeduc Assess* 2019;37(5):603–616.
34. Anderson PJ, De Luca CR, Hutchinson E, et al. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med* 2010;164(4):352–356.
35. Serenius F, Källén K, Blennow M, et al. Neurodevelopmental outcome in extremely preterm infants at 2.5 years after active perinatal care in Sweden. *Jama* 2013;309(17):1810–1820.
36. Bos AF. Bayley-II or Bayley-III: What do the scores tell us? *Dev Med Child Neurol* 2013;55(11):978–979.
37. Dwan K, Gamble C, Williamson PR, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PloS One* 2013;8:e66844.
38. Lin L, Chu H. Quantifying publication bias in meta-analysis. *Biometrics* 2018;74(3):785–794.
39. Duval S, Tweedie R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56(2):455–463.
40. Doleman B, Freeman SC, Lund, et al. Funnel plots may show asymmetry in the absence of publication bias with continuous outcomes dependent on baseline risk: presentation of a new publication bias test. *Res Synth Methods* 2020;11(4):522–534.