Child Maltreatment
http://cmx.sagepub.com/

Is a Structured, Manualized, Evidence-Based Treatment Protocol Culturally Competent and Equivalently Effective Among American Indian Parents in Child Welfare?

Mark Chaffin, David Bard, Dolores Subia Bigfoot and Erin J. Maher

Child Maltreat 2012 17: 242 originally published online 27 August 2012
DOI: 10.1177/1077559512457239

The online version of this article can be found at:
http://cmx.sagepub.com/content/17/3/242

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
APSAC
American Professional Society on the Abuse of Children

Additional services and information for *Child Maltreatment* can be found at:

Email Alerts: http://cmx.sagepub.com/cgi/alerts

Subscriptions: http://cmx.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://cmx.sagepub.com/content/17/3/242.refs.html

>> Version of Record - Sep 14, 2012

OnlineFirst Version of Record - Aug 27, 2012

What is This?

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

Child Maltreatment
17(3) 242-252
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1077559512457239
http://cmx.sagepub.com
SAGE

Is a Structured, Manualized, Evidence-Based Treatment Protocol Culturally Competent and Equivalently Effective Among American Indian Parents in Child Welfare?

Mark Chaffin[1], David Bard[1], Dolores Subia Bigfoot[1], and Erin J. Maher[2]

Abstract

In a statewide implementation, the manualized SafeCare home–based model was effective in reducing child welfare recidivism and producing high client satisfaction. Concerns about the effectiveness and acceptability of structured, manualized models with American Indians have been raised in the literature, but have rarely been directly tested. This study tests recidivism reduction equivalency and acceptability among American Indian parents. A subpopulation of 354 American Indian parents was drawn from a larger trial that compared services with versus without modules of the SafeCare model. Outcomes were 6-year recidivism, pre/post/follow-up measures of depression and child abuse potential, and posttreatment consumer ratings of working alliance, service satisfaction, and cultural competency. Recidivism reduction among American Indian parents was found to be equivalent for cases falling within customary SafeCare inclusion criteria. When extended to cases outside customary inclusion boundaries, there was no apparent recidivism advantage or disadvantage. Contrary to concerns, SafeCare had higher consumer ratings of cultural competency, working alliance, service quality, and service benefit. Findings support using SafeCare with American Indians parents who meet customary SafeCare inclusion criteria. Findings do not support concerns in the literature that a manualized, structured, evidence-based model might be less effective or culturally unacceptable for American Indians.

Keywords

SafeCare, home visitation, American Indians, child welfare, cultural competency

American Indian[1] parents in tribal, state, or county child welfare systems may receive home-based services as part of a child welfare service plan. A fully scaled up statewide controlled trial of the SafeCare home–based intervention model was recently completed in Oklahoma (Chaffin, Hecht, Bard, Silovsky, & Beasley, 2012). In the full study, SafeCare was found to reduce child welfare recidivism, yielding hazard ratios between 0.74 and 0.83, with the larger effect found among those participants meeting customary SafeCare inclusion criteria (i.e., at least one preschool age child and no current untreated substance use disorder). Of the 2,175 families enrolled in the Oklahoma trial, two thirds were Caucasian. There were 354 analyzable American Indian participants. This American Indian subpopulation represents perhaps the largest group of American Indian parents ever to participate in a child welfare comparative outcome study. Oklahoma is home to a number of indigenous and relocated American Indian tribes and nations that includes, but is not limited to, the Arapaho, Caddo, Cherokee, Cheyenne, Chickasaw, Choctaw, Creek, Delaware, Kaw, Kickapoo, Otoe-Missouria, Modoc, Ottawa, Pawnee, Ponca, Potawatomi, Quapaw, Sac and Fox, Seneca, Shawnee, Wyandotte, and Tonkawa. Whenever there is a sufficient ethnic or cultural minority subsample available within a clinical trial, it is important to examine whether subpopulation findings (i.e., effect sizes) are equivalent to those found in the overall study. It is always possible that trial findings may be driven by effects solely among a preponderance of majority culture participants and may not extend across diverse cultures. The main aim of this study was to examine whether recidivism reduction effects among American Indians were equivalent to those observed in the full study.

There has been very little home visiting outcome research of any kind among American Indian populations, and virtually all of what does exist comes from the primary prevention and

[1] University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
[2] Casey Family Programs, Seattle, WA, USA

Corresponding Author:
Mark Chaffin, University of Oklahoma Health Sciences Center, P.O. Box 26901, Oklahoma City, OK 73190, USA
Email: mark-chaffin@ouhsc.edu

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

health promotion literature, rather than from the child welfare services literature. Knowledge disparities may be related to the very small numbers of American Indian participants in most trials, a shortage of home-based services in Indian Country, or a lingering history of distrust about research participation (Barlow et al., 2006; Walkup, Barlow, Mullany, Pan, Goklish, Hasting et. al., 2009). In the primary prevention home visiting literature, studies have examined self-report outcomes among American Indian families including maternal stress and depression (Barlow et al., 2006; Walkup, Barlow, Mullany, Pan, Goklish, Hasting et. al., 2009). In the health prevention literature, one home visiting study has focused on obesity prevention (Harvey-Berino & Rourke 2003). The Healthy Start and Healthy Families home-visiting primary prevention evaluations in Alaska and Hawaii tested child maltreatment outcomes using randomized designs and included a significant proportion of Native people, but most were Native Hawaiian or Alaska Native people, not American Indians. Neither evaluation found reduced child welfare report outcomes (Duggan et al., 2007; Duggan et al., 2004). We were unable to locate any published trials, either specific trials or subpopulation equivalency studies, testing home visiting outcomes among American Indian parents in child welfare.

Based on our overall findings from the full study, we predicted a comparable SafeCare effect on downstream child welfare recidivism of about Hazard Ratio (HR) = .74 relative to similar home-based services without the structured SafeCare modules, and within customary model inclusion criteria. We also examined outcomes for two known malleable risk factors for child maltreatment—parental depression (Chaffin, Kelleher, & Hollenberg, 1996) and scores on the Child Abuse Potential Inventory (Milner, 1986), given that both of these measures tend to improve over the course of home-based services and are markers of general well-being and parenting distress (Chaffin & Bard, 2011). These additional outcomes were examined as supportive aims for the main recidivism equivalency question.

Equivalent outcomes are not certainty. The potential benefit of dominant culture service models for American Indian populations has been controversial, particularly when it comes to the highly manualized and structured evidence-based protocols being promoted for child welfare service systems (Barth et al., 2005; Chaffin & Friedrich, 2004). None of these evidence-based models were developed with American Indian culture in mind (Red Horse, Martinez, & Day, 2001; Red Horse et al., 2000). How well these models engage and are received by American Indian parents is virtually unstudied but will be important to learn before implementing evidence-based models in Native communities. In juvenile justice or mental health contexts, evidence-based models tend to deliver roughly equivalent benefits across majority and minority ethnic groups (Huey & Polo, 2008; Miranda, Schoenbaum, Sherbourne, Duan, & Wells, 2004), but we do not know whether this finding extends to American Indians or to parents in child welfare. Some observers cite the unique world views of American Indian people and argue that structured, manualized, behavioral, evidence-based models may be a poor fit or will prove unacceptable (Cruz & Spence, 2005). We would argue that the cultural fit of a model is best judged by asking actual consumers. SafeCare consumers have reported relatively better working alliance, cultural competency, and satisfaction (Damashek, Bard, & Hecht, 2012), but this has not been examined separately among American Indian consumers, and doing so is a second aim in the present study. Based on prior findings, we hypothesized that SafeCare would be perceived by American Indian consumers as having equivalent or better working alliance, quality, benefit, and cultural competency compared to services without SafeCare.

Method

Participants

Participants in this study were a subpopulation of 354 parents or caregivers who self-reported American Indian ethnicity drawn from the full Oklahoma SafeCare trial. The home-based services were operated by community-based agencies under contracts with child welfare which formed a managed home-based child welfare service network. Six administrative regions of the state, two urban and four rural, were served by a lead agency within each region, and all participated in the trial. Cases were enrolled in the study between September 9, 2003, and October 1, 2006. Eligible enrollees were maltreating caregivers (but not sexual abusers) referred by child welfare. One maltreating parent per household was enrolled, prioritizing the primary caregiver. Parents were recruited in their homes by an independent research assistant shortly after service enrollment and study enrollment was voluntary. Study participants and nonparticipants accessed the same services, and home-visitors were not told about their clients' study participation status. Separate enrollment flow for American Indian parents is unavailable, but overall 3,116 prospective participants were approached, 18 did not complete the recruitment process, 23 were determined to be ineligible, and 816 declined to enroll or did not complete baseline data collection, yielding an overall enrollment of 2,259 (72% of all individuals approached). Eighty-four participants were withdrawn after enrollment, leaving an analysis sample of 2,175 of which 354 reported American Indian ethnicity. The project was overseen by the University of Oklahoma Health Sciences Center institutional review board (IRB) and a study-specific Data and Safety Monitoring Board that included experts on research in tribal contexts. There were no study related adverse events.

Ninety-four percent of the 354 American Indian participants were female, with a mean age of 29 years ($SD = 8$; range = 18–71). Participants had a median of three children in their household, 78% of households had at least one preschool age child, and 7% of women reported being pregnant at baseline. Eighteen percent (18%) were urban, 68% lived in small communities, and 14% were rural. Eighteen percent (18%) had no telephone, and 31% had no access to a car. Residential instability was common, with 49% having lived in their current community less than 3 years, and 53% having moved more

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

than twice in the last 5 years. Twenty-six percent (26%) were married, 17% were cohabitating, 16% were separated, 15% were divorced, 2% were widowed, and 24% were never married. Thirty-eight percent (38%) had less than a high school education, 43% had a high school diploma or equivalent, 15% completed some college, and 5% had completed college. Median household income was $900 per month. Applying 2009 U.S. federal poverty line criteria for income and family size, 80% of households fell below the federal poverty line. Twenty-six percent (26%) indicated that they were currently unemployed, 26% were homemakers, 30% had a full-time job, 6% were students, 10% were working part-time, and 2% were self-employed. On items from the Family Resource scale, 15% indicated not having enough food for two consistent meals daily, 17% did not consistently have heat in their homes, 8% did not have indoor plumbing, 37% did not have consistent access to adult health care, and 48% did not consistently have money for necessities or bills. Forty-three percent (43%) self-reported being physically abused as a child and 41% self-reported being sexually abused as a child. Twenty-three percent (23%) reported being removed from their own parents at some point during their childhood. Forty-five percent (45%) reported some history of domestic violence in their current household. Participants had a mean of three and a median of two unduplicated prior household referrals to child welfare (range of 0–14; $SD = 2.7$). Ninety-three percent (93%) of all prior referrals involved child neglect, 19% involved physical abuse, and 5% involved sexual abuse (total is greater than 100% due to multiple allegations within a given referral).

Design and Procedure

A 2 × 2 cluster design was used. SafeCare versus Services as Usual was assigned at the agency/region level ($n = 6$ across the state). Within agencies/regions, small home visitor teams were randomized to coaching conditions. Thus, each home visitor in the state ($n = 219$) was assigned to deliver only one of the four design cells. Region/agency assignment to treatment condition began by randomizing the two urban regions, then all possible assignment permutations were evaluated and the solution with the best balance of pre-study case demographics was accepted. Once assigned, treatment conditions were formalized in agency funding contracts. For home visitors, 21% were assigned to Services As Usual/Uncoached, 22% to Services As Usual/Coached, 28% to SafeCare/Uncoached, and 28% to SafeCare/Coached. For the American Indian subpopulation, 18% were assigned to Services As Usual/Uncoached, 21.5% to Services As Usual/Coached, 30.5% to SafeCare/Uncoached, and 30% to SafeCare/Coached.

Treatments

Common elements. It is important to note that SafeCare and Services As Usual in this trial were identical in most respects with the exception of the SafeCare modules themselves. The SafeCare trial began with a well-established statewide network of homogeneous home-based service programs which study investigators had been evaluating and directly observing for over 5 years prior to the start of the experimental trial, including investigators making multiple site visits and directly observing in-home services on multiple occasions across all participating agencies. For those home visitors assigned to SafeCare, implementation involved migrating to SafeCare modules and materials for part of the overall in-home service (e.g., migrating to using SafeCare parenting materials rather than existing parenting materials). Both SafeCare and Services As Usual conditions offered case management and additional collateral services in common. Commonalities between conditions included the home-based format, caseloads, service duration (6 months), visit frequency (at least weekly), service goals, minimum workforce qualifications, case management practices, reporting requirements, administrative definitions, supervisory qualifications and frequency, assessment tools, and funding. All home visitors were trained in basic motivational interviewing and domestic violence safety planning skills and had access to emergency funds to help families meet basic concrete needs. All service providers (both SafeCare and Services As Usual) received brief classroom training and information about American Indian culture and cultural competency, but there were no American Indian cultural adaptations made to the SafeCare model itself or its curriculum materials.

SafeCare. SafeCare is a manualized, highly structured behavioral skill training model. SafeCare modules address (a) parent/child or parent/infant interaction, basic caregiving structure and parenting routines; (b) home safety; and (c) child health. SafeCare can be delivered as a free-standing program or as one component of a broader home-visiting service, and the latter was the case here. SafeCare model details can be found in the SafeCare manual (Lutzker & Bigelow, 2002) or through the SafeCare training institute (http://chhs.gsu.edu/safecare/). Initial SafeCare training was delivered by study investigators with participation and oversight from model developers. Training occurred in small 1-week workshops using live skill demonstration and role play to criterion, followed by one directly observed client session in the field, then regular clinical supervision. Home visitors assigned to the in vivo coaching condition received additional quality control and fidelity coaching beyond this. Customary SafeCare inclusion criteria are a preschool age child in the home and absence of an untreated substance use disorder. Because this study was conducted within a more inclusive service system, families with children up to age 12 were served irrespective of untreated substance abuse. A priori hypotheses were centered on participants falling within customary SafeCare inclusion criteria (which is the normal effectiveness test for any model), but we were also interested in whether effects might extend beyond this. Fifty-four percent of American Indian parents met customary SafeCare inclusion criteria.

Services As Usual. Services As Usual were periodically observed in the field by study investigators for descriptive

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

purposes. The term Services As Usual in behavioral studies often connotes diverse, ad hoc, or minimal services, none of which was case here. Services As Usual addressed comparable goals and issues as SafeCare, but in a less structured, and less protocol driven manner. Services As Usual providers tended to follow the lead of their clients in determining visit content.

Coaching. Half of all home visitors were randomized to receive in vivo coaching as a quality control strategy in the overall 2×2 experimental design. Coaches were selected from agency staff, favoring providers viewed as credible and influential by other home visitors. Coaching was structured to be distinct from regular supervision, being advisory rather than supervisory. All coaches were trained using Stoltenberg's developmental consultation model (Stoltenberg & McNeill, 2010). Coaches traveled with home visitors to the home at least monthly. SafeCare coaches used fidelity checklists, received additional advanced SafeCare model training, and met regularly with investigators and biannually with SafeCare developers. Services As Usual coaches focused on general service issues and problem solving rather than fidelity.

Data collection procedures. Client report data for the study were collected in the home by independent research assistants using Audio Computer Assisted Self-Interview (ACASI). Parents gave responses by touch screen. Computer interviews were conducted while the research assistant waited or supervised the children in order to provide the parent with uninterrupted private time to respond to items. Home visitors were not present during data collection. Data collectors normally did not view parents' responses unless the parent requested assistance. A federal Certificate of Confidentiality was obtained, and no individual research data were shared with child welfare authorities or service providers. Measures were collected at baseline (n = 354 or 100%), around the end of the services (median time = 202 days from baseline; SD = 111; n = 202 or 57%), and again at around 6 months after service exit for post-program follow-up (median = 393 days from baseline; SD = 75; n = 134 or 38%). The dominant reason for interwave attrition was participants who had disappeared and could no longer be located despite multiple attempts to follow-up using both official and unofficial contact sources (e.g., participants who moved and left no forwarding address, who had become incarcerated, or for whom no current location could be obtained from either their identified contact persons, the home visiting service agency, neighbors, or child welfare). Data collection efforts continued irrespective of service withdrawal (13% of cases), and the missing data pattern was unrelated to service completion (chi-square = 1.71, df = 2, p = .42).

Measures

Measurement evaluation strategy. Because none of the study measures was specifically developed for American Indian populations, extra steps were taken to check their measurement properties. Internal consistency was calculated and compared to published values. Construct validity was examined by testing expected correlations with related constructs (e.g., separate measures of distress and depression). Temporal stability expectations vary for many of these constructs, but positive correlations would be expected across time and simple pairwise correlations were performed over the three measurement points for repeated measures. Exploratory and confirmatory factor analyses were used to examine subscale structure in some instances described below.

Beck Depression Inventory-2. The Beck Depression Inventory (Beck, Steer, & Brown, 1996) is a 21-item multiple-choice self-report questionnaire designed to measure symptoms of depression. Published internal consistency of the scale is .93, and test–retest stability is .93 (Beck, Steer, & Brown, 1996). Observed alpha for the scale in the American Indian sample was .91. As expected, baseline Beck Depression Inventory scores were positively correlated with the Distress subscale of the Child Abuse Potential Inventory (r = .79, p < .001), and had substantial positive pairwise correlations across the three time points (r = .58, .59, .62; all p < .001).

Child Abuse Potential Inventory. The Child Abuse Potential Inventory (Milner, 1986) is a 160-item agree/disagree parent self-report questionnaire originally designed to predict child physical abuse risk. Item content includes parenting stress, attitudes, and family conflict. The measure has high internal consistency (KR-20 = .92 to .95), a 1-month test–rest stability of 0.83 (Milner, 1986), and actuarial predictive validity for discriminating maltreating from nonmaltreating parents and predicting future physical abuse as well as future neglect reports, although changes on the scale do not necessarily correspond to changes in actual recidivism risk (Chaffin & Valle, 2000; Milner 1986). Observed alpha in the American Indian subpopulation was .95. Scores were positively correlated over time (r = .67, .66, .76; all p < .001).

The Child Abuse Potential Inventory also includes an 18-item Lie Scale measuring social desirability response bias. The scale has been found to correlate significantly with other general social desirability measures and to discriminate between parents instructed to answer honestly versus to answer in a socially desirable manner (Milner, 1986). As expected, the Lie Scale was negatively correlated with self-report measures of problems or symptoms (Beck Depression Inventory and Child Abuse Potential Inventory; r = −.38 and r = −.44; both p < .001) and Lie Scale scores were positively correlated over time (r = .75, .69, .72; all p < .001).

Working Alliance Inventory. The Working Alliance Inventory (Tracey & Kokotovic, 1989) measures client–provider agreement on intervention goals and steps, feelings of mutual liking, collaboration, affiliation, and trust. The measure was administered at posttreatment only (n = 202). Alphas reported in the literature range from .68 to .87. Observed scale alpha in the American Indian study sample was .91. As expected, working alliance was positively correlated with the Client Cultural

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

Competency Inventory ($r = .80$, $p < .001$) and the Client Satisfaction Survey (CSS; $r = .79$, $p < .001$).

Client Cultural Competency Inventory. The Client Cultural Competency Inventory (Switzer, Scholle, Johnson, & Kelleher, 1998) is a client report measure of perceptions about service cultural competency. This measure was selected because American Indian people are culturally diverse and the Client Cultural Competency Inventory items are designed to measure cultural competency across diverse cultures rather than using items with culturally specific content. The measure asks about respectfulness, appreciation for cultural differences, and congruence with whatever cultural traditions and beliefs are held by the individual. The measure was administered at posttreatment only ($n = 202$). Initial examination of item properties did not support a single internally consistent factor. Some items reflected client perception of provider respect and sensitivity, while others reflected opinions about referrals to outside services and service scheduling. An exploratory factor analysis was performed for 1, 2, and 3 factor solutions and favored the 2-factor solution (Bayesian Information Criterion = 3,465; 3,411; and 3,414 for the 1-, 2-, and 3-factor solutions, respectively). Examining items with loadings greater than .60, the first factor was comprised of 4 items describing respect for the consumer's culture and customs. Item content included (a) respect for family beliefs and customs (loading = 0.84); (b) use of understandable language (loading = 0.66); (c) absence of negative judgments because of cultural difference (loading = 0.79); and (d) accepting and respectful behavior (loading = 0.91). Provider/client American Indian ethnic match did not load strongly on the factor (loading = 0.24), and so was dropped. A confirmatory factor analysis fit the data well (overall model fit chi square = 406, $df = 603$, $p = 1.0$). The factor score was saved for subsequent analyses. The expected correlation with the Working Alliance Inventory was reported above, and the Client Cultural Competency Inventory was also positively correlated with the CSS ($r = .73$, $p < .001$).

Client Satisfaction Survey. The CSS was developed for the study to measure parents' perceptions of how much home-based services have helped their family. The questionnaire is comprised of items reflecting the process and outcome goals of this home visiting service system and was administered at the end of services. Item content focused on three general dimensions. The first was quality of the program (e.g., "How knowledgeable did you find the home visitor who worked with your family?" "How clearly were the goals of the program explained to you?" "How would you rate the quality of services received?"). The second was satisfaction with services (e.g., "If you were to seek help again, would you come back to this program?" "How satisfied are you with the amount of help you received?" "Did you get the kind of services you wanted?"). The third was realized benefits from the services (e.g., "My relationship with my child has improved;" "I am better able to care for my child when he or she is sick;" I am better able to prevent behavior problems in my children"). The overall α for the items in the American Indian study population was .94. Although this α value would support using the scale as single measure, content differences among the items led us explore multifactor solutions. An exploratory factor analysis for ordinal data was conducted, examining one- to four-factor solutions. Bayesian Information Criterion values for the one- to four-factor solutions respectively were 5,482, 5,242, 5,232, and 5,302, respectively. The third factor in the three-factor solution had no items loading greater than .40 and was conceptually muddled, so we opted to accept the two-factor solution. The first factor (which we termed Client Satisfaction Survey—Quality) was comprised of the 13 quality and satisfaction content items from the CSS (all loadings between .75 and .94). The second factor (which we termed Client Satisfaction Survey—Benefit) was comprised of the five benefit items (all loadings between .65 and 1.0). A confirmatory factor analysis for the two factors fit the data well and factor scores were saved for subsequent analyses.

Prior child welfare reports and recidivism. Past and future child welfare reports were drawn from administrative data held in the statewide child welfare report database. Matches were executed to identify all child welfare reports involving the study subject as the perpetrator using a combination of database identifiers additionally verified by date-of-birth match. Following the lead of Drake and colleagues (Drake, Jonson-Reid, Way & Chung, 2003; Kohl, Jonson-Reid & Drake, 2009), we did not infer meaningful report differences based on substantiation. During the course of the study and follow-up period, the state child welfare agency made procedural changes in how reports were handled and classified, including changes in rates of screening out and in rates of differential response. We opted to include all types of reports in the recidivism analyses so that recidivism data would remain less affected by these procedural changes. Report events were aggregated across maltreatment types, across children in the family and across report dates in order to yield unduplicated event counts and temporal sequences. Mean follow-up time (i.e., from enrollment to data extract) for future reports was 6.2 years. Surveillance reports (i.e., reports made by the home visitor) accounted for 3.5% of all recidivism incidents, and were retained in the recidivism data given that they were infrequent, all participants had similar levels of surveillance, and the proportion of reports that involved surveillance was not significantly related to treatment conditions.

Data Analysis Methods

Covariate control and propensity stratification solutions. In subpopulation studies drawn from cluster trials, covariate imbalance between treatment conditions is a risk, even if the full study sample is reasonably balanced. The covariate imbalance between treatments in our American Indian subpopulation was substantially greater than the full study. Consequently, we opted to develop specific balancing solutions tailored to the American Indian subpopulation. Two quasi-experimental

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

strategies were used. Covariate control was used as the primary method. For recidivism outcomes, the study hypothesis was equivalency to the full study, so we constructed a covariate control strategy mirroring that used in the full study. This began, as in the full study, by constructing an historical recidivism risk estimate. Historical risk was estimated using a sample of 527 American Indian parents in child welfare seen at the same agencies and in the same home-based service programs prior to the start of the SafeCare trial, plus our 354 study parents. Twelve shared covariates for the historical and study subpopulation parents were used as predictors. These included scores on the Beck Depression Inventory and Family Resources Scale, marital status, gender, education, age, number of children, age of youngest child, number of prior child welfare referrals, plus the service agency, time, and agency × time trends for clients served by the agency. In the prediction model, these 12 covariates predicted observed recidivism hazard for the historical subjects, and estimated (i.e., missing) recidivism hazard for the study subpopulation subjects. The predicted hazard for study subpopulation subjects was outputted for use as a covariate, and might be conceptualized as an estimate of recidivism risk under the counterfactual condition that the study was never conducted. A second recidivism risk covariate also was used and was taken directly from the full study. This was the county-level population-based reporting rate for each study subpopulation subject's county of residence, calculated as the number of reports made in the county divided by the number of children living in the county. This covariate reflects the report proneness of the subject's county of residence. Consequently, for the covariance strategy in the recidivism analyses, two covariates were used—this historical risk covariate (derived from historical American Indian cases and the 12 raw covariates) and county-level report proneness. This mirrors the covariate approach used in the full study and allows a better direct comparison of recidivism outcomes.

For covariate control among the psychometric outcomes (for which we are hypothesizing differences rather than comparability), we began constructing the covariate controls by testing baseline balance between treatment conditions on 45 covariates, examining each of the six pairwise comparisons among the four design cells, for a total of 270 comparisons examined. Based on the comparisons, we selected 10 client-level covariates and 8 home-visitor-level covariates to be included in future models. Client covariates were age, gender, education, number of children, number of prior child welfare reports, any alcohol or drug disorder, history of sexual abuse, and scores on the Beck Depression Inventory, Social Provisions scale, and Child Abuse Potential Inventory. Home visitor covariates were age, gender, education level, percent of time devoted to nonclient duties, number of clients seen, years of experience in the program, full-time work status, and American Indian ethnicity.

The covariance strategies estimated treatment effects at the multivariate mean of the covariates by adjusting the overall effect estimate. Covariance offers the advantage, when combined with multiple imputation, of modeling outcomes using the full sample. Covariance models can provide excellent estimates of true causal effects under many but not all circumstances (see Cook, Shadish, & Wong, 2008; Shadish, Clark, & Steiner, 2008). No one quasi-experimental approach is consistently superior for capturing true causal effects (Coalition for Evidence Based Policy, 2012), and so testing the same hypotheses using different but complimentary approaches can be an informative strategy. The complimentary approach we selected for a second set of models was propensity stratification (Rubin, 2006). The propensity approach follows a very different logic from the covariance approach, and estimates treatment effects separately among homogeneous strata, and these estimates are then weighted and pooled. Stratification came at the expense of excluding those cases that could not be grouped or matched into a stratum that was both reasonably homogeneous and reasonably populated with members of all treatment conditions. This resulted in substantially reduced sample sizes for the propensity models (see Table 1) and consequently reduced power. However, it will allow us to check estimates from the full-sample covariance approach to see if they are supported by estimates from the propensity approach.

Separate propensity stratification solutions were formed for each individual outcome variable following four steps. First, each of the 45 baseline demographic covariates was tested, then selected if it univariately predicted the outcome variable at $p < .15$. Between 7 and 12 predictors were identified for each outcome. Second, the set of outcome predictors was entered into two forward selection logistic models for predicting each of the two treatment dimensions (SafeCare vs. Services As Usual/ Coached vs. Uncoached), requiring inclusion of at least four predictors. Propensity scores (i.e., predicted probabilities of treatment group membership) for each dimension in the 2 × 2 design were outputted, and then cases were grouped into six strata for each outcome using K-means clustering of the two respective propensity scores. Cartesian scatter plots were constructed for each of the four cells of the experimental design, and strata observed to be imbalanced with respect to experimental condition representation were excluded (e.g., fewer than 4 members of any condition, or fewer than 10 members of any dimension). The number of propensity strata and the total number of cases included differed for each outcome's solution.

Handling missing data. Recidivism outcomes had no missing data, for either the outcome or the two risk covariates. For the other covariate models, baseline missing data were not common, but longitudinal missing waves were common due to attrition (see Data Collection section). First, we checked for selection factors in the missing data, modeling missingness as a function of experimental conditions. There were no significant attrition related missing data pattern differences between SafeCare versus Services as Usual (Wald = .44, $df = 1$, $p = .51$), between Coached versus Uncoached (Wald = 0.003, $df = 1$, $p = .96$) or in the interaction (Wald = .02, $df = 1$, $p = .90$). Next, we

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

Table 1. Effects From Propensity and Covariate Models

Outcome	Effect	Propensity strata (total *n*)	Covariance model *N*	Propensity model effect	Covariance model effect	Propensity model effect size[a]	Covariance model effect size[a]
Recurrent recidivism survival							
Full subpopulation	SafeCare	4 (218)	354	−0.01	−0.03	0.99	0.97
	Coaching	4 (218)	354	0.04	−0.01	1.04	0.99
SafeCare inclusion subpopulation	SafeCare	4 (135)	219	−0.31	−0.32	0.73	0.73
	Coaching	4 (135)	219	0.05	0.04	1.05	1.04
Beck Depression Inventory slope							
Full subpopulation	SafeCare	4 (251)	354	−2.70	−3.84*	−0.24	−0.33
	Coaching	4 (251)	354	−0.05	−0.92	0.00	−0.08
SafeCare inclusion subpopulation	SafeCare	4 (133)	219	−3.40	−3.40*	−0.32	−0.32
	Coaching	4 (133)	219	0.88	−0.49	0.08	−0.05
Child Abuse Potential Inventory slope							
Full subpopulation	SafeCare	4 (234)	354	0.41	0.01	0.00	0.00
	Coaching	4 (234)	354	−0.56	0.57	−0.01	0.01
SafeCare inclusion subpopulation	SafeCare	4 (148)	219	−0.37	−0.53	0.00	−0.01
	Coaching	4 (148)	219	−0.41	−8.25	0.00	−0.08
Working Alliance Inventory item mean							
Full subpopulation	SafeCare	4 (145)	202	0.51+	0.49*	0.40	0.38
	Coaching	4 (145)	202	0.26	0.32+	0.20	0.25
SafeCare inclusion subpopulation	SafeCare	4 (88)	129	0.86*	0.74**	0.40	0.58
	Coaching	4 (88)	129	0.19	0.17	0.26	0.13
Consumer Satisfaction—Quality item mean							
Full subpopulation	SafeCare	4 (139)	202	0.25	0.33*	0.26	0.35
	Coaching	4 (139)	202	0.19	0.16	0.20	0.17
SafeCare inclusion subpopulation	SafeCare	4 (91)	129	0.46*	0.54**	0.47	0.55
	Coaching	4 (91)	129	0.21	0.15	0.21	0.15
Consumer Satisfaction—Benefit item mean							
Full subpopulation	SafeCare	3 (136)	202	0.30**	0.34*	0.33	0.37
	Coaching	3 (136)	202	0.11	0.06	0.12	0.07
SafeCare inclusion subpopulation	SafeCare	3 (91)	129	0.33+	0.41*	0.35	0.44
	Coaching	3 (91)	129	−0.17	−0.06	−0.18	−0.06
Client Cultural Competency Inventory Mean for respect items							
Full subpopulation	SafeCare	3 (120)	202	0.17	0.26+	0.20	0.31
	Coaching	3 (120)	202	0.09	0.07	0.11	0.08
SafeCare inclusion subpopulation	SafeCare	3 (77)	129	0.21	0.49**	0.24	0.57
	Coaching	3 (77)	129	−0.03	0.14	−0.03	0.16

[a]Survival effects sizes presented as Hazard Ratios. Continuous outcome effect sizes presented as effects/pooled baseline standard deviation for the measure, analogous to Becker's *G*.
$^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

employed a multiple imputation approach to manage the limited covariate missingness in covariance modeling approach, which would allow us to use the full sample for these models. Imputation relied on the Bayesian multiple imputation Markov chain Monte Carlo (MCMC) procedure of Asparouhov and Muthén (2010a, 2010b) available through Mplus version 6.1 (Muthén & Muthén, 2010). A total of 45 covariates were included in the imputation procedure. Treatment conditions, observed recidivism survival, and observed covariate values were allowed to predict missing values. Beck Depression Inventory and Child Abuse Potential Inventory scores were to be used in growth models, which can accommodate missingness, but we opted to impute the significant amount of missing values for these outcomes, given that imputation allows more precise missing value estimation than the normal growth modeling approach. We did not impute missing values for the univariate outcomes (i.e., Client Satisfaction Survey, Working Alliance Inventory, and Client Cultural Competency Inventory). A set of 30 imputation data sets were generated and used for all the covariance based models with results combined following Rubin's (1976) rules.

Adjustment for clustering. The data structure in the study was complex with clients nested within home visitors and cross-classified within counties. Home visitors were themselves nested within supervisory teams and supervisory teams were nested within agencies and regions. There were 134 of the

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

219 total home visitors from the full study who served one or more American Indian parents, but more than half (52%) of these 134 home visitors had three or fewer American Indian clients. Given these constraints, we opted to simplify the nesting structure and treat cases as clustered within home visitors using a population average strategy, and employed a robust maximum likelihood estimation approach with a sandwich estimator in order to manage clustering.

Recidivism outcome analyses. For child welfare recidivism outcomes, Cox proportional hazard models were fitted using a recurrent event framework similar to the approach used in the full study. Up to four separate recurrent events were coded using a gap-time strategy, then modeled as reflecting a single latent hazard with a random (i.e., frailty) effect (Masyn, 2009). Equality constraints were imposed for loadings after the initial recurrence. The latent hazard was then modeled as a function of treatment conditions and their interaction. If the interaction did not approach significance, it was dropped and a main effects model was tested.

Growth models. For multiwave continuous psychometric outcomes (i.e., Child Abuse Potential Inventory and the Beck Depression Inventory), piecewise latent growth curve models were used regressing latent intercepts, slopes and fixed piecewise terms on the treatment effects, and their interaction. If the interaction did not approach significance, it was dropped and a main effects model was tested. Random intercept and slope terms were modeled alongside a fixed piecewise adjustment to means at follow-up to capture any nonlinearity in the observed outcome trajectories. Latent intercepts and slopes were allowed to covary, and equality restraints were imposed on residuals, both of which are common growth modeling assumptions. Treatment effects on slopes were the outcome of main interest. Child Abuse Potential Inventory and Beck Depression Inventory scores were adjusted at each wave for the corresponding Child Abuse Potential Inventory Lie scale score to correct for self-report response bias, with equality constraints imposed on the bias correction coefficient across waves.

Posttreatment service satisfaction and cultural competency models. Under the covariance approach, posttreatment client ratings of services on the Client Satisfaction Survey—Quality factor score, the Client Satisfaction Survey—Benefit factor score, the Working Alliance Inventory, and the Client Cultural Competency Inventory—Respect score were tested using a single multivariate model, modeling these four outcomes jointly as a function of the covariate set, the treatment conditions, and the treatment interaction. If the interaction did not approach significance, it was dropped and a main effects model was tested. Because separate propensity strata were constructed for each of these four outcomes, the propensity approach tested each measure univariately.

Results

Child Maltreatment Recidivism

In the full study, a significant hazard ratio (.74) in favor of SafeCare was found within the customary SafeCare inclusion population (Chaffin et al., 2012). We hypothesized that the American Indian subpopulation recidivism reduction effect would be approximately equivalent, which we operationalized as at least falling within the 95% confidence interval for the full study effect (.58 to .95). Note that this does not mean comparing significance levels. As an initial step, we performed a power analysis for the subpopulation study $n = 354$ based on observed full study effect sizes. Power was estimated using the Cox Proportional Hazard module of the Power Analysis and Sample Size (PASS) software package. Power was estimated to be between 0.16 and 0.26. Given these very low power estimates, we concluded that effects comparable in size to those in the full study would be unlikely to be significant at $p = .05$ in the subpopulation analysis. The low observed power supported not using p values to assess equivalency.

There were 862 total unduplicated post-enrollment reports among the 354 subjects across an average 6.2-year follow-up; 247 or 70% had at least one post-enrollment report to child welfare, 185 (52%) had two or more, 137 (39%) had three or more, and 99 (28%) had four or more. Model findings for recidivism are displayed in the topmost rows of Table 1. Using the covariance approach, recurrent event hazards were modeled as a function of treatment conditions, and a hazard ratio of .73 was found among the customary SafeCare inclusion population. Covariance-based models and propensity models yielded virtually identical estimates (see Table 1). For the extended population, including those outside customary inclusion criteria, the effect among the American Indian subpopulation was almost nonexistent and was barely within the 95% confidence interval from the full study (.70 to .98).

Beck Depression Inventory

Using the covariance approach, growth models for the Beck Depression Inventory found between 3.4 and 3.8 points greater decreases in scores for SafeCare participants, corresponding to a Becker's G analogue treatment effect size of −.32 to −.33 (treatment effect on slope/pooled baseline standard deviation for the measure). Models for the full subpopulation and the customary SafeCare inclusion population reached significance, no coaching or interaction effects approached significance, and the estimates were supported by similar propensity-based estimates (see Table 1).

Child Abuse Potential Inventory

No treatment effects approached significance for the Child Abuse Potential Inventory. The model-based slope estimate for change on the measure was a decrease of 43 points from baseline to posttreatment after adjusting for response bias. This corresponds to a Becker's G effect size estimate of .41 (mean

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

slope/baseline pooled standard deviation for the measure), but this decrease did not differ among any of the treatment conditions. The covariate approach effects were supported by similar propensity approach effects (see Table 1).

Working Alliance Inventory, Client Satisfaction Inventory, Client Cultural Competency Inventory

Significant main effects in favor of SafeCare were found for the Working Alliance Inventory in both the full and customary SafeCare inclusion groups (Becker's $G = .38$ and .58 respectively; see Table 1), along with a trend ($p = .06$) for the Coaching effect among the full sample (Becker's $G = .25$). These were supported by similar size effects using the propensity approach. The mean item response on the Working Alliance Inventory was 4.75 ($SD = 1.3$; median = 5.1) on a 0–6 scale with higher scores indicating greater working alliance. The model estimate indicated that SafeCare raised the Working Alliance Inventory score by between .49 and .74 points in the full and customary inclusion groups, respectively.

Significant main effects in favor of SafeCare were also found for the Client Satisfaction Survey—Quality factor, among both the full and the customary inclusion groups (Becker's $G = .35$ and .55, respectively; see Table 1). Significant main effects in favor of SafeCare were also found for the Client Satisfaction Survey—Benefit factor among both groups (Becker's $G = .37$ and .44, respectively). The mean item response on the Client Satisfaction Scale was 3.45 ($SD = 0.53$; median = 3.6) on a 1–4 scale with higher scores indicating greater satisfaction, suggesting that in general services were highly rated.

A significant main effect in favor of SafeCare was found on the Client Cultural Competency Inventory—Respect factor for the customary inclusion group, and a trend ($p = .07$) in the full subpopulation group (Becker's $G = .57$ and .31, respectively). The mean item response on the Client Cultural Competency Inventory was 4.15 ($SD = .77$; median = 4.38) on a 1–5 scale, suggesting that services in general were highly rated.

Discussion

Among the customary SafeCare inclusion criteria group, which is the normal group in which effectiveness would be tested, the recidivism reduction hazard ratio was .73, which is virtually identical to the .74 value among the corresponding group in the full study. This strongly supported the main study prediction of equivalent recidivism reduction. There were significant effects in multiple domains, consistently pointing in favor of SafeCare. These included (a) reduced parental depression symptoms; (b) better working alliance ratings; (c) better service quality ratings; (d) better service benefit ratings; and (e) higher ratings of cultural competency. Although substantial overall improvement was noted on the Child Abuse Potential Inventory, the amount of improvement was equivalent across treatment conditions.

Almost no recidivism reduction was observed outside the customary SafeCare inclusion population (HR = .97), although the estimate was barely within the full study 95% confidence interval. Although it appears that the SafeCare model performs well with its intended population (parents of preschoolers), it does not appear to confer either advantage or disadvantage outside its design limits with American Indian parents. Additional work to adapt the model to parents of older children should be explored before extending the model outside its intended service population.

The second study hypothesis concerned the acceptability of a structured, manualized, behavioral model like SafeCare among American Indian parents. Contrary to some concerns in the literature about manualized dominant culture based models being a poor fit with American Indian world views or unacceptable to American Indian parents, the basic unmodified SafeCare model was not associated with any observed cultural competency or engagement problems. In fact, SafeCare was rated by American Indian parents as higher quality, more beneficial, more culturally competent and producing a better working alliance with their home visitor. Consumer ratings were not only better than for Services As Usual, but were high in absolute terms. This is consistent with the general literature on cross-cultural application of evidence-based models (Huey & Polo, 2008). We might speculate that the greater acceptability of the model could be related to role of structure and protocol itself. Ceremony and protocol are strong themes among many American Indian cultures and perhaps SafeCare was congruent with this tradition. Or it may be that American Indian parents, like many others, simply tend to favor services with organized content that focuses on practical everyday skills that they can use with their children. A strength of the study is that the acceptability and cultural competency of the services was measured directly from consumers themselves, rather than inferred conjecturally. Consumers made their ratings under conditions of high privacy, confidentiality, using computerized interviews, and without their home visitor present. This data collection strategy was designed to reduce any demand characteristics or bias in this type of data. Concerns might also be raised about the intrusiveness of in vivo coaching or fidelity monitoring as an implementation quality control strategy, because it interjects an outsider into the home visitor–client relationship. No support for this was found. In fact, working alliance tended to improve ($p = .06$) with the in vivo coaching quality control strategy.

There is a final secondary finding of note in this study. To our knowledge, this is the first study to examine the measurement properties of several standard psychometric measures with a moderate size sample of American Indian parents in child welfare. This included testing internal consistency, conducting exploratory and confirmatory factor analyses, some construct validity testing, and examining longitudinal correlations. Measures used in the study exhibited good internal consistency, stability, and construct validity. In fact, most indices were very close to those reported in the literature with other populations. This suggests that it is reasonable to use these measures in research with American Indian parents in child welfare.

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

Some limitations of the study need to be considered. The design did not randomize individual cases or home visitors to conditions. Allocation was at aggregate levels which were complexly structured, creating a practical necessity for simplification in the analysis. The efficiency of these methods in managing all possible dependencies cannot be proven or disproven. As is often a risk with subpopulation studies from cluster design trials, there was covariate imbalance which needed to be managed in order to estimate causal effects. Some analyses also may have been affected by missing data. Covariate control with multiple imputation of missing values was used to improve rigor in the context of this design and was supported with propensity stratification. More involved missing data approaches, such as pattern mixture models, are generally limited to larger sample studies. However, we have employed pattern mixture modeling in a study using the full Oklahoma trial data set and found that it produced estimates that were generally congruent with approaches similar to or more limited than those used here (Chaffin & Bard, 2011). We observed no treatment related difference in missing data patterns. As always with these types of methods and issues, unobserved covariate imbalances and non-ignorable missingness may still exist. The sufficiency of our methods for managing missing data and covariate imbalance can be neither proven nor disproven.

It is important to recognize that this study did not assess individuals' specific tribal affiliations or level of cultural traditionality. Obviously, capturing sufficient samples for analyzing specific cultural tenets would involve a much more focused study and very large (or even unfeasibly large) samples. Oklahoma is home to 39 federally recognized American Indian tribes, as well as individuals from over 100 other tribes, so even a large scaled-up statewide study such as this one would be unlikely to capture a sufficient sample of from each cultural tradition and level of traditionality. The results of this study should be considered a broad or macro level view, rather than an indication of SafeCare fit with any specific culture. This broader and more macro view of cultural fit will be an important consideration for state and tribal child welfare policy makers who often must adopt models based on their ability to fit diverse cultures.

Several strengths of the study should be considered. The study tested SafeCare outcomes in a fully scaled-up context. Evidence-based treatments often show attenuated benefit when rolled-out into large scale field implementations, so it is encouraging to see effects in this context. It is also important to note that the study comparisons were inherently conservative. The comparison condition, which we believe was a good quality home-based program, was comparable or even identical in almost all respects to the SafeCare condition, with the exception of the SafeCare content and curriculum modules themselves. This offers a strong and specific test of the SafeCare elements. Comparisons with less credible or minimal services might well show larger relative SafeCare benefits. Finally, to our knowledge, this is the largest American Indian parent sample in a controlled child welfare outcome trial that has heretofore been examined.

In summary, the findings suggest that the SafeCare model is a reasonable fit with American Indian parents in child welfare who meet customary SafeCare inclusion criteria. In contrast to concerns that structured, manualized, or evidence-based models are a poor cultural fit with American Indian populations, services in general were highly rated by consumers, with SafeCare being rated higher on average in these regards. The study also points to areas for future adaptations of the SafeCare model if it is to be extended to broader child welfare populations, including needed adaptations for parents of older children.

Acknowledgments

Additional in-kind support was provided by the Violence Prevention Branch of the U.S. Centers for Disease Control and Prevention. The opinions expressed are those of the authors and do not necessarily reflect those of Casey Family Programs, the NIMH, or the CDC. The authors wish to recognize the contributions of John Lutzker; Randy Campbell; Kathy Bigelow; Jill Filene; Dan Whitaker; Steve Ross; Gina Carrier; Tyler Corwin; Peter Pecora; Sue Steib; staff and leadership at the Oklahoma Department of Human Services including Howard Hendrick, B. K. Kubiak, J. J. Jones, John Gelona and Kathy Simms; and the leadership and staff of the Oklahoma Children's Services network agencies

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This project was supported by a contract with Casey Family Programs and by grants R01MH065667 (M. Chaffin, PI) and R01MH072961 (G. Aarons, PI) from the National Institute for Mental Health

Note

1. We use the term American Indian throughout this manuscript, though other language is sometimes used including the broader term, Native Americans which can include Alaska Natives and other peoples. In Oklahoma, where this study took place, the more specific term, American Indian, is commonly used.

References

Asparouhov, T. & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. Technical Report. Version 3. Downloaded from www.statmodel.com 8/08/2012.

Asparouhov, T. & Muthén, B. (2010). Multiple imputation with Mplus. Technical Report. Version 2. Downloaded from www.statmodel.com 08/08/2012.

Barth, R. P., Landsverk, J., Chamberlain, P., Reid, J. B., Rolls, J. A., Hurlburt, M. S., & Kohl, P. L. (2005). Parent training programs in child welfare services: Planning for a more evidence-based approach to serving biological parents. *Research on Social Work Practice*, *15*, 353–371.

Barlow, A., Varipatis-Baker, E., Speakman, K., Ginsburg, G., Friberg, I., Goklish, N., & Walkup, J. (2006). Home-visiting intervention to

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014

improve child care among American Indian adolescent mothers. *Archives of Pediatrics & Adolescent Medicine*, *160*, 1101–1107.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of beck depression inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment*, *67*, 588–597.

Chaffin, M., & Bard, D., (2011). Changes in parental depression symptoms during family preservation services. *Child Abuse and Neglect*, *35*, 448–458.

Chaffin, M., & Friedrich, W. (2004). Evidence-based practice in child abuse and neglect. *Children and Youth Services Review*, *26*, 1097–1113.

Chaffin, M., Hecht, D., Bard, D., Silovsky, J., & Beasley, W. H. (2012). A statewide trial of the safecare home-based services model with parents in child protective services. *Pediatrics*, *129*, 509–515.

Chaffin, M., Kelleher, K., & Hollenberg, J. (1996). Onset of physical abuse and neglect: Psychiatric, substance abuse, and social risk factors from prospective community data. *Child Abuse & Neglect*, *2*, 191–203.

Chaffin, M., & Valle, L. A. (2003). Dynamic prediction characteristics of the child abuse potential inventory. *Child Abuse & Neglect*, *27*, 463–481.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*, 724–750.

Cruz, C. M., & Spence, J. (2005). Oregon tribal evidence based and cultural best practices. *Oregon Department of Human Services: Office of Mental Health and Addiction Services*. Retrieved December 15, 2010, from http://www.oregon.gov/DHS/mentalhealth/ebp/tribal-ebp-report.pdf?ga=t

Coalition for Evidence-Based Policy. (2012, February). *Which comparison-group ("quasi-experimental) study designs are most likely to produce valid estimates of a program's impact?* Washington, DC: Coalition for Evidence Based Practice.

Damashek, A., Bard, D., & Hecht, D. (2012). Provider cultural competency, client satisfaction, and engagement in home-based programs to treat child abuse and neglect. *Child Maltreatment*, *17*, 56–66.

Drake, B., Jonson-Reid, M., Way, I., & Chung, S. (2003). Substantiation and recidivism. *Child Maltreatment, 8*, 248–260.

Duggan, A., Caldera, D. L., McFarlane, E., Fuddy, L., Burrell, L., Higman, S. M., . . . Sia, C. (2007). Impact of a statewide home visiting program to prevent child abuse. *Child Abuse & Neglect 31*, 801–827.

Duggan, A., McFarlane, E., Fuddy, L., Burrell, L., Higman, S. M., Windham, A., & Sia, C. (2004). Randomized trial of a statewide home visiting program: Impact in preventing child abuse and neglect. *Child Abuse & Neglect*, *28*, 597–622.

Harvey-Berino, J., & Rourke, J. (2003). Obesity prevention in preschool Native-American children: A pilot study using home-visiting. *Obesity Research*, *11*, 606–611.

Huey, S. J., & Polo, A. J. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child and Adolescent Psychology*, *37*, 262–301.

Kohl, P.L., Jonson-Reid, M., & Drake, B., (2009). Time to leave substantiation behind: Findings from a national probability study. Child Maltreatment, 14, 17-26.

Lutzker, J. R., & Bigelow, K. M. (2002). *Reducing child maltreatment: A guidebook for parent services*. New York, NY: Guilford Press.

Masyn, K. E. (2009). Discrete-time survival factor mixture analysis for low-frequency recurrent event histories. *Research in Human Development*, *6*, 165–194.

Milner, J. S. (1986). *The child abuse potential inventory manual* (2nd ed.). DeKalb, IL: Psytec.

Miranda, J., Schoenbaum, M., Sherbourne, C., Duan, N., & Wells, K. (2004). Effects of primary care depression treatment on minority patients' clinical status and employment. *Archives of General Psychiatry*, *61*, 827–834.

Muthén, B. O., & Muthén, L. K. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthen & Muthen.

Red Horse, J. G., Martinez, C., & Day, P. (2001). *Family preservation: A case study of Indian tribal practice*. Seattle, WA: Casey Family Programs, National Indian child welfare Association.

Red Horse, J., Martinez, C., Day, P., Day, D., Poupart, J., & Scharmbarg, D. (2000). *Family preservation concepts in American Indian communities*. Portland, OR: National Indian child welfare Association.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Rubin, D. B. (2006). Estimating treatment effects from nonrandomized studies using subclassification on propensity scores. In D. A. Hantula (Ed.), *Advances in social and organizational psychology* (pp. 41–59). Mahwah, NJ: Lawrence Erlbaum.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*, 1344–1356.

Stoltenberg, C. D., & McNeill, B. W. (2010). *IDM supervision: An integrative developmental for supervising counselors and therapists* (3rd ed.). New York, NY: Routledge/Talyor & Francis Group.

Switzer, G. E., Scholle, S. H., Johnson, B. A., & Kelleher, K. J. (1998). The client cultural competence inventory: An instrument for assessing cultural competence in behavioral managed care organizations. *Journal of Child and Family Studies*, *7*, 483–491.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the working alliance inventory. *Psychological Assessment*, *1*, 207–210.

Walkup, J.T., Barlow, A., Mullany, B.C., Pan, W., Goklish, N., Hasting, R., Cowboy, B., Fields, Pl, Baker, E.V., Speakman, K., Ginsburg, G., & Reid, R. (2009). Randomized controlled trial of a paraprofessional-delivered in-home intervention for young reservation-based American Indian mothers. *Journal of the American Academy of Child Psychiatry, 48*, 591–601.

Downloaded from cmx.sagepub.com at GEORGIA STATE UNIVERSITY on March 31, 2014