Have Parenting Programs for Disruptive Child Behavior Become Less Effective?

Manuscript in preparation for invited resubmission for publication in

Journal of Child Psychology and Psychiatry

Patty Leijten, PhD (University of Amsterdam)

G.J. Melendez-Torres, DPhil (University of Exeter)

Sophia Backhaus, DPhil (University of Amsterdam)

Frances Gardner, DPhil (University of Oxford)

Annabeth P. Groenman, PhD (University of Amsterdam)

Tycho J. Dekkers, PhD (Accare Child Study Center)

Barbara J. van den Hoofdakker, PhD (University Medical Center Groningen)

Liina Björg Laas Sigurðardóttir, MSc (University of Amsterdam)

Marjolein Luman, PhD (VU Amsterdam)

Merlin Nieterau, MSc (University of Amsterdam)

Saskia van der Oord, PhD (KU Leuven)

Geertjan Overbeek, PhD (University of Amsterdam)

Constantina Psyllou, MSc (University Medical Center Groningen)

Karen Rienks, MSc (University of Amsterdam)

Susanne Schulz, PhD (University of Amsterdam)

John R. Weisz, PhD (Harvard University)

Correspondence to

Patty Leijten, Research Institute of Child Development and Education, University of Amsterdam. Nieuwe Achtergracht 127, 1018 WS Amsterdam, the Netherlands.

p.leijten@uva.nl

Conflicts of interest

The authors disclose no conflicts of interest in the work conducted for this manuscript.

Acknowledgements

We thank Janina Jochim (University of Oxford) for her valuable help with the data extraction.

Funding

This work was supported by grants from the Netherlands Organisation for Health Research and Development (ZonMw; 636320007) and the Dutch Research Council (NWO; VI.Vidi.201.065) awarded to Patty Leijten.

Abstract

Background: We examined how estimates of parenting program effects on disruptive child behavior have evolved over time. In addition, we examined if any time trends in effect sizes can be explained by trial, sample, or intervention characteristics. **Methods:** We systematically searched 22 international and regional databases, grey literature, and 4 trial registries for randomized controlled trials of behavioral parenting programs. **Results:** We identified 235 eligible trials (1,003 effect sizes) from 35 countries representing all six WHO regions. We found that intervention effects have grown smaller over time, with an average reduction of $d = 0.09$ per decade (95% CI [0.03, 0.16]). The trend was non-linear ($\beta = -0.06$, 95% CI [−0.10, −0.01]): effects reduced, from $d = -0.83$ in the 1980s to $d = -0.57$ in the 1990s to $d = -0.35$ in the 2000s, and then stabilized. Reductions in effect size were partly explained by trial characteristics (i.e., increased sample sizes and active control conditions, and reduced risk of selective outcome reporting) and sample characteristics (i.e., less severe samples and fewer boys). Other changes over time, unrelated to effect sizes, include older samples, less use of time-out, reduced use of observational outcome measures, less program developer involvement, and less risk of bias regarding random sequence generation and concealment, and incomplete outcome data. **Conclusions:** Our findings suggest that estimates of parenting program effect sizes are currently stable: Effect sizes are no longer reducing but there is also no evidence of recent increases. Experimentation with the content, delivery, and personalization of parenting programs is needed to identify ways to increase program effects.

Keywords: Behavioral parenting program; disruptive child behavior; systematic review; meta-analysis; time trends.

Abbreviations: Not applicable.

Have Parenting Programs for Disruptive Child Behavior Become Less Effective?

The first randomized controlled trials of behavioral parenting programs to reduce disruptive (i.e., oppositional, defiant, aggressive) child behavior were published in the late 1970s and early 1980s (Forgatch & Toobert, 1979; Webster-Stratton, 1982). These studies reflected a paradigm shift in strategies to reduce disruptive child behavior—from child-focused psychodynamic play-therapy to training parents to redirect children's behavior based on operant and social learning theory principles (Patterson et al., 2002). In the following decades, an increasing number of randomized trials and dozens of meta-analyses would be published (e.g., Beelmann et al., 2023; Mingebach et al., 2018). This ever-expanding literature raises the question as to what extent parenting program effect estimates have changed over time, and how this may be influenced by changes in the programs, in the samples that they target, or in the trials that evaluate them. In the present study, we examined how estimates of parenting program effects have evolved over time. In addition, we examined whether any time trends can be explained by trial, sample, or intervention characteristics.

In the early 1960s, researchers at the University of Oregon broke with the tradition to treat disruptive child behavior with psychodynamic play-therapy. Simple reinforcement techniques to strengthen non-disruptive behavior yielded rapid and strong improvements in children's behavior (Patterson, 1965), especially when combined with non-violent negative consequences for disruptive behavior, such as time-out and taking away privileges (Wells, 1997). This led to programs that trained parents to change their behavior in order to reduce disruptive child behaviors. In the following years, the Oregon team and investigators from other parts of the United States built theories and programs on how parents can alter the development of disruptive child behavior. A notable example was Patterson's model of coercive interaction cycles, explaining how parents and children can unwittingly reinforce

aversive behavior in each other, leading to interaction cycles that become increasingly difficult to break (Patterson, 1982). This model, which has been widely supported empirically (e.g., Lunkenheimer et al., 2016; Smith et al., 2014), combined with Hanf's model stressing the need to strengthen parent-child relationship quality in families with disruptive child behavior (Hanf, 1969), forms the basis of most established parenting programs to prevent or treat disruptive child behavior (Kaehler et al., 2016; Reitman & McMahon, 2013; Weisz & Kazdin, 2017).

Now, more than half a century later, behavioral parenting programs are implemented across the globe (Backhaus et al., 2023) and every year new meta-analyses synthesize the findings of an increasing number of trials (e.g., Beelmann et al., 2023). This wealth of research places us in the unique position to examine how parenting program effects have evolved over time. Insights in such time trends can help us better understand the factors that determine the magnitude of parenting program effect sizes and can guide future research and intervention development to increase the impact of parenting programs.

Evidence from trials on interventions for child and adolescent conduct problems suggests that intervention effects have gradually reduced over five decades (Weisz et al., 2019). Similar trends, with mainly initial reductions in effect sizes, are observed in other fields, such as cognitive behavioral therapy for adult depression (Ljótsson et al., 2017). The reasons for declining intervention effects are not well understood. Suggested explanations include that trials have become more rigorous over time, with larger samples, less risk of bias and more active control groups. Several meta-analyses indeed show that less rigorous trials yield larger effect sizes (e.g., Cuijpers et al., 2010; Gellatly et al., 2007). This may be because small samples increase the likelihood of chance findings and these larger positive effects may have been more likely to get published, and because active control groups tend to yield smaller between-group effects (Rifkin, 2007). Researcher allegiance may have increased

effect sizes especially in some of the earlier trials (Leykin & DeRubeis, 2009; Ljótsson et al., 2017). Researcher allegiance effects can be caused by biases—distortion of findings because of program developers' preferences (Luborsky et al., 1999), but also by better expertise and skill with the target treatment, and by accumulation of evidence from prior studies indicating that a particular treatment is in fact more effective (Hollon, 1999).

Alternative explanations include that the traditional structure of interventions, often weekly in-person sessions, may fit less well with societal life today than it did several decades ago (Weisz et al., 2019). The use of smart phones may have made individuals accustomed to having help at hand more immediately. Although an increasing number of interventions use technology to support flexible delivery (e.g., Jones et al., 2021), many still use the traditional structure. In addition, most parents nowadays have better access to online information about children's development and parenting (Canário et al., 2022), making the content of parenting programs less novel and therefore perhaps less likely to have a big impact. Also, attempts to increase program effects may have led scholars to add new elements to programs, while evidence suggests that stacking elements is often counter-effective (Bakermans-Kranenburg et al., 2003; Leijten et al., 2022). Last, it may be that the factors contributing to youth mental health problems are changing faster than our treatment are able to change to address those factors (Weisz et al., 2019).

For behavioral parenting programs specifically, time trends in effect sizes are unknown. There may be similar declines as in the broader field of interventions for youth conduct problems (Weisz et al., 2019). In addition, some effective techniques to increase children's compliance, most notably time-out, have been under scrutiny (Readdick & Chapman, 2000; Siegel & Bryson, 2014) and may be used less in more recently evaluated programs. Also, given the global dissemination of parenting programs in the past decades (Backhaus et al., 2023), there may be changes in sample characteristics that influence

program effect sizes (e.g., baseline problem severity or cultural fit; Leijten et al., 2020; Morelli et al., 2018).

However, there may also be reasons to be hopeful that parenting programs have become more effective in recent decades. The field has invested in improving our understanding of the program elements that are most likely to contribute to effective change (e.g., Kaminski et al., 2008; Leijten et al., 2019), in adapting transported programs to local contexts while maintaining functional fidelity (Baumann et al., 2015), and in technological developments allow for more versatile delivery formats that may better fit families' needs (e.g., Duppong-Hurley et al., 2016). The goal of these investments was to provide more effective programs.

The Present Study

We examined how our estimates of the effects of parenting programs on disruptive child behavior have evolved over time. Specifically, we examined (1) linear and non-linear time trends in estimated effects; (2) how trial, sample, and intervention characteristics relate to time and effect size; and (3) whether any time trends in effect sizes could be accounted for by time trends in trial, sample, and intervention characteristics. Our analyses were exploratory, guided by research questions rather than hypotheses.

Methods

Information Sources and Search Strategy

Our primary source of studies was the pre-registered systematic literature review by Backhaus et al. (2023; preregistered on PROSPERO CRD42019141844; we did not pre-register separately), based on an extensive search in 26 international databases (e.g., PsycINFO and MEDLINE), including various non-English databases (e.g., Biomedical journals from India and Russia), grey literature (e.g., ProQuest, China Doctoral Dissertations

Full-Text Database), and trial registries (e.g., ClinicalTrials.gov and WHO International Clinical Trials Registry Platform), last updated in August 2022.

Eligibility Criteria and Selection Process

We included (i) randomized controlled trials that compared the effects of (ii) a behavioral parenting program on (iii) disruptive child behavior against (iv) any control. We focused on trials including samples of (v) children with a mean age between 2 and 10 years, because conduct problems in adolescence require a more systemic approach (Boldrini et al., 2023). We defined behavioral parenting programs as the program content being primarily based on operant and social learning principles. These programs typically teach parents to use differential attention and reinforcement (e.g., reinforcing positive behavior and avoiding reinforcement of disruptive behavior). They vary in the extent to which they complement this approach with other intervention components (e.g., increasing parent-child relationship quality, emotion regulation, and parental self-management; Kaehler et al., 2016).

Data Items and Collection Process

Three team members (SB, JJ, PL) independently extracted data items using a piloted data extraction form, and independently coded numerous study variables, with coder disagreements resolved through discussion. In addition to the study variables, we coded trial characteristics for descriptive purposes (e.g., country and its World Bank income status). We also coded whether programs were implemented as universal prevention (i.e., targeting the general population), selective prevention (i.e., targeting families with risk factors for disruptive child behavior problems), indicated prevention (i.e., targeting families with elevated levels of disruptive child behavior problems), or treatment (i.e., targeting families referred for treatment of disruptive child behavior problems).

Time

We coded publication year for each trial. We then divided publication year by 10, to facilitate interpretation of regression coefficients. Last, we grand mean centered publication year and computed a squared publication year, to explore the possibility of non-linear trends.

Effect Size

We computed Cohen's *d* based on post-intervention means and standard deviations, subtracting children's mean level of disruptive behavior in the control condition from children's mean level of disruptive behavior in the intervention condition, and dividing this difference by the pooled standard deviation of both conditions. Where means and standard deviations were not reported, we used relevant model statistics such as regression coefficients and *F*-values and converted these to Cohen's *d*.

Trial Characteristics

We coded (i) sample size (i.e., total number of participants at baseline), (ii) active control condition (0 = *no intervention or waitlist*; 1 = *low intensity intervention or unstandardized treatment as usual*), (iii) developer involvement (0 = *no intervention developer co-authored the publication of the trial findings*; 1 = *a developer of the intervention co-authored the publication*), (iv) preregistration in a trial register (e.g., clinicaltrials.gov) or preregistered analysis plan on a preregistration platform (e.g., Open Science Framework or AsPredicted) (0 = *no*; 1 = *yes*), and (v) risk of bias using the Cochrane Risk of Bias Tool for randomized controlled trials (Higgins et al., 2011): random sequence generation, random sequence concealment, blinding of participants, blinding of outcome assessors, missing outcome data, and selective outcome reporting. For each indicator, we rated risk of bias as 0 (*low*), 1 (*unclear*), or 2 (*high*), except for blinding of outcome assessors. We rated blinding of outcome assessors as 0 (*low*) when the study included at least one observational measure of disruptive child behavior and as 2 (*high*) when the study did not include an observational measure of disruptive child behavior. We used version 1.0 for the

tool, rather than version 2.0, because version 1.0 focuses on trial level rigor and version 2.0 focuses on outcome level rigor; the focus of our study is on trial characteristics.

Sample Characteristics

We coded (i) children's mean age in years at baseline, (ii) child sex (i.e., percentage of boys in the sample), (iii) baseline severity of disruptive child behavior, coded in the subset of trials (48%) that used the Eyberg Child Behavior Inventory (ECBI) Intensity Scale, all other measures were used less frequently, (iv) socioeconomic disadvantage (0 = *sample not primarily disadvantaged*; 1 = *sample primarily disadvantaged*); and (v) whether the sample was culturally and/or ethnically different from the sample for which the intervention was originally developed (0 = *no*; 1 = *yes*).

Intervention Characteristics

We coded (i) number of sessions, (ii) delivery mode (0 = *group*; 1 = *individual* as well as 0 = *self-directed*; 1 = *therapist-led*), including separate codes for the use of in vivo coaching and video feedback (0 = *no;* 1 = *yes* for each):, and (iii) whether the intervention included the following elements (each coded as 0 = *no;* 1 = *yes*): reinforcement of positive child behavior, avoiding reinforcement of disruptive child behavior (including separate codes for time-out, ignore, and the use of logical/natural consequences or removing privileges), relationship enhancement techniques (including a separate code for child-led play), parental self-management (e.g., relaxation or increasing social support), and teaching parents how to enhance children's broader skills (e.g., teaching children social or problem solving skills).

Data Synthesis

We used robust variance estimation (Hedges et al., 2010) to estimate (1) an intercept only model with the overall program effect sizes; (2) a series of single predictor models with publication year predicting effect size, as well as each of the individual trial, sample, and intervention characteristics predicting publication year and effect size; and (3) a series of

multi-predictor models, with both publication year and trial, sample, and intervention characteristics predicting effect sizes. Robust variance estimation weighs the multiple effect sizes included trials using an approximate variance-covariance matrix (Hedges et al., 2010). Within-trial correlation of effect sizes was set at the default of $\rho = .80$.

Results

Included Trials

Our systematic search yielded 21,020 unique hits (Figure 1). Of the 1,248 full-text papers, dissertations, and trial registrations examined, we included 235 eligible trials with 1,003 effect sizes based on 27,023 families (see Appendix for included trials). Of these 235 trials, 21% had been included by Weisz et al. (2019), indicating the different scope of the two reviews (e.g., our inclusion of prevention trials, searching 26 international databases, and including non-English language reports).

The overall effect of parenting programs on disruptive child behavior was $d = -0.37$ (95% CI [−0.43, −0.32]). Converting this to a common language effect size (Magnusson, 2022) suggests that any family participating in a parenting program had a likelihood of 60.3% to experience significantly fewer disruptive behavior than any family not participating in a parenting program. In terms of numbers needed to treat (Cook & Sackett, 1995), this reflects that between eight and nine families needed to participate for one family to experience significant benefit.

Trials were conducted in 35 different countries from all six World Health Organisation regions. Initial dominance by the United States gradually shifted to dominance by northern and western Europe. Since the 2020s, 27% of the trials have been conducted in middle income countries. This is an encouraging increase from 3% in the 2000s and 10% in the 2010s, but still a severe underrepresentation of the 75% of the world population living in middle income countries (World Bank, 2022). Low-income countries (6% of the world

population; World Bank, 2022) are also seriously underrepresented with only one trial (0.4%) (Sim et al., 2014). Effect sizes in high income countries did not differ from those in middle income countries (differential $d = 0.19$ (95% CI [−0.19, 0.56]).

In the 2000s and 2010s, universal and selective prevention studies made up 50% to 57% of the trials. In both earlier and recent years around 70% of the trials were indicated or treatment trials and 30% were universal or selective prevention trials. Universal prevention yielded an overall effect of $d = -0.27$ (95% CI [−0.40, −0.13]; $k = 40$; $n = 160$), selective prevention $d = -0.33$ (95% CI [−0.41, −0.25]; $k = 76$; $n = 281$), indicated prevention $d = -0.42$ (95% CI [−0.55, −0.28]; $k = 62$; $n = 314$), and treatment $d = -0.49$ (95% CI [−0.62, −0.36]; $k = 58$, $n = 248$).

Time Trends in Effect Sizes

Estimates of parenting program effects on disruptive child behavior have become smaller over time, with an average reduction of $d = 0.09$ per decade (95% CI [0.03, 0.16]). Adding a quadratic term to the model revealed that this trend is in fact non-linear ($\beta = -0.06$, 95% CI [−0.10, −0.01]): Effect sizes reduced during the first decades and then stabilized around the year 2000 (Figure 2). There was no evidence of increases in program effects in any of the later years.

Time Trends in Trial, Sample, and Intervention Characteristics (Table 1)

Trial Characteristics

There was no overall time trend for sample size, but trials with larger samples yielded smaller effect sizes: with every additional 100 participants, effect sizes became $d = 0.10$ weaker (95% CI [0.05, 0.15]). Trials used more active (rather than passive) control conditions in recent decades: 20% in the 1980s, 9% in the 1990s, 18% in the 2000s, 22% in the 2010s, and 38% in the 2020s. Active control conditions yielded smaller effects (on average $d = 0.18$ smaller, 95% CI [0.08, 0.28]). The first preregistered trial was published in 2007 (Van den

Hoofdakker et al., 2007). In the 2010s, about a third of the trials were preregistered and in the 2020s around two thirds. Preregistered trials did not yield smaller effects (differential d = 0.11, 95% CI [−0.01, 0.23]). In 64% of the trials, program developers co-authored publications on program evaluations. Program developer involvement was less common in more recent trials (β = –4.22, 95% CI [–6.35, –2.09]), but was not associated with effect size (β = –0.004, 95% CI [−0.13, 0.12]).

Of the six indicators of risk of bias, four improved over time: bias regarding randomization sequence (β = −5.56, 95% CI [−7.91, −3.20]), randomization concealment (β = −3.51, 95% CI [−5.52, −1.50]), incomplete data (β = −3.39, 95% CI [−5.78, −1.00]), and selective outcome reporting (β = −6.78, 95% CI [−8.90, −4.66]). There was no change in blinding of participants (β = −1.51, 95% CI [−9.74, 6.73]). Blinding of outcome assessors worsened over time (β = 9.33, 95% CI [4.80, 13.87]), but effect sizes based on observational outcomes were not smaller than effect sizes based on unblinded ratings (differential d = 0.09, 95% CI [−0.12, 0.31]). Only higher risk of selective outcome reporting predicted larger effect sizes (β = −0.11, 95% CI [−0.21, −0.01]. Changes in risk of selective outcome reporting did not merely reflect reporting more outcomes (average number of effect sizes reported in the 1980s = 6.42, 1990s = 3.94, 2000s = 4.53, 2010s = 3.95, 2020s = 4.11).

Sample Characteristics

More recent trials included fewer boys and children were on average older. Samples with larger proportions of boys yielded larger effects (β = –0.51 [–1.02, –0.01]); child age was not associated with program effects. The subset of trials that use the ECBI (48%) suggests that samples have become less severe in terms of baseline disruptive child behavior (β = −0.06, 95% CI [−0.10, −0.003]). Trials with lower baseline ECBI scores yielded smaller effects: for every 10 points on the ECBI (possible range 36–252) d became weaker with .04 (95% CI [0.08, 0.10]). Sample socioeconomic disadvantage was not associated with time or

effect size. Cross-cultural transportation was more common in recent years (β = 5.50 (95% CI [3.39, 7.60]), but did not predict effect size (differential d = −0.01, 95% CI [−0.12, 0.011]).

Intervention Characteristics

There were no time trends regarding the number of sessions or delivery format. Time-out was used less often in more recently evaluated programs (β = −4.13, 95% CI [−7.75, −0.51]), but the use of time-out was not associated with effect size (differential d = −0.09, 95% CI [−0.27, 0.09]). None of the other intervention elements tested was associated with either time or effect size, except for the use of video feedback, which was not associated with time, but associated with smaller effect size (differential d = 0.23, 95% CI [0.03, 0.43]). This might be because video feedback was mainly used in programs primarily focused on increasing parental sensitivity, with disruptive child behavior being a secondary outcome (e.g., Stolk et al., 2008).

Do Time Trends in Trial, Sample, and Intervention Characteristics Account for Time Trends in Effect Sizes?

We tested whether the five study characteristics that predicted effect size accounted for the time trend in effect sizes. This included sample size, active control, risk of selective outcome reporting, disruptive child behavior severity, and proportion of boys. We excluded the use of video-feedback from this because its association with effect size seemed confounded by intervention characteristics. Multivariate meta-regressions (Table 2) showed that when all five study characteristics were combined in one model, the time trend in effect size was no longer significant. When the five study characteristics were tested individually, the time trend in effect sizes remained, indicating that none of the individual characteristic fully accounted for the time trend. This suggests that reductions in effect sizes over time are at least partly due to more recent trials including some combination of larger and less severe samples with fewer boys, more active control conditions, and less risk of selective outcome

reporting. Including the use of video-feedback in these models did not change any of the findings.

Discussion

We examined time trends in the estimated effect sizes of behavioral parenting programs for disruptive child behavior. Effect sizes declined during the first decades (1980s and 1990s) and then stabilized (2000s–2020s). Declines in effect sizes were partly explained by trial characteristics (larger sample size, more active control conditions, and less risk of selective outcome reporting) and sample characteristics (more severe samples and more boys). There was no evidence of increases in program effects in later years.

Our finding that estimates of effect sizes were reduced over time matches earlier findings on the effects of different types of psychotherapy for child and adolescent conduct problems (Weisz et al., 2019). In addition, the identified non-linear trend matches the trend observed for psychotherapy for adult depression (Ljótsson et al., 2017). Three factors partly explaining the time trend reflect increased rigor in terms of increases in sample size, active control conditions, and complete outcome reporting. Publications relying on small samples and selective reporting may be more likely to present chance findings favoring the intervention (Kühberger et al., 2014), and such positive findings may be more likely to be reported and ultimately published. Also, active control conditions are known to yield smaller between-group effect sizes (Mohr et al., 2014; Weisz et al., 2017). The two other factors reflect study samples having become less severe and less dominated by boys. Problem severity is indeed the most robust moderator of parenting program effect size (McMahon et al., 2021) and it is therefore not surprising that if samples become less severe, effect sizes decrease. Child sex does not tend to moderator parenting program effect size (McMahon et al., 2021), but it may be that because prevalence rates of disruptive behavior tend to be higher

in boys (Demmer et al., 2017), child sex confounds with problem severity in more prevention-oriented samples.

There was no evidence that program effects are on the rise in recent years. Insights into, for example, the most effective ingredients of parenting programs (e.g., Leijten et al., 2019) and the use of technology to support parents' participation (Jones et al., 2021), are still relatively recent and may need more time to start influencing program effects. Also, it might be that as long as programs rely on the same core principles, effect sizes will remain largely the same (Jones et al., 2019; Weisz et al., 2019). That said, alternative parenting program approaches to reducing disruptive child behavior have been evaluated (e.g., emotion coaching, Havighurst et al., 2013; attachment enhancement, Stattin et al., 2015; mindfulness, Coatsworth et al., 2015), but none of these has been shown to outperform conventional behavioral parenting programs in terms of their effects on disruptive child behavior.

Much has changed in society, family life and parenting, and information transmission in the past decades. It might well be that a variety of societal, cultural, and technical changes in the advent of the current millennium have contributed to the relatively stable reduction in the effects of behavioral parent training programs. For example, how parents perceive and implement support from experts may be influenced by parents' access to online information about disruptive child behavior and parenting (Dworkin et al., 2013), by trends towards intensive and perfectionistic parenting in some countries (Faircloth, 2023), and by parental feelings of confidence or burn-out in their parenting role, which—seemingly paradoxically—are both higher in current than in previous generations (Mikolajczak et al., 2019; Pew Research Center, 2015). Future investigations of time trends in the effects of parenting programs ideally incorporate data on historical societal trends to complement the findings from the present study.

Strengths of our study include that the data span more than four decades, reflect studies in 35 different countries from all six WHO regions, were not limited to English-language reports, and included more than 1,000 effect sizes from 235 trials of programs based on the same underlying principles. This yielded meaningful variation in key factors (i.e., time and trial characteristics), while keeping other factors relatively constant (e.g., the main content of the intervention). In addition, our examination of both linear and non-linear associations, revealed that effect sizes declined mainly in the earlier decades. Last, we considered potentially confounding factors at the level of trial design, sample characteristics, and intervention content and delivery that have been shown to predict intervention effect sizes (Leijten et al., 2019; Luborsky et al., 1999).

Limitations of our study include that our findings reflect associations between developments over time; we cannot rule out the possibility that other, or additional, historical developments caused effect sizes to decrease. In addition, we were limited by the trial, sample, and intervention characteristics reported in the original studies. Especially for sample characteristics, aggregate level meta-analysis like the one presented here masks meaningful variation within samples, because it relies on trial level data (e.g., percentage of boys and mean age) rather than the characteristics of individual children. Last, because the number of trials published in the first decades is much smaller than the number of trials published in the last few decades (Figure 2), estimates of program effects in the earliest years may be less reliable.

Our findings warrant reservation, at least from a scientific perspective, about conducting more trials of the same parenting programs in similar setting—they will likely only confirm what that hundreds of randomized trials and dozens of meta-analyses have already established. If we aspire to increase the effects of parenting programs, experimentation with program content, delivery, and personalization, based on insights from

basic mental health science (Holmes et al., 2014), is needed. In terms of program content, our understanding of the key ingredients of behavioral parenting programs could lead to leaner versions of programs that are more versatile for experimenting with the change, addition, or removal of elements (Leijten et al., 2021). In addition, experimenting can be done with the delivery of parenting programs, such as using a more flexible structure (e.g., yearly check-ups; Dishion et al., 2008), and with assigning program content based on individual family characteristics. Such scientific exercises can only be done if program manuals are made available via open access, which is increasingly common, but still not the norm (Cuijpers et al., 2014).

Furthermore, rigorous experimentation with intervention content, delivery, and personalization requires a diverse set of evaluation designs. Randomized trials with pre- and post-intervention assessments are powerful for estimating overall program effects but less useful for identifying subtle but meaningful impact of changes to program content or delivery (Leijten et al., 2021). Technology has made frequent assessments easier, allowing for research designs that track families' progress during the intervention to better understand when changes occur. Tracking families' progress during interventions for youth mental health is not new (Chorpita et al., 2008; Deković et al., 2012), but underused. Another promising development is integrative data analysis (i.e., pooling data from different trials) which can yield new insights into different mechanisms underlying change in different families (Laas Sigurðardóttir et al., 2023). These insights can be used to better tailor and target programs. The next few decades will show whether these innovations translate into improvements in the effects of parenting programs on disruptive child behavior.

Key Points and Relevance

- It has been 45 years since the first randomized evaluations of behavioral parenting programs.
- We found that estimates of parenting program effects on disruptive behavior initially reduced and have remained stable since the new millennium.
- Reductions were at least partly explained by increased trial rigor and by samples that have become less severe and include and fewer boys.
- Experimentation with the content, delivery, and personalization of parenting programs is needed to identify ways to increase their effects.

References

Backhaus, S., Leijten, P., Jochim, J., Melendez-Torres, G. J., & Gardner, F. (2023). Effects over time of parenting interventions to reduce physical and emotional violence against children: A systematic review and meta-analysis. *EClinicalMedicine, 60*. https://doi.org/10.1016/j.eclinm.2023.102003

Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129*(2), 195–215. https://doi.org/10.1037/0033-2909.129.2.195

Baumann, A. A., Powell, B. J., Kohl, P. L., Tabak, R. G., Penalba, V., Proctor, E. K., ... & Cabassa, L. J. (2015). Cultural adaptation and implementation of evidence-based parent-training: A systematic review and critique of guiding evidence. *Children and Youth Services Review, 53*, 113–120. https://doi.org/10.1016/j.childyouth.2015.03.025

Beelmann, A., Arnold, L. S., & Hercher, J. (2023). Parent training programs for preventing and treating antisocial behavior in children and adolescents: A comprehensive meta-

analysis of international studies. *Aggression and Violent Behavior, 68*, 101798. https://doi.org/10.1016/j.avb.2022.101798

Boldrini, T., Ghiandoni, V., Mancinelli, E., Salcuni, S., & Solmi, M. (2023). Systematic review and meta-analysis: Psychosocial treatments for disruptive behavior symptoms and disorders in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 62*(2), 169–189. https://doi.org/10.1016/j.jaac.2022.05.002

Canário, A. C., Byrne, S., Creasey, N., Kodyšová, E., Kömürcü Akik, B., Lewandowska-Walter, A., Modić Stanke, K., Pećnik, N., & Leijten, P. (2022). The use of information and communication technologies in family support across Europe: A narrative review. *International Journal of Environmental Research and Public Health, 19*(3), 1488.

Chorpita, B. F., Bernstein, A., Daleiden, E. L., & Research Network on Youth Mental Health. (2008). Driving with roadmaps and dashboards: Using information resources to structure the decision models in service organizations. *Administration and Policy in Mental Health and Mental Health Services Research, 35*(1–2), 114–123. https://doi.org/10.1007/s10488-007-0151-x

Coatsworth, J. D., Duncan, L. G., Nix, R. L., Greenberg, M. T., Gayles, J. G., Bamberger, K. T., Berrena, E., & Demi, M. A. (2015). Integrating mindfulness with parent training: Effects of the Mindfulness-Enhanced Strengthening Families Program. *Developmental Psychology, 51*(1), 26–35. https://doi.org/10.1037/a0038212

Cook, R. J., & Sackett, D. L. (1995). The number needed to treat: A clinically useful measure of treatment effect. *British Medical Journal, 310*(6977), 452–454. https://doi.org/10.1136/bmj.310.6977.452

Cuijpers, P., Boyce, N., & Van Ommeren, M. (2024). Why treatment manuals of psychological interventions should be freely available. *The Lancet Psychiatry, 11*(5), 325–326. https://doi.org/10.1016/S2215-0366(24)00071-3

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine, 40*(2), 211–223. https://doi.org/10.1017/S0033291709006114

Deković, M., Asscher, J. J., Manders, W. A., Prins, P. J., & van der Laan, P. (2012). Within-intervention change: Mediators of intervention effects during multisystemic therapy. *Journal of Consulting and Clinical Psychology, 80*(4), 574–587. https://doi.org/10.1037/a0028482

Demmer, D. H., Hooley, M., Sheen, J., McGillivray, J. A., & Lum, J. A. (2017). Sex differences in the prevalence of oppositional defiant disorder during middle childhood: A meta-analysis. *Journal of Abnormal Child Psychology, 45*, 313–325. https://doi.org/10.1007/s10802-016-0170-8

Dishion, T. J., Shaw, D., Connell, A., Gardner, F., Weaver, C., & Wilson, M. (2008). The family check-up with high-risk indigent families: Preventing problem behavior by increasing parents' positive behavior support in early childhood. *Child Development, 79*(5), 1395–1414. https://doi.org/10.1111/j.1467-8624.2008.01195.x

Doffer, D. P., Dekkers, T. J., Hornstra, R., Van der Oord, S., Luman, M., Leijten, P., Hoekstra, P. J., Van den Hoofdakker, B. J., & Groenman, A. P. (2023). Sustained improvements by behavioural parent training for children with attention-deficit/hyperactivity disorder: A meta-analytic review of longer-term child and parental outcomes. *JCPP Advances, 3*(3), e12196.

Duppong-Hurley, K., Hoffman, S., Barnes, B., & Oats, R. (2016). Perspectives on engagement barriers and alternative delivery formats from non-completers of a community-run parenting program. *Journal of Child and Family Studies, 25*, 545–552. https://doi.org/10.1007/s10826-015-0253-0

Dworkin, J., Connell, J., & Doty, J. (2013). A literature review of parents' online behavior. *Cyberpsychology, 7*(2), 1–10. https://doi.org/10.5817/CP2013-2-2

Faircloth, C. (2023). Intensive parenting and the expansion of parenting. In E. Lee, J. Bristow, C. Faircloth, J. Macvarish (Eds.). Parenting culture studies (pp. 33–67). Palgrave Macmillan

Forgatch, M. S., & Toobert, D. J. (1979). A cost-effective parent training program for use with normal preschool children. *Journal of Pediatric Psychology, 4*(2), 129–145. https://doi.org/10.1093/jpepsy/4.2.129

Gellatly, J., Bower, P., Hennessy, S., Richards, D., Gilbody, S., Lovell, K. (2007). What makes self-help interventions effective in the management of depressive symptoms? Meta-analysis and meta-regression. *Psychological Medicine, 37*(9), 1217–1228. https://doi.org /10.1017/S0033291707000062

Havighurst, S. S., Wilson, K. R., Harley, A. E., Kehoe, C., Efron, D., & Prior, M. R. (2013). "Tuning into kids": Reducing young children's behavior problems using an emotion coaching parenting program. *Child Psychiatry & Human Development, 44*, 247–264. https://doi.org/10.1007/s10578-012-0322-1

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, *343*, d5928. https://doi.org/10.1136/bmj.d5928

Hollon, SD (1999). Allegiance effects in treatment research: A commentary. *Clinical Psychology: Science and Practice, 6*(1), 107–112. https://doi.org/10.1093/clipsy.6.1.107

Holmes, E. A., Craske, M. G., & Graybiel, A. M. (2014). Psychological treatments: A call for mental-health science. *Nature, 511*(7509), 287–289. https://doi.org/10.1038/511287a

Jones, D. J., Loiselle, R., Zachary, C., Georgeson, A. R., Highlander, A., Turner, P., Youngstrom, J. K., Khavjou, O., Anton, M. T., Gonzalez, M., Lafko Bresland, N., & Forehand, R. (2021). Optimizing engagement in behavioral parent training: Progress toward a technology-enhanced treatment model. *Behavior Therapy, 52*(2), 508–521. https://doi.org/10.1016/j.beth.2020.07.001

Jones, P.J., Mair, P., Kuppens, S., & Weisz, J.R. (2019). An upper limit to youth psychotherapy benefit? A meta-analytic copula approach to psychotherapy outcomes. *Clinical Psychological Science, 7*(6), 1434–1449. https://doi.10.1177/2167702619858424

Kaehler, L. A., Jacobs, M., & Jones, D. J. (2016). Distilling common history and practice elements to inform dissemination: Hanf-Model BPT programs as an example. *Clinical Child and Family Psychology Review, 19*, 236–258. https://doi.org/10.1007/s10567-016-0210-5

Kaminski, J., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*, 567−589. https://doi.org/10.1007/s10802-007-9201-9

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLOS ONE, 9*(9), e105825.

Laas Sigurðardóttir, L. B., Melendez-Torres, G. J., Backhaus, S., Gardner, F., Scott, S., European Parenting Program Research Consortium, & Leijten, P. (2023). Individual participant data meta-analysis: Individual differences in mediators of parenting program effects on disruptive behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*. https://doi.org/10.1016/j.jaac.2023.11.005

Leijten, P., Gardner, F., Melendez-Torres, G. J., Van Aar, J., Hutchings, J., Schulz, S., Knerr, W., & Overbeek, G. (2019). Meta-analyses: Key parenting program components for disruptive child behavior. J*ournal of the American Academy of Child & Adolescent Psychiatry, 58*(2), 180−190. https://doi.org/10.1016/j.jaac.2018.07.900

Leijten, P., Melendez-Torres, G. J., & Gardner, F. (2022). The most effective parenting program content for disruptive child behavior: A network meta-analysis. *Journal of Child Psychology and Psychiatry, 63*(2), 132−142. https://doi.org/10.1111/jcpp.13483

Leijten, P., Scott, S., Landau, S., Harris, V., Mann, J., Hutchings, J., Beecham, J. & Gardner, F. (2020). Individual participant data meta-analysis: Impact of conduct problem severity, comorbid attention-deficit/hyperactivity disorder and emotional problems, and maternal depression on parenting program effects. *Journal of the American Academy of Child & Adolescent Psychiatry, 59*(8), 933−943. https://doi.org/10.1016/j.jaac.2020.01.023

Leijten, P., Weisz, J. R., & Gardner, F. (2021). Research strategies to discern active psychological therapy components: A scoping review. *Clinical Psychological Science, 9*(3), 307−322. https://doi.org/10.1177/2167702620978615

Leykin, Y., & DeRubeis, R. J. (2009). Allegiance in psychotherapy outcome research: Separating association from bias. *Clinical Psychology: Science and Practice, 16*, 54–65. http://dx.doi.org/10.1111/j.1468-2850.2009.01143.x

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S., & Schweizer, E. (1999). The researcher's own therapy allegiances: A “wild card” in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*(1), 95–106. https://doi.org/10.1093/clipsy.6.1.95

Ljótsson, B., Hedman, E., Mattsson, S., & Andersson, E. (2017). The effects of cognitive–behavioral therapy for depression are not falling. *Psychological Bulletin, 143*(3), 321–325. http://dx.doi.org/10.1037/bul0000055.

Lunkenheimer, E., Lichtwarck-Aschoff, A., Hollenstein, T., Kemp, C. J., & Granic, I. (2016). Breaking down the coercive cycle: How parent and child risk factors influence real-time variability in parental responses to child misbehavior. *Parenting, 16*(4), 237−256. https://doi.org/10.1080/15295192.2016.1184925

McMahon, R. J., Goulter, N., & Frick, P. J. (2021). Moderators of psychosocial intervention response for children and adolescents with conduct problems. *Journal of Clinical Child & Adolescent Psychology, 50*(4), 525−533. https://doi.org/10.1080/15374416.2021.1894566

Magnusson, K. (2022). Interpreting Cohen's d effect size: An interactive visualization (Version 2.6.0) [Web App]. R Psychologist. https://rpsychologist.com/cohend/

Mikolajczak, M., Gross, J. J., & Roskam, I. (2019). Parental burnout: What is it, and why does it matter?. *Clinical Psychological Science, 7*(6), 1319−1329. https://doi.org/10.1177/2167702619858430

Mingebach, T., Kamp-Becker, I., Christiansen, H., & Weber, L. (2018). Meta-meta-analysis on the effectiveness of parent-based interventions for the treatment of child externalizing behavior problems. *PLOS ONE, 13*(9), e0202855. https://doi.org/10.1371/journal.pone.0202855

Mohr, D. C., Ho, J., Hart, T. L., Baron, K. G., Berendsen, M., Beckner, V., Cai, X., Cuijpers, P., Spring, B., Kinsinger, S. W., Schroder, K. E., & Duffecy, J. (2014). Control condition design and implementation features in controlled trials: A meta-analysis of trials evaluating psychotherapy for depression. *Translational Behavioral Medicine, 4*(4), 407−423. https://doi.org/10.1007/s13142-014-0262-3

Morelli, G., Quinn, N., Chaudhary, N., Vicedo, M., Rosabal-Coto, M., Keller, H., Murray, M., Gottlieb, A., Scheidecker, G., & Takada, A. (2018). Ethical challenges of parenting interventions in low-to middle-income countries. *Journal of Cross-Cultural Psychology, 49*(1), 5−24. https://doi.org/10.1177/0022022117746241

Patterson, G. R. (1965). An application of conditioning techniques to the control of a hyperactive child. In L. Ullman and L. Krasner (Eds.), Case studies in behavior modification. Holt, Rinehart, & Winston

Patterson, G. R. (1982). Coercive family process. Castalia

Patterson, G. R. (2002). The early development of coercive family process. In J. Reid, G. R. Patterson, D. Snyder (Eds.), Antisocial behavior in children and adolescents: A developmental analysis and model for intervention (pp. 25−44). American Psychological Association

Pew Research Center (2015). Parenting in America. https://www.pewresearch.org/wp-content/uploads/sites/20/2015/12/2015-12-17_parenting-in-america_FINAL.pdf

Readdick, C. A., & Chapman, P. L. (2000). Young children's perceptions of time out. *Journal of Research in Childhood Education, 15*(1), 81−87. https://doi.org/10.1080/02568540009594777

Reitman, D., & McMahon, R. J. (2013). Constance “Connie” Hanf (1917–2002): The mentor and the model. *Cognitive and Behavioral Practice, 20*(1), 106−116. https://doi.org/10.1016/j.cbpra.2012.02.005

Rifkin, A. (2007). Randomized controlled trials and psychotherapy research. *American Journal of Psychiatry, 164*(1), 7−8. https://doi.org/10.1176/ajp.2007.164.1.7

Siegel, D., & Bryson, T. P. (2014). “Time-outs” are hurting your child. Time. Retrieved from http://time.com/3404701/disciplinetime-out-is-not-good

Sim, A., Puffer, E., Green, E., Chase, R., Zayzay, J., Garcia-Rolland, E., & Boone, L. (2014). Parents make the difference: Findings from a randomized impact evaluation of a parenting program in rural Liberia. International Rescue Committee

Smith, J. D., Dishion, T. J., Shaw, D. S., Wilson, M. N., Winter, C. C., & Patterson, G. R. (2014). Coercive family process and early-onset conduct problems from age 2 to school

entry. *Development and Psychopathology, 26*(4pt1), 917−932. https://doi.org/10.1017/S0954579414000169

Stattin, H., Enebrink, P., Özdemir, M., & Giannotta, F. (2015). A national evaluation of parenting programs in Sweden: The short-term effects using an RCT effectiveness design. *Journal of Consulting and Clinical Psychology, 83*(6), 1069–1084. https://doi.org/10.1037/a0039328

Stolk, M. N., Mesman, J., van Zeijl, J., Alink, L. R., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Juffer, F., & Koot, H. M. (2008). Early parenting intervention: Family risk and first-time parenting related to intervention effectiveness. *Journal of Child and Family Studies, 17*(1), 55–83. https://doi.org/10.1007/s10826-007-9136-3

Webster-Stratton, C. (1982). The long-term effects of a videotape modeling parent-training program: Comparison of immediate and 1-year follow-up results. *Behavior Therapy, 13*(5), 702−714. https://doi.org/10.1016/S0005-7894(82)80026-9

Weisz, J. R., & Kazdin, A. E. (2017). Evidence-based psychotherapies for children and adolescents. Guilford Press.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., Jensen-Doss, A., Hawley, K. M., Krumholz, L. S., Chu, B.C., Weersing, V. R., & Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist, 72*, 79−117. https://dx.doi.org/10.1037/a0040360

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science, 14*(2), 216−237. https://doi.org/10.1177/17456916188054

Wells, K. C. (1997). The death of discipline: is the requiem premature?. *Aggression and Violent Behavior, 2*(4), 337−341. https://doi.org/10.1016/S1359-1789(97)00019-0

World Bank (2022). The world bank in middle income countries. Retrieved from https://www.worldbank.org/en/country/mic/overview#1 Accessed on March 28th, 2023

Table 1

Associations of Trial, Sample, and Intervention Characteristics with Time and Effect Size

	Association with time β [95% CI]	Association with effect size β [95% CI]
Trial characteristics		
Sample Size *N* at baseline	0.01 [–0.001, 0.03]	**0.001 [0.001, 0.002]**
Active control condition *0=no; 1=yes*	**3.91 [1.15, 6.66]**	**0.18 [0.08, 0.28]**
Developer involvement *0=no; 1=yes*	**–4.22 [–6.35, –2.09]**	–0.004 [–0.13, 0.12]
Preregistration *0=no; 1=yes*	**9.36 [7.71, 11.02]**	0.11 [–0.01, 0.23]
Risk of bias *0=low; 1=unclear; 2=high*		
Random sequence allocation	**–5.56 [–7.91, –3.20]**	–0.01 [–0.12, 0.10]
Random sequence concealment	**–3.51 [–5.52, –1.50]**	–0.07 [–0.17, 0.03]
Blinding of participants	–1.51 [–9.74, 6.73]	–0.03 [–0.27, 0.21]
Blinding of outcome assessors	**9.33 [4.80, 13.87]**	0.09 [–0.12, 0.03]
Incomplete data	**–3.39 [–5.78, –1.00]**	–0.09 [–0.17, 0.03]
Selective outcome reporting	**–6.78 [–8.90, –4.66]**	**–0.11 [–0.21, –0.01]**
Sample Characteristics		
Child age *Mean sample age in years*	**0.71 [0.08, 1.34]**	0.02 [–0.01, 0.04]
Child sex *Proportion of boys*	**–16.92 [–26.03, –7.82]**	**–0.51 [–1.02, –0.01]**

Child disruptive behavior severity	**−0.06 [−0.10, −0.003]**	**−0.004 [−0.01, −0.001]**
Baseline ECBI in intervention condition (48% of trials)		
Family socioeconomic status	−0.75 [−0.26, 1.75]	−0.03 [−0.14, 0.08]
0=primarily low; 1=not primarily low		
Cross-cultural transportation	**5.50 [3.39, 7.60]**	−0.01 [−0.12, 0.11]
0=similar setting; 1=different country/culture/ethnic group		
Intervention characteristics		
Number of sessions	−0.15 [−0.32, 0.02]	0.003 [−0.01, 0.01]
Number of sessions offered		
Delivery mode		
Individual *(vs group)*	−2.17 [−5.36, 1.01]	−0.08 [−0.23, 0.06]
Therapist-led *(vs self-directed)*	−0.93 [−4.88, 3.02]	0.03 [−0.14, 0.20]
In vivo coaching *(1=yes; 0=no)*	−1.75 [−6.83, 3.34]	−0.19 [−0.45, 0.08]
Video feedback *(1=yes; 0=no)*	−2.14 [−16.78, 12.50]	**0.23 [0.03, 0.43]**
Components *(1=yes; 0=no)*		
Positive reinforcement	Insufficient variation	Insufficient variation
Avoiding reinforcement	−5.52 [−13.78, 2.73]	−0.02 [−0.56, 0.53]
Time-out	**−4.13 [−7.75, −0.51]**	−0.09 [−0.27, 0.09]
Ignore	−1.81 [1.14, 4.76]	−0.12 [−0.25, 0.01]
Consequences	0.83 [−3.30, 4.97]	−0.04 [−0.20, 0.13]
Clear rules and instructions	−0.61 [−5.51, 4.29]	0.05 [−0.14, 0.24]
Relationship building	2.37 [−0.67, 5.41]	0.06 [−0.08, 0.19]
Child-led play	−2.39 [−4.93, 0.15]	−0.05 [−0.18, 0.08]
Parental self-management	−1.06 [−3.54, 1.42]	0.14 [0.03, 0.25]
Parents teaching children skills	−5.52 [−13.78, 2.73]	0.06 [−0.06, 0.17]

Note. Coefficients in **bold** are significant at $\alpha = .05$. Coefficients should be interpreted as the increase in effect size (in Cohen's *d*) or year per one unit increase in the predictor. Because publication year was divided by 10, an increase of one unit on year reflects an increase of 10 years. [a]risk of bias coded as 0 = low, 1 = unclear, 2 = high. [b]blinding of outcome assessors coded as 0 = parent-reported measure of disruptive child behavior, 2 = blinded observational measure of disruptive child behavior.

Table 2

Trial and Sample Characteristics Explaining Time Trends in Effect Sizes

	β [95% CI]
Individual predictor models	
Indicates that none of the individual predictors explains away the time trend	
Year (linear)	0.04 [–0.03, 0.11][a]
Year (quadratic)	–0.04 [–0.09, 0.001][a]
Sample size	0.001 [0.0004, 0.001]
Year (linear)	0.04 [–0.04, 0.11]
Year (quadratic)	–0.07 [–0.11, –0.02]
Active control	0.17 [0.06, 0.27]
Year (linear)	0.04 [–0.04, 0.12]
Year (quadratic)	–0.06 [–0.10, –0.01]
Risk of selective outcome reporting	–0.07 [–0.18, 0.04]
Year (linear)	0.18 [0.07, 0.29]
Year (quadratic)	–0.01 [–0.09, 0.07]
Sample severity	–0.003 [–0.01, –0.0003]
Year (linear)	0.06 [–0.03, 0.14]
Year (quadratic)	–0.06 [–0.10, –0.01]
Proportion of boys	–0.32 [–0.84, 0.20]
Combined predictor model	
Indicates that when combined in one model, the predictors explain away the time trend	
Year (linear)	0.11 [–0.05, 0.27]
Year (quadratic)	–0.02 [–0.10, 0.07]
Sample size	0.001 [–0.001, 0.002]
Active control	0.14 [–0.05, 0.37]
Risk of selective outcome reporting	–0.03 [–0.24, 0.17]

Sample severity	−0.002 [−0.01, 0.002]
Proportion of boys	−0.91 [−1.83, 0.01]

Note. [a]The individual year variables (i.e., linear and quadratic year) were not significant, but the joint contribution of time was ($\chi^2 = 10.92$, $p = 0.004$). This was not the case for the combined predictors model ($\chi^2 = 3.56$, $p = 0.169$). In the combined predictors model, the time trend in effect sizes was no longer significant.

Figure 1

Flowchart of Data Collection Procedure

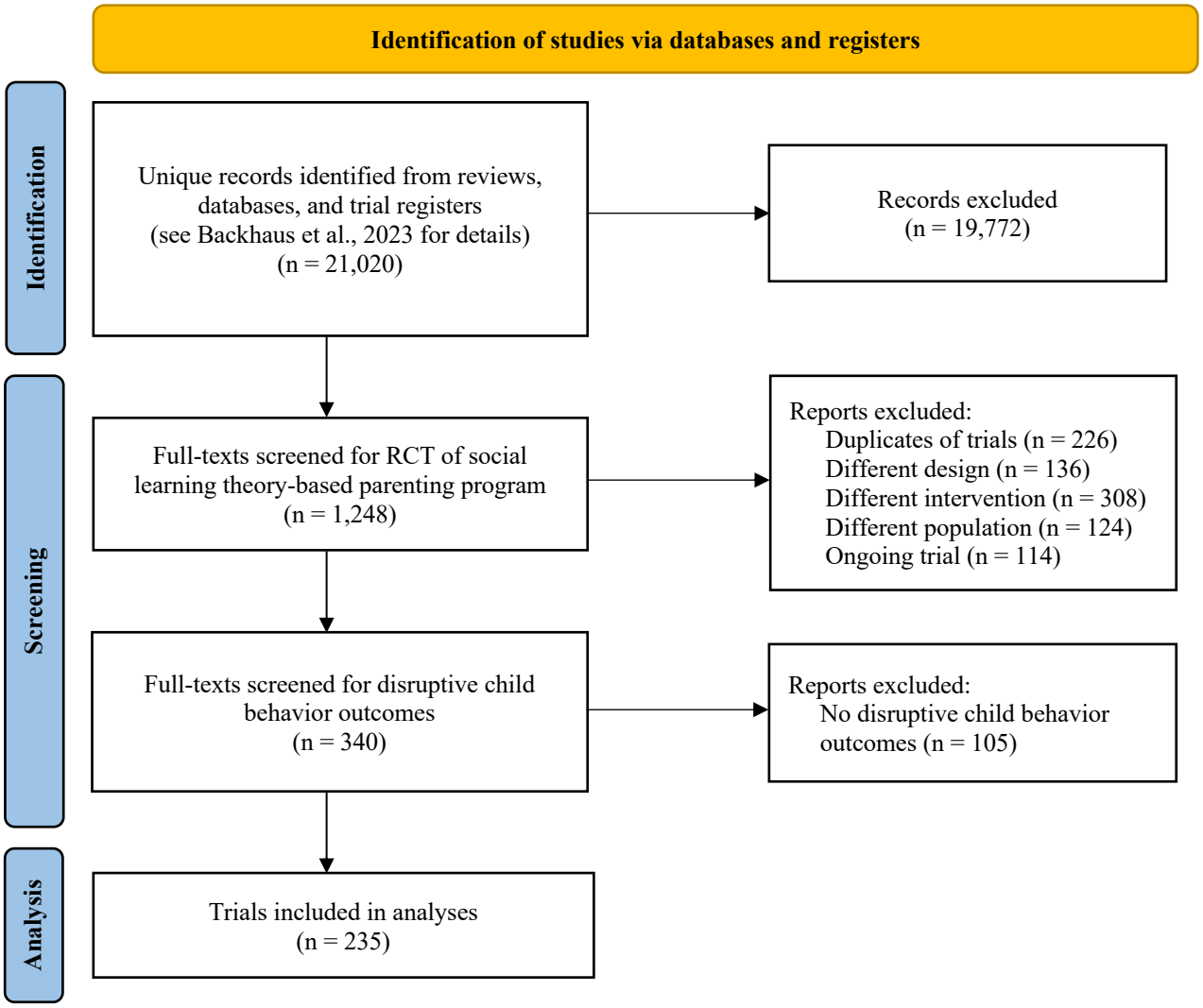

Figure 2

Time Trends in Parenting Program Effects (Cohen's d) on Disruptive Child Behavior (Bubble Size Reflects Number of Trials)