

Addendum to “Population-Based Prevention of Child Maltreatment: The U.S. Triple P System Population Trial”

Ronald J. Prinz¹ · Matthew R. Sanders² · Cheri J. Shapiro³ · Daniel J. Whitaker⁴ · John R. Lutzker⁴

Published online: 16 January 2016
© Society for Prevention Research 2016

Abstract A previous article published several years ago (Prinz et al. *Prevention Science*, 10, 1–12, 2009) described the main results of a place-randomized-design study focused on the prevention of child-maltreatment-related outcomes at a population level through the implementation of a multilevel system of parenting and family support (the Triple P—Positive Parenting Program). The current report, prepared at the encouragement of the journal, provides additional details about procedures, measures, and design-related decisions, presents an additional analysis of the main outcome variables, and poses questions about the study and its implications. We also offer guidance about how the field can move forward to build on this line of research. From the outset, the three designated primary child maltreatment outcomes were county-wide rates for substantiated child maltreatment cases, out-of-home placements, and hospital-treated child maltreatment injuries, derived from independent data sources available through administrative archival records. Baseline equivalence between the two intervention conditions was reaffirmed. The additional analysis, which made use of a 5-year baseline (replacing a 1-year baseline) and ANCOVA, yielded large effect

sizes for all three outcomes that converged with those from the original analyses. Overall, the study underscored the potential for community-wide parenting and family support to produce population-level preventive impact on child maltreatment. Issues addressed included (1) the need for replication of population-oriented maltreatment prevention strategies like the one tested in this randomized experiment, (2) the need to demonstrate that a parenting-based population approach to maltreatment prevention can also impact children’s adjustment apart from child abuse, and (3) the role of implementation science for achieving greater population reach and maintenance over time.

Keywords Child maltreatment · Prevention · Archival records · Place randomization · Implementation science

Seven years ago, we published an article (Prinz et al. 2009) reporting the main results of a place-randomized-design study focused on the prevention of child-maltreatment-related outcomes at a population level. As with any trial, some questions have been raised regarding the procedures and the robustness of the findings, and as a result, we were asked by the journal to write an article addressing several questions about the trial related to design, methods, analysis, and results. The current report furnishes additional details about procedures used and design-related decisions, presents an additional analysis of the main findings, and poses questions about the study that provide clues as to how the field can move forward to build on this line of population-level research.

Place randomization studies are rare in the social and behavioral sciences, and to our knowledge, the previous article (Prinz et al. 2009) described the first place randomization outcome study in the area of child maltreatment (CM) prevention. The intervention involved the community-wide

This comment refers to the article available at: <http://dx.doi.org/10.1007/s11121-009-0123-3>.

✉ Ronald J. Prinz
prinz@sc.edu

¹ Psychology Department, University of South Carolina, Columbia, SC 29208, USA

² University of Queensland, Brisbane, Australia

³ Institute for Families in Society, University of South Carolina, Columbia, SC, USA

⁴ Georgia State University, Atlanta, GA, USA

implementation of the Triple P—Positive Parenting System (Sanders 2012). Thus, with this innovative effort, there were few precedents in the field and we recognized that much was still to be learned about this emerging approach. We identified some of these issues in the discussion section of the 2009 article, where we acknowledged that, “There is much to learn about how to conduct and interpret population trials...” and that this “trial should be viewed as the beginning of a line of population research in the prevention of CM...” (p.10). The focus of our 2009 article was on CM indicators derived from archival data systematically and routinely collected and reported to the state data repository by hospitals, the foster care system, and the child protective services system. The critical consideration for this measurement strategy was that these data were full and complete for every county, with the same reporting methods used in each of the counties in the study. With respect to the independent variable, namely, training of a wide range of practitioners from different agencies and service sectors to deliver the program, we provided specific data on assessment of reach, which was assessed by independent follow-up interviews of almost all of the practitioners who had been trained.

Measurement, Design Details, and Procedural Considerations

We have been asked to provide additional information about the measurement plan, design details, and procedures pertinent to our 2009 article. By way of background, an earlier article (Prinz and Sanders 2007) described the nature, rationale, and empirical foundation for a population-level approach to parenting and family support intervention and provided an introduction to the intervention project. However, that article was not written as a study protocol for the randomized trial, but rather as a conceptual description of the goals of the study as an illustrative example. Decisions about necessary procedural adjustments to the study at various stages were made in consultation with Centers for Disease Control and Prevention (CDC) scientists who were members of the research team, as the study was funded as a cooperative agreement.

One question to address pertains to what outcomes were targeted from the outset of the study and what adjustments if any were made to the plan during implementation. The main thrust of the study was to test the impact of the parenting-based intervention on child maltreatment prevention, as assessed by archival-administrative data records. A second goal was to examine the impact of the same intervention on general parenting and child behavior population-wide, as assessed by random household telephone surveys. This second measurement domain, which was not included in the previous article (Prinz et al. 2009), is discussed later in this paper.

The previous article (Prinz et al. 2009) focused exclusively on the archival-administrative data records and CM prevention. We were asked to clarify what CM-related outcomes were intended when the study was launched, and whether outcome variables were added or dropped. From the outset, three primary outcomes for CM were targeted, each derived from an independent source:

1. Substantiated child maltreatment cases. This indicator, which was generated by Child Protective Services (CPS) in each county and reported to a central data repository, included cases of substantiated (founded) CM for any child under age 8 years during the given year. A case could involve one or more than one type of maltreatment (e.g., physical abuse and neglect) but regardless of the number of maltreatment types/categories and the number of substantiation opportunities, no child was counted more than once in a calendar year; that is, the data were unduplicated.
2. Out-of-home placements. This indicator, which was generated by the Foster Care System in each county and reported to a central data repository, was a count of children under age 8 years placed into foster care during the given year. No child was counted more than once in a calendar year.
3. CM injuries (hospital treated). This indicator, which was generated by hospitals in each county and reported to a central data repository, included any hospital-treated injury or other medical condition that received an ICD code linked to possible CM and pertained to a child under age 8 years. Such cases came from both emergency room treatment and inpatient hospitalization. No child was counted twice in the same calendar year.

Clarification is provided here regarding another variable, namely, reports of suspected CM, and how this variable was handled within the prevention study. By definition, reports of suspected CM include both unfounded and substantiated/founded cases. Because substantiated CM cases were entirely subsumed within reports, and further because these data (i.e., reports and substantiated cases) were derived from the same single source (i.e., CPS), it was felt that only one category should be a designated or primary outcome. While the maltreatment field has generally embraced reports over substantiations, there were specific reasons why substantiated CM was selected as the targeted outcome for this study: (1) The supervisor for data management within the state department of social services, as well as programmers at the state data repository, advised that CM reports included some cases containing inaccurate maltreatment categorization, incorrect ages which would make age filtering less reliable, and key missing information that would make unduplication of records less accurate. (2) Only medium to large counties (i.e., population

greater than 50,000) were included in the design, which mitigated the concern that reports are preferred over substantiations to overcome low base rates. (3) Substantiated CM cases invoke many costs and burdens associated with casework, treatment and other services, the legal system, and foster care, whereas a subset of reports, namely, the unsubstantiated cases, do not result in these costs, which makes substantiated CM a better choice for indexing societal impact.

In addition to evaluating intervention impact on CM, the population trial sought to gauge potential impact on general parenting and child behavior problems using random household telephone survey methodology. We explain here why this data source could not be used in the outcome evaluation presented in the article of Prinz et al. (2009). Serious obstacles were encountered that invalidated the telephone survey as a means of reliably assessing county-level outcome effects, a conclusion reached by the research team in consultation with additional CDC scientists. Specifically, several issues undermined the utility of the telephone survey for precise measurement at the county level. First, the unanticipated, sudden, and steady rise in the percentages of adults and children living in households with only mobile phone service occurred at the time the study started. This phenomenon was highest among adults living in poverty or near poverty and produced substantial coverage bias (Blumberg and Luke 2007). Second, there was a concomitant increase in households with both mobile and landline phone service where the landline was virtually ignored, adding to survey non-response (e.g., average county response rate was 11.4 %). Third, efforts to adequately sample lower socioeconomic households, the very households most likely to be involved in CM incidents, as well as to yield adequate representation of African American parents, were unsuccessful in a number of the counties. All of the aforementioned factors adversely affected the telephone survey and representativeness. For example, the telephone surveys for all 18 counties substantially underrepresented lower SES families, but this was most pronounced for 9 of the counties which underrepresented by over two thirds the proportion of families with annual household incomes below \$30,000. Similarly, the telephone surveys unrepresented African American families in all 18 counties and especially in 5 counties where the surveys achieved representation of less than a third of the proportion of African Americans residing in those counties. Consequently, the random household telephone survey method was not considered valid for reliable county-level assessment of the dependent variables (i.e., outcomes) necessary to conduct outcome analysis for this domain.

However, it was possible to make use of the household telephone survey to provide relevant data about one aspect of the independent variable. That is, the telephone survey could assess whether or not the media programming facet of the independent variable, in the form of awareness of the

intervention program, had increased in the intervention counties collectively. The program awareness data were presented in Prinz et al. (2009), but the caveat not discussed was that the verification of media exposure could not be evaluated with respect to either reach into lower SES households or analysis at the individual-county level.

More information is provided here regarding the design, method of county random assignment, and procedural sequence. As described in Prinz et al. (2009), the 18 selected counties were randomly assigned to intervention or comparison conditions. Matched random assignment was planned from the beginning and implemented as planned, use of the term “block randomization” in the 2009 article notwithstanding. According to Keele, “The matched pairs design is simply a form of a block design where each block contains only two units” (Keele et al. 2008). Matched random assignment consistent with recommendations by Graham et al. (1984) was employed to reduce the likelihood of “unhappy randomization” (i.e., important chance differences at baseline). Matching took into account three variables: county population size, county poverty rate, and county CM rate (per 1000 children birth to 17 years). Nine pairs of counties were then randomized to condition. The 5-year period from 1999 to 2003 provided the backdrop for gauging subsequent hypothesized effects on the three archival outcome indicators. Triple P training of service providers in the nine intervention counties began midway through 2003. The whole 2005 calendar year was designated as the period of outcome to gauge the impact of population exposure from 2 years of program implementation. In 2006, Triple P practitioner training was initiated in the comparison counties consistent with the original grant plan, marking the end of the randomized study.

Additional Analysis

The outcome analyses in our 2009 article made use of the 1-year baseline period. We were asked to consider whether using the 5-year baseline period (i.e., an average of the five baseline years) would have been preferable. An argument can be made that the 5-year period provides a more stable estimate of baseline. Consequently, we undertook this additional analysis and have reported the results here.

A preliminary issue was baseline equivalence. Using the 5-year baseline, the two sets of counties were compared using *t* tests, which yielded no significant baseline differences for the 5-year averages: substantiated CM cases, $t(16)=0.22, p=.83$; out-of-home placements, $t(16)=0.30, p=.77$; and CM injuries, $t(16)=0.99, p=.34$. The means and standard deviations for the baseline 5-year averages are found in Table 1 in the original Prinz et al. (2009) article. Baseline equivalence, then, was clearly established for all three outcome variables. Additionally, it was already reported that the two sets of

Table 1 Pre-post means and standard deviations on the primary outcomes for intervention and control counties

Outcome variables	Triple P System counties ($n=9$)		Control counties ($n=9$)	
	Pre-intervention(5-year avg)	Post-intervention	Pre-intervention (5-year avg)	Post-intervention
Substantiated CM cases	10.82 (4.36)	10.86 (3.79)	11.40 (6.75)	16.30 (9.13)
Out-of-home placements	4.02 (1.59)	3.90 (2.11)	3.76 (1.91)	4.64 (2.02)
CM injuries (hosp & ER)	1.72 (0.66)	1.54 (0.84)	1.44 (0.53)	1.78 (0.99)

Each mean is a county rate of the respective occurrence per 1000 children. The source of the data was the South Carolina central data repository for the Child Protective Services system, the Foster Care System, and hospitals, respectively

counties were comparable on the matching variables with t test p values ranging from .73 to .87 (see Table 1 in the original 2009 article). The randomization procedure avoided unhappy randomization but more importantly yielded two sets of counties that were comparable on both the matching variables and the baselines for all three outcome variables.

In undertaking the additional analysis using the 5-year baseline, we also faced the issue of whether or not to retain the pre-post difference score approach used in Prinz et al. (2009). Although legitimate, the use of pre-post difference scores in randomized designs is somewhat controversial and not without its disadvantages (Rogosa 1988; Senn 2006). There is also asymmetry in creating a difference score where the pre-score is a 5-year average and the post-score is a single year. Consequently, for the present analysis, the difference score approach was replaced by analysis of covariance (ANCOVA) for each of the three outcome variables, controlling for the corresponding (5-year) baseline for each variable. ANCOVA is an appropriate analytic approach when randomization has taken place and baseline equivalence has been established (Cohen et al. 2002). Furthermore, ANCOVA can be used either as an alternative or a complement to matched random assignment (Shadish et al. 2002).

We were asked to clarify what significance level and direction were hypothesized, and why, with respect to the outcome analyses in Prinz et al. (2009). Regarding the analyses undertaken for Prinz et al. (2009) and for this paper, of greatest importance was the magnitude or strength of effect (i.e., effect size), in the predicted direction of course. This is consistent with what many researchers/methodologists (e.g., Cook and Campbell 1979; Cohen et al. 2002; Kazdin 2003; McCartney and Rosenthal 2000) have advocated, namely, an emphasis on effect size over p values. We felt going into this population-focused study that practical significance, which could only be confirmed by large effects, was of utmost importance. Effect size considerations notwithstanding, a one-tailed significance level (with an alpha of .05) was chosen a priori for testing of the three outcomes in the 2009 article and this paper for a number of reasons. We were conducting what was essentially the first place randomization study on prevention of CM. We clearly had directional hypotheses for all three outcomes and did not want to miss any potential effects in this crucial area.

For feasibility and resource reasons, there could only be 18 units (counties) for the design, which heightened risk for a type II error. It is acknowledged that a one-tailed approach did not take into account the possibility of an iatrogenic effect, but there was no basis on which to expect such effects in this study because evidence-based parent/family interventions such as the Family Check-Up, Parent-child Interaction Therapy, SafeCare, The Incredible Years, and Triple P across many outcome studies have *not* shown iatrogenic effects, so there was no basis on which to expect such effects in this study. Finally, there is precedence in other community-level trials for the use of one-tailed tests (e.g., Spoth et al. 2011). However, p values for both one-tailed and two-tailed significance are reported in the analyses presented here.

ANCOVA was conducted for each of the three outcome variables, controlling for baseline on the respective outcome variable (5-year average for baseline period). Means and standard deviations as a function of intervention condition and pre-post measurement time are found in Table 1 in the current article for the three outcome variables. For substantiated CM cases, the ANCOVA result was as follows: overall model $F(2, 15)=21.77, p<.001$, coefficient $B_{\text{cond}}=-4.836$ ($S.E. 1.866$), $t(16)=-2.592$, one-tailed $p=.02$ (two-tailed $p=.04$), effect size=1.30 (Cohen's d). For out-of-home placements, the ANCOVA result was as follows: overall model $F(2, 15)=16.95, p<.001$, coefficient $B_{\text{cond}}=-0.990$ ($S.E. 0.569$), $t(16)=-1.741$, one-tailed $p=.05$ (two-tailed $p=.10$), effect size=.87 (Cohen's d). For CM injuries (hosp & ER), the ANCOVA result was as follows: overall model $F(2, 15)=11.96, p<.002$, coefficient $B_{\text{cond}}=-0.581$ ($S.E. 0.289$), $t(16)=-2.014$, one-tailed $p=.03$ (two-tailed $p=.06$), effect size=1.01 (Cohen's d).

The effect sizes found in the additional analyses are large for all three outcome variables. The magnitude of these effects converges with what was reported in the previous article (Prinz et al. 2009) on all three outcomes. The observed p values reflect the relatively small number of units (i.e., 18 counties) rather than small effects. Even though the three outcomes were all significant using one-tailed tests, there is still the possibility that the observed large effects happened by chance. For this and other reasons, replication of the findings

and associated prevention strategy is critically important to increase confidence in the conclusion.

When a study employs a large number of statistical tests and perhaps also includes post hoc comparisons, statistical methods are used to control the potential for chance findings. With respect to the results reported here and in Prinz et al. (2009), for which no statistical adjustments were made, it should be kept in mind that (a) only three outcome tests were conducted, with all three outcome variables derived from independent data sources (i.e., CPS, foster care system, and hospitals), and (b) results for all three outcomes reflect quite large effect sizes.

The pattern of intervention effects for both out-of-home placements and hospital-treated CM injuries was reflected in lower rates in the intervention counties relative to control counties as well as baseline. For substantiated CM cases, the mean rate for the intervention counties was held constant relative to baseline, while the control counties showed a rate increase during the same period. Randomization (including equivalence at baseline) supported a conclusion of preventive impact on substantiated CM cases. Additionally, however, it was noted in Prinz et al.'s (2009) discussion section that the increase "in the control counties mirrored similar increases across the other 28 counties." Data for the non-study counties came from the same central repository used for the study counties. More precisely, the 28 non-study counties in the state showed an increase of 37 % in mean rate of substantiated CM cases (from 9.79 to 13.40) during the same period (baseline 5-year average to post-intervention), compared to that found in Table 1 (in this current paper) with an increase of 43 % for the control counties and only 0.4 % for the intervention counties. It was the case that substantiated CM rates for the intervention counties did not decrease from baseline. However, given randomization, equivalence at baseline, and a general pattern of increase for the control and non-study counties during the intervention time period, one would expect a similar increase in the intervention counties had the intervention not taken place, which all taken together supports the conclusion of a true preventive effect on substantiated CM. Furthermore, the intervention counties showed a decrease in variance for substantiated CM rates from pre to post, compared with a variance increase for the control counties (see SDs in Table 1 in the current paper).

The robustness of scientific findings is ultimately determined by replication (Valentine et al. 2011). Since the publication of Prinz et al. (2009), independent studies have emerged further supporting the promise and viability of a population approach to CM prevention. For example, a four- to seven-session universal postnatal nurse home visiting program in Durham County (NC) has shown promise for impact on CM cases and emergency room treatment (Dodge et al. 2014). In another study, with 15-year follow-up using case-linked administrative data in a quasi-experimental design,

Smith (2015) found that implementation of Level 4 Group Triple P delivered when children were pre-schoolers significantly reduced the rate of hospital emergency department visits over childhood and adolescence.

The previous study (Prinz et al. 2009) was not able to determine the impact of the Triple P system on general parenting practices and child behavioral problems. Recently, however, a number of investigators have been examining this issue (Fives et al. 2014; Frantz et al. 2015). For example, an evaluation of the Triple P system in Ireland, reported by Fives et al. (2014), found population-level effects on children's emotional and behavioral problems and on a range of parenting variables. This study overcame the aforementioned challenges of using a random-dial telephone survey—by employing face-to-face interviews with parents in randomly selected households. Newer studies such as Fives et al. (2014) provide clues as to potential mediators for population impact beyond what Prinz et al. (2009) was able to address. Further to the matter of replication, the core foundational elements of the intervention system used in Prinz et al. (2009) have undergone extensive replication in numerous studies across applications, settings, and investigators (with and without developer involvement), and have demonstrated robust effects (see Sanders et al. 2014).

The Prinz et al.'s (2009) place-randomization study provided evidence that community-wide implementation of parenting and family support can positively impact child-maltreatment-related indicators in a preventive manner. These outcomes might have resulted due to (a) effects on coercive parenting in families accessing the intervention; (b) effects resulting from mobilization of a social contagion via media as well as conversations among parents and practitioners, such as described in Fives et al. (2014); (c) broader impact from training many practitioners serving high-risk segments of the population; or (d) some combination of all of these factors. These putative mediators provide cogent hypotheses for much needed future studies.

Although Prinz et al. (2009) provided initial evidence for proof of concept relative to prevention of CM, this early investigation did not include or address strategies to optimize penetration and impact, nor cogent procedures for sustaining program utilization. Since the undertaking of this study, the field has seen the rapid development of implementation science, which now offers indispensable guidance (Fixsen et al. 2013). In retrospect, some of the recommended practices that the Prinz et al. (2009) study might have instituted include the following: (1) extended preparatory planning with supervisors and managers prior to training, (2) building of supportive organizational climate and structural environment, (3) institutionalizing quality assurance processes (e.g., fidelity assessment and promotion, peer support networks, and active evaluation process with feedback loops), and (4) initiating multiple action steps to achieve sustainability. However, more

recent large-scale deployments of the Triple P system have paid close attention to optimizing implementation processes through the use of a structured implementation framework in working with partner organizations (e.g., Fives et al. 2014).

Since the publication of Prinz et al. (2009) which demonstrated viability and efficacy, work has emerged that underscores both the promise of a population-level approach to family-based prevention and the need for future research in several areas articulated in this report. Refinement and expansion of the population, public health strategies associated with this line of work are critical to realizing society's aspirations for child well-being (Biglan 2015; National Research Council and Institute of Medicine 2009).

Compliance with Ethical Standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The Institutional Review Board at the University of South Carolina granted an exemption for the research protocol.

Conflict of Interest The Triple P—Positive Parenting Program is owned by The University of Queensland. The University, through its technology transfer company Uniquist Pty Ltd., has licensed Triple P International Pty Ltd. to disseminate the program worldwide. Royalties stemming from this dissemination work are paid to UniQuest, which distributes payments to the University of Queensland Faculty of Health and Behavioural Sciences, School of Psychology, Parenting and Family Support Centre, and contributory authors in accordance with the University's intellectual property policy. No author has any share or ownership in Triple P International. Matthew Sanders is the founder and lead author of the Triple P—Positive Parenting Program and is a consultant to Triple P International. Ronald Prinz is a consultant to Triple P International. Cheri Shapiro is a consultant to Triple P America. Daniel Whitaker and John Lutzker have no competing interests. The University of South Carolina's Office of Research Compliance in conjunction with the CDC assisted the investigators with protection of the integrity of the research on which the original article was based.

Funding The research on which this article was based was support by Grant #U17/CCU422317 from the National Center for Injury Prevention and Control, Centers for the Disease Control and Prevention. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References

- Biglan, A. (2015). *The nurture effect: How the science of human behavior can improve our lives and our world*. Oakland: New Harbinger.
- Blumberg, S. J., & Luke, J. V. (2007). Coverage bias in traditional telephone surveys of low-income and young adults. *Public Opinion Quarterly*, *71*, 734–749. doi:10.1093/poq/nfm047.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: Erlbaum.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.
- Council, N. R., & Institute of Medicine. (2009). Preventing mental, emotional, and behavioral disorders among young people: progress and possibilities. In M. E. O'Connell, T. Boat, & K. E. Warner (Eds.), *Board on children, youth, and families, division of behavioral and social sciences and education*. Washington, DC: National Academies Press.
- Dodge, K. A., Goodman, W. B., Murphy, R. A., O'Donnell, K., Sato, J., & Guptill, S. (2014). Implementation and randomized controlled trial evaluation of universal postnatal nurse home visiting. *American Journal of Public Health*, *104*, s136–s143. doi:10.2105/ajph.2013.301361.
- Fives, A., Purcell, L., Heary, C., Nic Gabhainn, S., & Canavan, J. (2014). *Parenting support for every parent: a population-level evaluation of Triple P in Longford Westmeath. Final Report*. Athlone: Longford Westmeath Parenting Partnership.
- Fixsen, D., Blasé, K., Metz, A., & VanDyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, *79*, 213–230. doi:10.1177/001440291307900206.
- Frantz, I., Stemmler, M., Hahlweg, K., Pluck, J., & Heinrichs, N. (2015). Experiences in disseminating evidence-based prevention programs in a real-world setting. *Prevention Science*, *16*, 789–800. doi: 10.1007/s11121-015-0554-y.
- Graham, J. W., Flay, B. R., Johnson, C. A., Hansen, W. B., & Collins, L. M. (1984). Group comparability: a multiattribute utility measurement approach to the use of random assignment with small numbers of aggregated units. *Evaluation Review*, *8*, 247–260. doi:10.1177/0193841X8400800206.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.
- Keele, L., McConaughy, C., & White, I. (2008). Matching in randomized experiments. *Society for Political Methodology Conference*, University of Michigan.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*, 173–180. doi: 10.1111/1467-8624.00131.
- Prinz, R. J., & Sanders, M. R. (2007). Adopting a population-level approach to parenting and family support interventions. *Clinical Psychology Review*, *27*, 739–749. doi:10.1016/j.cpr.2007.01.005.
- Prinz, R. J., Sanders, M. R., Shapiro, C. J., Whitaker, D. J., & Lutzker, J. R. (2009). Population-based prevention of child maltreatment: the U.S. Triple P System Population Trial. *Prevention Science*, *10*, 1–12. doi:10.1007/s11121-009-0123-3.
- Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. M. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–209). New York: Springer.
- Sanders, M. R. (2012). Development, evaluation, and multinational dissemination of the Triple P—Positive Parenting Program. *Annual Review of Clinical Psychology*, *8*, 1–35. doi:10.1146/annurev-clinpsy-032511-143104.
- Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The triple P—positive parenting program: a systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review*, *34*, 337–357. doi:10.1016/j.cpr.2014.04003.
- Senn, S. (2006). Change from baseline and analysis of covariance revisited. *Statistics in Medicine*, *25*, 4334–4344. doi:10.1002/sim.2682.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont: Wadsworth, Cengage Learning.
- Smith, G. (2015). *15-year follow-up of WA Triple P trial, collaboration for applied research and evaluation*. Western Australia: Telethon

- Institute for Child Health Research under contract with the Department of Health.
- Spoth, R., Redmond, C., Clair, S., Shin, C., Greenberg, M., & Feinberg, M. (2011). Preventing substance misuse through community-university partnerships: randomized controlled trial outcomes 4 ½ years past baseline. *American Journal of Preventive Medicine*, *40*, 440–447. doi:[10.1016/j.amepre.2010.12.012](https://doi.org/10.1016/j.amepre.2010.12.012).
- Valentine, J. C., Biglan, A., Boruch, R. F., Castro, F. G., Collins, L. M., Flay, B. R., et al. (2011). Replication in prevention science. *Prevention Science*, *12*, 103–117. doi:[10.1007/s11121-011-0217-6](https://doi.org/10.1007/s11121-011-0217-6).