



UNIVERSITY OF
CAMBRIDGE

Institute of Criminology
Violence Research Centre

Violence Research Centre - Working Paper

25 May 2014

The South Carolina Triple P System Population Trial to Prevent Child Maltreatment: Seven Reasons to be Sceptical about the Study Results

Manuel Eisner

Professor of Comparative and Developmental Criminology

Institute of Criminology - Violence Research Centre

University of Cambridge

CB3 9DA Cambridge

Email: mpe23@cam.ac.uk

This is a working paper. Working papers are in draft form and are distributed to stimulate discussion and critical comment. It may not be reproduced without permission of the copyright holder.

Copyright © 2014 by Manuel Eisner

Abstract

In 2009, Prinz, Sanders, Shapiro, Witzker and Lutzker published the paper “Population-Based Prevention of Child Maltreatment: The U.S. Triple P System Population Trial” in *Prevention Science*. The study suggested a highly significant impact of a relatively inexpensive universal parent training intervention on child maltreatment. The findings have been received very positively amongst prevention practitioners and researchers, and have become the basis for recommendations to disseminate the programme widely. In the present paper I argue that considerable weaknesses in the original paper make it impossible to assess with confidence whether the intervention has had any effects. It recommends that the authors make the primary data available for replication analyses by other researchers.

In 2009, Prinz, Sanders, Shapiro, Witaker and Lutzker published the paper “Population-Based Prevention of Child Maltreatment: The U.S. Triple P System Population Trial” in the journal *Prevention Science*. The study reported findings from a population-level randomized trial conducted in South Carolina. It claimed to conclusively demonstrate that implementing a low-cost parenting intervention (The Triple P System) in nine of the 18 participating counties resulted in large positive effects at the county level for three population indicators, namely substantiated child maltreatment ($d = 1.09$), child out-of home placements ($d = 1.22$), and child maltreatment injuries ($d = 1.14$).

Prinz et al. (2009) is one of the most influential studies in prevention science over the past five years: It had a very positive reception in the academic community as a milestone in population level approaches to the reduction of child maltreatment; it attracted substantial attention in the media including a call by New York Times commentator David Bornstein (2013) to introduce Triple P as a program that has proven population-level effectiveness; and it led to the quick acceptance of the Triple P system in several influential recommendation lists in the United States including SAMSHA’s National Registry of Evidence-Based programs (SAMSHA, 2013) and the Centres for Disease Control and Prevention, CDC (Centres for Disease Control and Prevention, 2013). The *Washington State Institute for Public Policy* recently calculated, based on the study, that every US-Dollar invested into Triple P saves an average of six dollars in costs (Lee et al., 2012).

In the United States, the case for disseminating Triple P at a population level largely rests on the South Carolina study. In what follows I will present seven arguments why academics and policy-makers should be sceptical about its findings.

The Study

While several papers have been published on the Triple P System Population Trial (TPSPT) in South Carolina (Prinz et al., 2009; Sanders, 2008, Shapiro et al., 2008) – a single study was involved. The main details on its design and outcomes can be found in Prinz and Sanders (2007) and Prinz et al. (2009). The study tested the effectiveness of the Triple P--Positive Parenting Program, a parent training program developed by Sanders and colleagues at the University of Queensland, Australia. The Triple P system comprises five levels representing levels of program intensity, from media and information strategies (level 1) to individual and group-based series of 4-10 sessions of parent training (levels 4 and 5). The South Carolina TPSPT was designed to test whether the implementation of the full Triple P system would have population-level beneficial effects on child maltreatment.

The study population were all 18 counties in South Carolina with a population of between 50,000 and 175,000 (out of a total of 42 counties). The counties were first matched in pairs by population size, county poverty rate, and county child abuse rate (Prinz and Sanders, 2007). Counties were then randomly allocated to the treatment and the control condition, creating 9 pairs of counties. In the treatment condition, which comprised about 85,000 parents with children aged 0-8, the Triple P System was disseminated over a two year period, using a variety of dissemination strategies. In the process 21 Triple P newspaper articles were published, 26,000 newsletters were sent to parents, and approximately 650 Triple P service providers were trained.

Baseline and post measures reported in Prinz et al. (2009) are based on official records of child maltreatment, in particular substantiated cases of child maltreatment, out-of-home placements, and child maltreatment injuries. According to their analyses the authors claim that the study demonstrates “a positive impact of Triple P on population indicators of CM” and that “the results of this study demonstrate the feasibility of implementing a large-scale parenting intervention” (Prinz et al., 2009: 9). Furthermore, they calculate that in a community with 100,000 children under 8 years of age, these effects would translate into 688 fewer cases of CM, 240 fewer out-of-home placements, and 60 fewer children with injuries requiring hospitalization or emergency room treatment (Prinz et al., 2009: 9).

Seven Reasons to be Sceptical

In what follows I argue that these conclusions are not supported by convincing evidence, and that it is impossible to assess whether the intervention was effective. I will organize my comments around seven main points. In doing this, I will use criteria for assessing the quality of evaluation studies laid out in Shadish, Cook and Campbell (2002), Farrington (2003), and the reporting guidelines specified in the CONSORT statement (e.g., Altman et al., 2010).

1. Selective reporting on study design and outcomes

The first issue relates to inconsistencies in the description of the study design and the collected outcome measures. Two publications describe the study design: A project overview was published in 2007 in *Clinical Psychology Review* (Prinz and Sanders, 2007:746-748). It reports the study design, the data collection, and the planned outcome measures at a time when the outcomes were not yet known but the study was already underway since several years. The outcome study Prinz et al.

(2009), published in *Prevention Science*, similarly comprises a description of the study design, data collection, and outcome measures. A comparison of the two publications reveals substantial unexplained discrepancies. They relate to the study design, the data collection, and the planned outcomes.

Table 1 *TPSPT study design as reported in Prinz and Sanders (2007) and findings reported in Prinz et al (2009).*

Criterion	Reported in Prinz and Sanders (2007:746-748)	Reported in Prinz et al. (2009)
Randomization	Pair-wise matching	Stratified random assignment
Age range	0-7	0-8
Population Survey	Annual assessments	A pre and a post intervention survey
Baseline reference period for official data	5 years before the intervention	Last year before the intervention
Outcomes – Survey based	Parent knowledge of Triple P	Reported
	Parent involvement in Triple P	[Not reported]
	Relationship between demographics and participation	[Not reported]
	Parenting practices	[Not reported]
	Parent confidence and stress	[Not reported]
	Child maladjustment	[Not reported]
Outcomes – Archival Records	Reported CM	(supplementary analysis)
	Substantiated physical maltreatment	[not reported]
	Substantiated neglect	[not reported]
	Substantiated sexual abuse	[not reported]
	Substantiated overall CM	Reported
	Out-of-home placements	Reported
	[not planned]	Child CM injuries

First, Prinz and Sanders (2007: 8) describe the study design as a matched pair approach, whereby counties “were matched up in pairs”, while Prinz et al. (2009) describe the study as based on “stratified random assignment”. The difference matters because it influences the appropriate statistical approach (see below, point 4).

Second, Prinz and Sanders (2007) report that the baseline period to collect archival records is “the five years preceding the start of the trial to provide an extended baseline”. In Prinz et al. (2009) the statistical analyses were conducted using as baseline “the year just prior to the initiation of intervention”. 5-year averages were reported, but not used in the statistical analyses. This difference is relevant because the change in the baseline period has – as I will explain below, see point 3 - likely inflated the estimated treatment effect.

Third, the two publications refer to vastly different outcome measures. The description of the ‘measurement process’ in Prinz and Sanders (2007: 9) specifies that “a telephone survey of caregivers in households with children ages one to seven years is conducted annually in the 18 counties to assess media and informational exposure to Triple P, parent involvement in parenting consultation and support (generally and also specifically through Triple P), parenting practices, parental confidence and stress, and reports of child adjustment.” In Prinz et al. (2009) this extensive survey is no longer mentioned. Instead, the authors refer to a short pre- and post-test survey “to determine relative growth in public awareness of Triple P” (Prinz et al., 2009: 6). This discrepancy matters because no information is given about what happened to the survey. Thus, the survey described in Prinz and Sanders (2007) would permit analyses of whether the high-risk target groups were effectively reached by Triple P, whether self-reported parenting practices could be changed, and whether child adjustment problems were improved. No such analyses have been reported to date.

Also, the list of outcome variables based on archival data changed between the two publications. None of the planned analyses by sub-dimensions of child abuse (i.e. physical maltreatment, neglect, sexual abuse, and combinations of these) described in Prinz and Sanders (2007) are reported in Prinz et al. (2009). Also, “reported cases of child maltreatment” were a main planned outcome in Prinz and Sanders (2007), but become relegated to a separate section on ‘supplementary analyses’ in Prinz et al. (2009: 8). No significant effect was found for this outcome. In contrast, a new outcome indicator (child maltreatment injuries) is reported in Prinz et al. (2009), which was not mentioned in Prinz and Sanders (2007). This new outcome had a positive effect. In sum, out of 12 outcomes mentioned in Prinz and Sanders (2007) only three were reported in Prinz et al. (2009), and one new indicator was introduced that was not initially planned.

2. Poor information on data collection and data-sources

The second point relates to the lacking clarity of the study description. Readers of the Prinz et al. (2009) paper will try in vain to find information about the calendar years in which the intervention was delivered, the year or years in which the baseline measures were taken, the period in which the post measures were taken, and how many years elapsed between the pre and the post measures. The missing information makes it difficult to understand the precise sequence of measurement, randomization, dissemination, and outcome measurement; it prevents readers from examining whether history effects (i.e. relevant events in any of the counties during the trial period) may have affected the results; and it is an obstacle to attempts to access the data needed to replicate the study. In what follows data on the timing of the intervention and the baseline and post assessments were therefore taken from Prinz (2006), a PowerPoint presentation that shows the respective information.

The TPSPST-study also provides no references on the source of the socio-demographic characteristics of the counties and no source on the origins of the three main archival population indicators. The only information provided is that the data “were derived from independent data-collection systems deposited with a state-run statistical division” (Prinz et al. 2009: 6). Full information on the data sources constitutes an important principle of academic research. Lacking information is a barrier to replicating the study findings by accessing the official sources used by the authors.

3. Dubious claims for baseline equivalence

Demonstrating the extent to which baseline equivalence has been achieved is important for any randomized trial. For the TPSPST-study is currently impossible to assess whether equivalence was effectively achieved. Table 1 in the TPSPST-study (Prinz et al., 2009: 7) presents the control and treatment means for a number of variables averaged over five years prior to the intervention (1999-2003). The authors claim that the data shown that the TG (treatment group) and CG (control group) counties “did not differ significantly”. However, the baseline means presented in table 1 are different from the values that were effectively used as baseline measures for calculating the treatment effects (see Table 3 in Prinz et al., 2009: 8). For some variables the differences between the two tables are substantial: In particular, the 5-year average rate of out-of-home placements in Table 1 hardly differs between the control and the treatment condition ($TG = 4.02$ v. $CG = 3.76$ per 1000 children). In

contrast, the means used for calculating the intervention effects suggest a substantial difference at baseline ($TG = 4.27$ v. $CG = 3.10$).

The most likely reason for the discrepancy between the two tables is that Table 3 refers to the rates from “the year just prior to the initiation of intervention” (Prinz et al., 2009: 8), while Table 1 refers to five-year averages. This casts doubt on the validity of the findings for several reasons: First, testing equivalence for one set of baseline variables and then using a different set of variables as baseline covariates is not advisable - baseline equivalence should be assessed on the same variables used in the statistical model that tests treatment effects. Second, the large difference at baseline for out-of-home placements makes regression to the mean a methodological issue and the estimation of effect sizes based on difference scores problematic. Third, if there was not similarity at baseline, then the issue arises of how to control for it. If there were a stable time-invariant group difference, then the 5-year average is the best indicator of that. But if there were a time-varying group difference prior to treatment, then slopes would have to be included in the model of the outcome. (The difference between a CG mean of 3.76 in the year immediately prior to treatment and of 3.10 averaged over 5 years suggests that time-varying group differences are not totally implausible). Prinz et al. (2009) did neither of the above; they used only the immediate group pre-test means. In a small sample RCT, baseline equivalence is often a problematic issue for which a clear resolution is not at hand.

4. Contradictory description of outcome analysis

According to Prinz and Sanders (2007) the 18 counties were first matched in pairs, and then one county was randomly allocated to either treatment or control condition. If this was indeed the procedure, then the main results table 3 reports the wrong significance tests (Martin et al., 1993). For matched pairs the paired version of Student's *t* test needs to be used, which has $n/2-1$ degrees of freedom. In the present study this means that the study had 8 degrees of freedom rather than 16 as reported in Table 3 (Prinz et al. 2009: 8). This is important because the smaller degrees of freedom would result in different significance levels and confidence intervals for the outcome results.

Prinz and Sanders (2009) also offer two different descriptions of how the data were analysed. In the main text the authors say that “for each of the indicators, post-intervention rates for the Triple P System and Control conditions were compared, controlling for pre-intervention rate from the year just prior to the initiation of intervention” (Prinz et al., 2009: 8). This probably suggests an

ANCOVA design whereby the slope coefficients for TG v CG were estimated. The findings in the main results table 3 relate to a different test, namely Student's t-tests that "compared the two conditions with respect to pre-post difference scores" (Prinz et al., 2009: 8). This limits the ability to draw firm conclusions about the intervention effects. Baseline covariate-adjusted tests and difference-score tests can lead to different results and findings are impossible to assess without clarity about whether both tests lead to the same or to different conclusions.

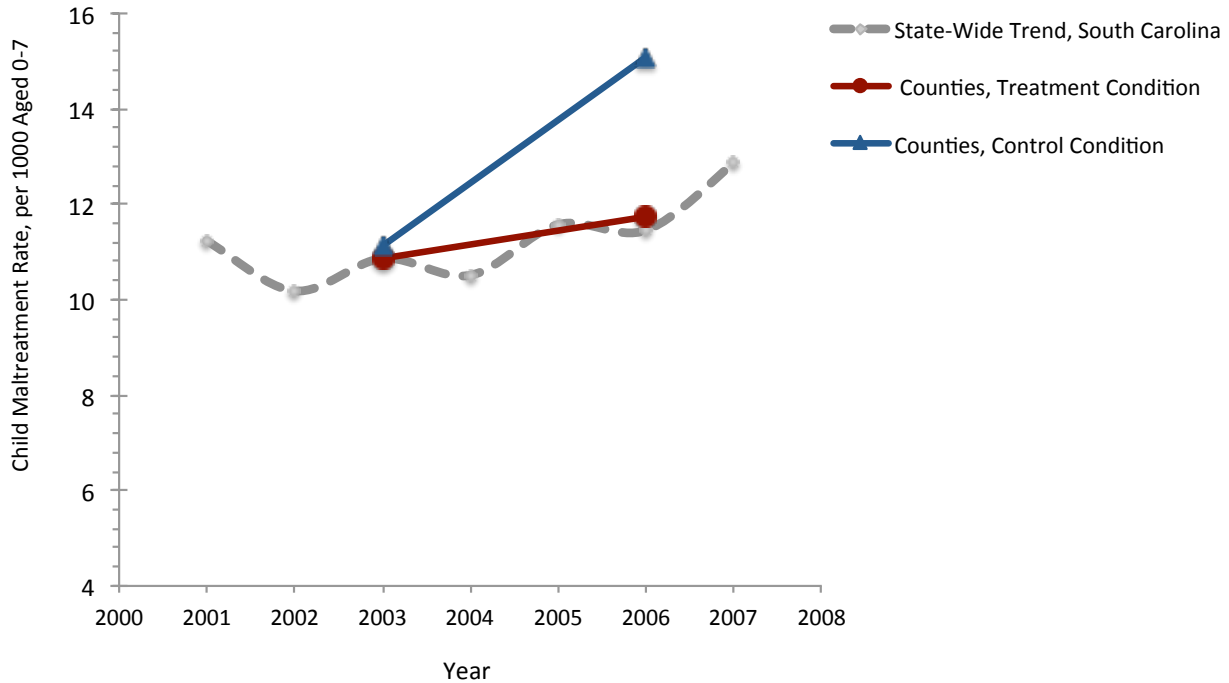
5. Lacking comparison with state-wide trends

It is important to note that in the Triple P counties the number of substantiated cases of child maltreatment did not effectively decrease. Rather, the Triple P counties experienced a slight increase from 10.86 to 11.74 cases per 1000 children at ages birth to 8 (Prinz et al., 2009: 8). The treatment effect on child maltreatment exclusively results from the stronger increase in the counties that were allocated to the control condition, where child maltreatment cases grew from 11.12 cases per 1000 at baseline to 15.06 cases at post.

The authors of the study address this issue in the discussion section. They claim "that the increase in substantiated CM in the control counties mirrored similar increases across the other 28 counties in the same state that were not part of the TPSP" (Prinz et al., 2009: 9). However, they do not provide the data to substantiate their claim. I therefore assessed the claim by examining age-specific data on substantiated child maltreatment cases for the whole of South Carolina (42 counties), which are publicly available in the *Child Maltreatment* statistical volumes published by the U.S. Department of Health and Human Services, Administration on Children, Youth and Families. The state wide trends for substantiated child maltreatment of children aged 0-7 are displayed in Figure 1, along with the means reported for the TG and the CG in Prinz et al. (2009).

Figure 1 shows that the counties in the Triple P condition developed perfectly in line with the state-wide trend in substantiated child maltreatment cases. There is no evidence that the rate of child maltreatment cases in the Triple P counties increased less than what would be expected on the basis of state-wide trends. The treatment effect therefore seems to be entirely due to an atypical increase of child-maltreatment cases in the control condition, where the rate of child maltreatments increased by over 35%. It is unclear why these counties experienced an increase that was above the state-wide trend, but the pattern suggests that their value as a plausible counterfactual is limited.

Figure 1 *Substantiated child maltreatment cases (per 1000 children) in Triple P counties, control counties, and South Carolina as a whole*



Sources: See Appendix I for the detailed figures used in this graph.

Note: State-wide child maltreatment rates have been multiplied by a constant factor of .85 so that they are indexed to be identical to the child maltreatment rate in the treatment condition in 2003.

It is important to note that there may be limits in the analysis conducted here. In particular, Prinz et al. (2009) report that they eliminated double counts in their data, which may not be the case in the published statistics. However, there is no reason to believe that this would significantly alter the picture.

6. *Lacking face validity*

Experts working in the field agree that child maltreatment is a complex condition that has multiple roots, which are very difficult to address by way of preventive interventions and treatment (e.g. MacMillan et al., 2009). Many experts would also expect that substantial parts of a target

population must be reached in order to achieve large population-level effects (e.g. Glasgow et al., 2009). The TPSPT-study presents data on the percentage of respondents who ‘had heard’ about Triple P. In the treatment condition 17.1% of families with 0-7 year old children had ‘heard of’ Triple P, as compared to 5.5% in the control condition. The study also reports the estimated number of families exposed to Triple P *according to the service providers*: “It was estimated that between 8,883 and 13,560 families participated in Triple P within the Triple P System counties (Prinz et al., 2009: 6). This corresponds to about 10-15% of the target population, if the dissemination exclusively targeted parents with children at ages 0-7. However, the providers “were trained in Triple P for a broader age range (birth to 12 years) consistent with the inherent breadth of Triple P so that providers who straddled the target age (e.g., 5–12 years) in terms of the families they served would find Triple P of greater utility in their work.” (Prinz et al., 2009: 4). This implies that some training was delivered to parents outside the target group, diluting the maximum effective exposure in the target group. In any event, it is not easy to argue that treatment exposure was widespread either to the targeted age 0-7 population or to the age 0-12 population.

TPSPT claims large benefits from treatment exposure of 10-15% of a universal target population. Yet Triple P is a short intervention that mainly entails tip sheets, seminars to large audiences, and brief consulting sessions. So the claim is that this modestly intensive intervention delivered to at most one out of eight families is sufficient to reduce child maltreatment by between 28% (substantiated child abuse) and 44% (out of home placements) at the population. While logically not impossible, the claim lacks face validity: First, no theoretical or empirical arguments are presented about the causal mechanisms through which such a change amongst multi-risk parents can be achieved with such a low penetration of the target population. Second, data from the population survey, which could shed light on change in parenting practices and child problem behavior are not presented as supportive evidence. This raises concerns about whether core potential proximal mechanisms were effectively affected by the intervention.

7. *Failure to declare conflict of interest*

The journal *Prevention Science* and the *Society of Prevention Research* have a policy which requires that authors declare their conflicts of interest in work published in the journal. The website states that “such conflicts of interest may arise out of commitments involving honoraria, consultant relationships, participation in a speakers bureau, stock holdings or options, royalties, ownership of a company or patent, research contracts or grants, and, in some instances, being an official

representative of another organization” (Society of Prevention Research, 2013). For both Prof Prinz and Prof Sanders these conditions seem to be fulfilled. Prof Prinz is a member of the advisory board of *Triple P International* and acts as a consultant to Triple P. Prof Sanders is the developer of Triple P and director of the *Parenting and Family Support Centre* at the University of Queensland. *Triple P International* is a private company and royalties from the international dissemination of Triple P are distributed to, amongst others, the *Parenting and Family Support Centre* of the *University of Queensland* (Triple P International, 2013). It would therefore have been adequate to clarify the roles of the two principal investigators in the dissemination of Triple P.

Conclusions

Given the issues identified here it seems advisable to interpret the findings presented in Prinz et al. (2009) with scepticism. In fact, I believe that the information presently available does not permit any conclusions about whether the intervention had the claimed effects. To obtain a better picture it would be desirable that the authors made all study data, including all information collected as part of the population survey, publicly available for a replication study.

The arguments presented here lead to a number of difficult questions: Why did the reviewers fail to notice important limitations of the manuscript, allowing authors to report the findings more in line with current methodological and reporting standards? Why did the funding agency accept the study as the highest evidence for intervention effectiveness, seemingly without critically evaluating the study design against the reported outcomes? Why have hundreds of colleagues cited the study, giving policy makers and the wider public the impression of robust evidence for the effectiveness of a relatively low-dose preventive programme against the most serious forms of child maltreatment? And why did nobody compare the Prinz et al. (2009) paper with a related paper published two years before (Prinz and Sanders, 2007), the comparison of which highlights a range of possible inconsistencies between study design and measurement plans and realizations?

Several lessons can be learned from this, many of which are increasingly recognized amongst prevention researchers and promoted by professional organizations (e.g. Mayo-Wilson, 2013): Compulsory registration of trials including a study protocol will reduce the risk of post-hoc selective reporting; insistence by journals on adherence to reporting standards such as those described in the various CONSORT statements (e.g. Altman, 1996) can improve the descriptive validity of trials; compulsory reporting of conflicts of interests across all prevention science journals can make the relationship between the authors and the evaluated programme more transparent; compulsory

publication of all data used in a study will enable readers to critically examine and replicate the findings; funding agencies can insist on only funding studies that involve at least one external academic in the data-analysis as a control and monitoring mechanism; and scholarly journals have to somehow increase the level of scrutiny of papers, or at least the scrutiny of papers that make claims about particularly large effects, or that seem to push ahead the field in ways everybody wants but that can lull critical faculties.

References

- Altman, D. G. (1996). Better reporting of randomised controlled trials: the CONSORT statement. *BMJ: British Medical Journal*, 313(7057), 570.
- Bornstein, David (2013). The Benefits of Positive Parenting. The New York Times Opinion pages, <http://opinionator.blogs.nytimes.com/2013/02/20/helping-the-parents-to-spare-the-children> (last accessed 4 Sept 2013).
- Centers for Disease Control and Prevention (2013). Child Maltreatment: Prevention Strategies (<http://www.cdc.gov/violenceprevention/childmaltreatment/prevention.html>), (last accessed 4 Sept. 2013)
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *The Annals of the American Academy of Political and Social Science*, 587(1), 49-68.
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American Journal of Public Health*, 89(9), 1322-1327.
- Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., Anderson, L. (2012). *Return on investment: Evidence-based options to improve statewide outcomes (Report 12-04-1201)*. Olympia: Washington State.
- MacMillan, H. L., Wathen, C. N., Barlow, J., Fergusson, D. M., Leventhal, J. M., & Taussig, H. N. (2009). Interventions to prevent child maltreatment and associated impairment. *The Lancet*, 373(9659), 250-266.
- Martin, D. C., Diehr, P., Perrin, E. B., & Koepsell, T. D. (1993). The effect of matching on the power of randomized community intervention studies. *Statistics in Medicine*, 12(3-4), 329-338.
- Mayo-Wilson, E., Grant, S., Hopewell, S., Macdonald, G., Moher, D., & Montgomery, P. (2013). Developing a reporting guideline for social and psychological intervention trials. *Trials*, 14(1), 1-7.
- Prinz, R. J., Sanders, M. R., Shapiro, C. J., Whitaker, D. J., & Lutzker, J. R. (2009). Population-based prevention of child maltreatment: The US Triple P system population trial. *Prevention Science*, 10(1), 1-12.
- Prinz, R. (2006). CDC-Funded Triple P System Population Trial (powerpoint presentation, <http://www.cimh.org/downloads/handouts/Triple%20P%20Briefing%20May06%20PART3.ppt>).
- Prinz, R. J., & Sanders, M. R. (2007). Adopting a population-level approach to parenting and family support interventions. *Clinical Psychology Review*, 27(6), 739-749.
- SAMSHA (2013). Triple P Positive Parenting Program. National Repository of Evidence-Based Programs and Practices (<http://www.nrepp.samhsa.gov/ViewIntervention.aspx?id=1>), last accessed 4 Sept. 2013).
- Sanders, M. R., & Prinz, R. J. (2008). Ethical and Professional Issues in the Implementation of Population-Level Parenting Interventions. *Clinical Psychology: Science and Practice*, 15(2), 130-136.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA, US: Houghton, Mifflin and Company.
- Society of Prevention Research (2013). Conflict of Interest and Disclosure Statement Policy, http://www.preventionresearch.org/history_conflict.php (last accessed 2 Sept 2013).
- Triple P International (2013). Behind the Scenes - Triple P International and the University of Queensland, <http://www.triplep.net/glo-en/behind-the-scenes/triple-p-international-and-the-university-of-queensland> (last accessed 2 Sept 2013).

Appendix I Data used for estimating State-wide Trends of Child Maltreatment in South Carolina

Year	Child Maltreatment Cases, South Carolina, Ages 0-7		Population Age 0-7	Rate per 1000 Children		Prinz et al (2009) ^a	
	(1)	(2)		Based on (1)	Based on (2)	TG ^b	CG ^b
2001	5,692	N/A	433,000	13.15	N/A		
2002	5,192	N/A	435,000	11.94	N/A		
2003	5,571	N/A	437,622	12.73	N/A	10.86	11.12
2004	5,385	5,133	437,077	12.32	11.74		
2005	6,033	5,731	444,722	13.57	12.89		
2006	6,070	5,743	454,060	13.37	12.65	11.74	15.06
2007	N/A	7,032	466,180	N/A	15.08		
Change, 2003-6				5.0%		8.1%	35.4%

- (1) This series is based on the data published in “Child Maltreatment”: Department of Health and Human Services. Administration in Children, Youth and Families, Child Maltreatment. Washington DC: U.S. Government Printing Office. Years 2001 to 2006. From 2007 onwards the publication ceases to publish figures by age for individual states.
- (2) This series is based on a special analysis for this study by the National Data Archive for Child Abuse and Neglect (NDACAN), Cornell University, based on the National Child Abuse and Neglect Data System (NCANDS). Data prior to 2004 were not available on NCANDS.

Notes:

- a Prinz et al (2009) don’t show the years in which baseline and post were measured. Prinz (2006) shows that Triple P was disseminated in 2004 and 2005. According to Prinz et al (2009) ‘baseline was measured ‘just prior to the initiation of the intervention’, there was a ‘2-year intervention period’, and ‘post’ was measured after the intervention. I therefore assume that the baseline year was 2003 and the post year was 2006. The assumed years are line with Prinz (2006), which suggests that the initially planned 5-year baseline period (which was, as shown above, not used for computing effect sizes) was 1999-2003. Note that overall levels reported by Prinz et al. (2009) are slightly lower than those published in either ‘Child Maltreatment’ or the special analyses of NCANDS. The reasons are not known. Prinz et al. (2009) report that they ‘unduplicated’ child maltreatment cases, which may account for the difference.
- b TG = Treatment group counties; CG = Control group counties.