American Journal of Orthopsychiatry
2007, Vol. 77, No. 4, 497–505

Copyright 2007 by the American Psychological Association
0002-9432/07/$12.00 DOI: 10.1037/0002-9432.77.4.497

Outcomes for Youth Receiving Intensive In-Home Therapy or Residential Care: A Comparison Using Propensity Scores

Richard P. Barth, PhD, MSW
University of Maryland

Johanna K. P. Greeson, MSS, MLSP,
Shenyang Guo, PhD, MA, and
Rebecca L. Green, MSW
University of North Carolina at Chapel Hill

Sarah Hurley, MA and Jocelyn Sisson, MS
Youth Villages

This study compares outcomes for behaviorally troubled children receiving intensive in-home therapy (IIHT) and those receiving residential care (RC). Propensity score matching is used to identify matched pairs of youth ($n = 786$) with equivalent propensity for IIHT. The majority of pretreatment differences between the IIHT and RC groups are eliminated following matching. Logistic regression is then conducted on outcome differences at 1 year postdischarge. Results show that IIHT recipients had a greater tendency (.615) toward living with family, making progress in school, not experiencing trouble with the law, and placement stability compared with RC youth (.558; $p < .10$). This suggests that IIHT is at least as effective for achieving positive outcomes. Given IIHT's reduced restrictiveness and cost, intensive in-home services should be the preferred treatment over RC in most cases.

Keywords: propensity score matching, multisystemic therapy, child welfare outcomes, residential care

The last seven decades have witnessed a movement away from residential care (Barth, 2005; Wolins & Piliavin, 1964), one that appears to be accelerating following critical reviews of residential care's efficacy (e.g., Burns & Hoagwood, 2002; Lyons & McCulloch, 2006) and the inability to meet the standards of evidence-based practice (Hair, 2005). Research is amassing that indicates that any short-term gains from residential treatment are often mitigated after discharge because of a lack of family involvement, inadequate teaching of adaptive skills, and problems with after-care planning (Quay, 1979, 1986; Whitaker & Pecora, 1984; Wilson & Lyman, 1983, as cited in Leichtman, 2006). Additionally, in light of concerns about the quality of care provided, protection of youth, and overall appropriateness of the treatment modality, there has been a call for action on various fronts to address the deficiencies in residential care (Pumariega, 2006).

Part of this concern stems from the emergence of home- and community-based alternatives to residential care. Optimally, these interventions are individualized, culturally competent, intensive home- and community-based options. Parents and professionals collaborate as allies in the treatment of troubled children and youth (Friedman, Pinto, Behar, Bush, Chirolla, Epstein et al., 2006). Such alternative models, including multisystemic therapy (MST), have already shown, under certain conditions, to result in outcomes that are at least as good as those of residential care (Henggeler et al., 2003).[1] Moreover, they have also been shown to be less costly than residential care (Hoagwood, Burns, Kiser, Ringeisen, & Schoenwald, 2001). Additionally, other forms of community-based treatment that engage parents intensively in the therapeutic process, like multidimensional-treatment foster care, have been shown to be superior to some forms of residential care, small-group home care in particular (Chamberlain & Reid, 1998; Eddy, Whaley, & Chamberlain, 2004; Leve & Chamberlain, 2005). This finding is supported by more general social science research showing that placing troubled youth together is likely to result in an increase in problem behavior (Dishion & Dodge, 2005).

Yet residential care services and group living arrangements remain quite common, and some states have more than 50% of their older adolescents in group care (Barth & Chintapalli, in press; F. Wulczyn, in press). The costs are staggering. Lyons and McCulloch (2006) indicate that the state of Illinois spent nearly 75% of their total mental health services budget for the 50,000 children in long-term residential treatment centers and psychiatric hospitals. Webster (1999) concluded that nearly 60% of California's out-of-home care budget was spent on children in residential care.

Richard P. Barth, School of Social Work, University of Maryland; Johanna K. P. Greeson, School of Social Work, Shenyang Guo, School of Social Work, and Rebecca L. Green, School of Social Work, University of North Carolina at Chapel Hill; Sarah Hurley and Jocelyn Sisson, Youth Villages.

This study was supported by a grant to Youth Villages (Memphis, TN) from the Hyde Family Foundations. Findings, opinions, and recommendations expressed in this article are those of the authors and not necessarily those of Youth Villages or the Hyde Family Foundations.

For reprints and correspondence: Richard P. Barth, School of Social Work, University of Maryland, 525 West Redwood Street, Baltimore, Maryland 21201. E-mail: rbarth@ssw.umaryland.edu

[1] A heated debate about the quality of the research that supports the efficacy claims of MST has surfaced (see Henggeler, 2006; Littell, 2006).

Although the evidence base for in-home services is mounting, the extent to which such services outperform residential treatment remains relatively untested. To our knowledge, there are no studies that rigorously compare outcomes for in-home services and residential care among youth. Lyons and McCulloch (2006) argue that this lack of research can be traced back to the relatively low rates of use of residential care and the diversity of reasons and presenting problems among those entering residential care. They argue that residential treatment should not be judged as ineffective because of the lack of randomized clinical trials and provide evidence that behavior problems do decline in well-administered programs. They contend that the development of large-scale databases on the outcomes for youth in residential treatment settings offers the best hope for advancing the field's understanding of the outcomes of residential care. We agree, in part, with this contention but assert that large and longitudinal data sets do not offer much additional information unless they are compared with other interventions. Yet, given the enormous cost of residential care and its restrictiveness, even a modest decline of symptoms during treatment must be considered suspect without a carefully selected counterfactual group to help pinpoint the likelihood that such improvements did not occur spontaneously and would not have occurred in community-based settings.

The purpose of the present analysis is to demonstrate whether intensive in-home therapy (IIHT) derived from MST is more effective than traditional residential care (RC) for behaviorally difficult youth. Because randomized clinical trials contrasting the use of RC and IIHT are relatively costly and difficult and remain rarely accomplished, other means to make comparisons between outcomes for IIHT and RC deserve consideration. To have credibility, however, these methods must contend with the likelihood that youth who are placed into RC have greater difficulties than youth placed into IIHT.

The literature on RC includes commentary that youth who enter RC have often failed at other placements (e.g., Baker, Wulczyn, & Dale, 2005), suggesting that the youth who enter group care are the worst cases that have previously been unable to be treated by community services. Yet there is also countervailing evidence that a sizable proportion of youth enter group care directly, without previous efforts to treat them in less restrictive settings (James et al., 2006). Further, data from a national probability study of children in out-of-home care indicate that youth who enter group care directly after child welfare intake are nearly twice as likely to have a clinical or borderline level of problem behaviors than those remaining at home and receiving conventional child welfare services or those entering kinship care or foster care (Wulczyn, Barth, Yuan, Jones Harden, & Landsverk, 2005). Differences in case characteristics between youth who receive IIHT and RC clearly need to be understood and controlled for, to the greatest extent possible, to allow for the most useful comparisons between IIHT and RC.

Part of the development of this understanding requires that we determine what differences between characteristics of children who enter IIHT and RC are associated with the outcomes of interest. After all, differences that are not related to the outcomes may not bias the results. Yet information about the characteristics of cases that predict outcomes in RC is scant and inconsistent. Lyons and McCulloch (2006) found that, after placement into RC, age, diagnosis, and race predicted symptom improvement. Baker, Wulczyn, and Dale (2005) concluded that youth in RC who ran away were likely to have had prior substance abuse histories, were older at the time of placement, and had incarcerated parents. Similarly, Gorske, Srebalus, and Walls (2003) observed that youth with antisocial behaviors were at greater risk for unsuccessful treatment, as defined by premature discharge, running away, or moving to an equally restrictive setting. Surprisingly, Stage (1999) conversely concluded that a history of antisocial behaviors, victimization, and family dysfunction did not predict discharge status from RC. Peterson and Scanlan (2002) also found divergent results. A primary diagnosis of conduct disorder was associated with better functioning in family or foster home environments than group homes. As these studies suggest, the effectiveness of RC for behaviorally difficult youth is still being debated.

A significant challenge to generating better comparisons between RC and intensive in-home services is the lack of randomized trials. When these are not feasible, because of cost or unwillingness of service providers or courts to agree to randomize, quasi-experimental alternatives are needed. In this study, propensity score matching (PSM), substantially developed through the work of Nobel Laureate James Heckman (1978), is used. To generate the most rigorous comparison of outcomes for youth who participate in RC and those who participate in IIHT, propensity score analysis provides statistical controls to generate matched groups of youth entering RC and IIHT (Rosenbaum & Rubin, 1983). The use of propensity score matching helps to mitigate selection factors that might result in different likelihoods of service receipt in RC versus in-home treatment (Guo, Barth, & Gibbons, 2006).

Method

Design and Sample

The agency involved in the present investigation is a large provider of behavioral health services for troubled children and their families, with 44 locations in seven states and the District of Columbia. The agency offers a complete continuum of programs and services. Standardized data from closed-case files were extracted from the agency's management information system database. Data used in this investigation were from an overall population of 1,369 youth, of whom 937 received IIHT (but had not been in RC with this agency) and 432 received RC (but had never received IIHT from this agency). These youth had both their intake and psychosocial assessments within 45 days of admission to the agency. Table 1 presents the demographic characteristics of the study sample.

This study used listwise deletion to handle missing data. Although this is not a perfect solution, listwise deletion is currently the best available approach under the PSM framework (Allison, 2001). Mean substitution was not selected because of the potential for creating selection bias, and multiple imputation of missing data was not selected because the effects of the approach on PSM are unknown. Sound methods for handling missing data under the context of PSM have yet to be developed.

Description of IIHT. Beginning in 1994, this agency became the first community-based agency to provide MST for dissemination purposes outside of MST Services' clinical trials. Since the origination of this relationship, supervision has been provided by MST Services based on the same sources of information used for

Table 1
Youth Treatment Group for Overall and Matched Samples (%)

Variable	Overall IIHT ($n = 937$)	Overall RC ($n = 432$)	Matched IIHT Scheme 1 ($n = 393$)	Matched RC Scheme 1 ($n = 393$)	Matched IIHT Scheme 2 ($n = 408$)	Matched RC Scheme 2 ($n = 408$)
Race (other)$^{+,***,b,e}$						
African American	21.9	36.6	32.3	43.5	34.6	44.1
Gender (female)$^{+,***}$						
Male	67.1	81.3	80.7	75.8	80.4	78.4
Age group$^{+,***,b,f}$						
0-11 years	31.7	16.4	17.3	26.7	17.2	33.3
12-15 years	47.2	61.3	61.6	50.9	61.5	43.4
16 years	21.1	22.2	21.1	22.4	21.3	23.3
Site (TN)*						
MS	13.8	9.9	9.9	13.7	10.1	14.0
Presenting problem of delinquency (No)$^{+,***}$						
Yes						
Number of delinquency types***	62.5	75.5	73.3	69.1	74.8	76.2
0	37.5	24.5	26.7	30.8	25.3	23.8
1-2	29.1	34.7	34.1	33.1	34.3	35.8
3+	33.4	40.7	39.2	36.1	40.4	40.4
Presenting problem of mental health problems (No)^{***,a}						
Yes	48.9	60.4	58.3	49.4	59.6	61.5
Number of mental health problems$^{+,***,a,d}$						
0	51.1	39.6	41.7	50.6	40.4	38.5
1-2	33.4	37.5	36.9	33.1	38.0	46.6
3+	15.5	22.9	21.4	16.3	21.6	15.0
Presenting problem of maltreatment (No)						
Yes	32.3	36.3	36.1	37.4	35.8	39.7
Number of maltreatment types						
0	67.7	63.7	63.9	62.6	64.4	60.3
1	21.6	22.0	21.1	24.9	21.8	28.2
2	8.7	12.5	13.2	10.7	12.3	9.3
3	2.0	1.8	1.8	1.8	1.7	2.2
Presenting problem of substance abuse (No)						
Yes	17.9	21.8	21.4	19.6	22.1	22.3
Commission of a Status Offense (No)$^{+,**,b,d}$						
Yes	10.9	17.8	17.1	26.0	16.4	22.1
Simple assault (No)						
Yes	4.4	6.5	6.1	8.1	6.4	7.1
Other criminal behavior (No)$^{+,***}$						
Yes	10.7	18.5	17.1	19.6	16.9	20.1
Presenting problem of runaway (No)						
Yes	22.8	27.3	26.2	26.7	27.9	27.5
Presenting problem of gang (No)						
Yes	3.9	4.6	4.1	3.6	4.7	4.7
Committed a sex offense (No)$^{+,***}$						
Yes	10.8	18.3	16.8	17.6	17.2	20.6
Siblings in out-of-home care (No)**						
Yes	1.5	4.4	2.8	3.6	4.2	2.0
Past mental health treatment (No)***						
Yes	45.9	70.4	68.7	72.3	68.9	65.0
Past inpatient treatment (No)***						
Yes	25.4	42.6	40.5	42.5	41.7	35.5
Past outpatient treatment (No)***						
Yes	30.0	46.1	43.8	46.8	45.1	40.7
Past foster care placement (No)						
Yes	4.6	6.7	6.1	7.9	6.6	7.4
Received special education services (No)$^{+,***}$						
Yes	15.7	25.5	24.7	26.0	23.0	27.5
Received financial assistance (No)						
Yes	35.2	40.3	39.7	39.2	39.7	40.9
Youth's prior legal charges (No)*						
Yes	26.0	32.6	32.1	30.5	31.6	30.9

Reference group for categorical variables is shown next to the variable name.
Scheme 1 conditioning variables.
$^{+}$Scheme 2 conditioning variables.
Prematched sample differences (χ^2 tests), $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.
Postmatched sample differences, *Scheme 1* (χ^2 tests), $^{a}p < .05$, $^{b}p < .01$, $^{c}p < .001$.
Postmatched sample differences, *Scheme 2* (χ^2 tests), $^{d}p < .05$, $^{e}p < .01$, $^{f}p < .001$.

all other MST replications, including information on therapist fidelity provided by families. The provider agency adheres to MST Services' protocol for monitoring model fidelity. They routinely complete two types of adherence measures (Therapist Adherence Measure and Supervisor Adherence Measure); these and other indicators are reviewed with MST staff to prevent model drift. During the entire study period, the agency was a licensed MST provider. This offers additional assurance of model fidelity within the agency.

The difference between IIHT and MST involves the population of clients served. IIHT has expanded services to include youth presenting primarily with mental health issues and children in state custody as a result of abuse, neglect, or delinquency. Yet there is considerable evidence that maltreated youth also have significant levels of antisocial behavior problems and delinquent behavior (e.g., Wall & Barth, 2005). Thus, although the IIHT population may differ from populations served in research studies done on MST, the behaviors targeted by the intervention are comparable.

Description of RC. The three open residential campuses operated by the agency provide the setting for intensive mental health treatment based on the re-education of emotionally disturbed (Re-ED) youth therapy model (Hobbs, 1982; Hooper, Murphy, Devaney, & Hultman, 2000). Youth live in groups of 8 to 12 (based on age, sex, presenting issues, functional level, and other relevant characteristics such as physical size) in houses on each residential campus and are cared for by a staff of counselors, teachers, nurses, and cottage staff on a 24-hr basis. Psychotherapies, recreational therapy, and life skills training (conducted primarily by master's-level staff), along with medical care, educational services, and basic needs (clothing, housing, meals), are provided in a therapeutic milieu that focuses on achievement of treatment goals. Family involvement is required; while the child is in treatment, a family counselor works with the family to address issues in order for the child to return home. A child and adolescent psychiatrist is available to see the child on a scheduled and as-needed basis. Following best practice standards, children's strengths are assessed and a comprehensive, individualized treatment plan is developed. The treatment plan is monitored biweekly and adjusted as progress is made.

Purpose of the Present Analysis

The purpose of the present analysis is to compare youth who received IIHT with those who received RC with the aim of assessing the effectiveness of IIHT vis-à-vis 1-year postdischarge outcomes for school status, trouble with the law, and the case status of placement at home or in a homelike environment. Four steps are involved in the analysis process: (a) evaluation of bivariate differences between IIHT and RC youth; (b) logistic regression modeling to predict a propensity score; (c) calculation of the predicted probability of receiving IIHT; and (d) multivariate analysis of the outcome variable.

Propensity Score Matching

The interest in determining program effectiveness has led to a surge in work focusing on estimating average treatment effects under various sets of assumptions. Statisticians (e.g., Rosenbaum & Rubin, 1983) and econometricians (e.g., Heckman, 1978, 1979; Heckman, Ichimura, & Todd, 1997) have made considerable progress by developing a new approach (PSM) for estimating causal effects from observational data. The rationale for PSM and the matching procedures are described next, in substantial detail, because the quality of the matching process is important to the meaningfulness of comparing matched groups.

The primary advantage of PSM is its effective control of covariates and extraneous factors threatening internal validity through predicting propensity scores, matching on propensity scores, and multivariate analysis based on a matched sample or resample. Specifically, this approach involves the following four steps (Guo, Barth, & Gibbons, 2006):

1. Using prior research, theory, and experience, identify case characteristics that are likely to be associated with the assignment to conditions, so that these can be addressed with the PSM processes. These characteristics are provided in Table 1.

2. Use these predictor (or conditioning) variables in a logistic regression in which the dependent variable is a dichotomous variable indicating IIHT receipt versus RC receipt. This helps to ascertain whether the hypothesized case characteristics do indeed distinguish between the groups. Because the results of the propensity score are highly sensitive to the choice of variables included in the logistic regression, multiple models specifying different sets of conditioning variables must be tested.

3. Calculate the probability of receiving IIHT based on the estimated logistic regression. The logit of the predicted probability (i.e., $\log[(1 - p)/p]$) is defined as a propensity score for receiving IIHT. Match all RC youth on the propensity score with IIHT youth to create a matched sample or resample. Matching is done without replacement and includes two conditions: Matches must be the nearest case in the RC group to the IIHT group and must also fall within a predetermined caliper (range). The latter step is intended to prevent cases that are nearest neighbors, but do not have very similar propensity scores, from being included in the final matched data set.

4. Based on the matched sample, complete a multivariate analysis of the outcome variable of interest. The analysis in this article uses a three-level ordinal outcome variable: desirable outcome, mixed outcome, and undesirable outcome. An ordinal logistic regression is used to predict the probability of falling into each level of the three outcomes. Different sets of covariates are tested in the analysis.

Measures

Measures used in the PSM procedure are presented first, followed by a description of the measure used in the analysis of youth's outcomes. Two sets of variables were created for the stepwise selection of conditioning variables. The selection of these variables and formation of two schemes were based on a combination of what seemed plausible and what the analyses revealed as viable. The following variables were in the first set, which contained all demographic variables, intake data from psychosocial

assessments, service-related variables, and poverty status of the parents: race (African American–other), gender (male–female), age group (in years: 0–11, 12–15, 16+), presenting problem of delinquency (yes–no), number of delinquency types (range = 0 to 3+), presenting problem of mental health issues (yes–no), number of mental health problems (range = 0 to 3+), presenting problem of maltreatment (yes–no), presenting problem of substance abuse (yes–no), committed of a status offense (yes–no), other criminal behavior (yes–no), committed of a sex offense (yes–no), siblings in out-of-home care (yes–no), past mental health treatment (yes–no), past inpatient treatment (yes–no), past outpatient treatment (yes–no), receipt of special education services (yes–no), parent's receipt of financial assistance (yes–no), and youth's prior legal charges (yes–no). The stepwise procedure identified eight conditioning variables for the logistic regression (see Table 1).

The second set of variables only contained demographic characteristics, intake data on youth's psychosocial assessments, and a few service-related variables. Precisely, this set of variables included the following: race (African American–other), gender (male–female), age group (in years: 0–11, 12–15, 16+), presenting problem of delinquency (yes–no), number of delinquency types (range = 0 to 3+), presenting problem of mental health issues (yes–no), number of mental health problems (range = 0 to 3+), presenting problem of maltreatment (yes–no), presenting problem of substance abuse (yes–no), committed of a status offense (yes–no), other criminal behavior (yes–no), committed of a sex offense (yes–no), past mental health treatment (yes–no), receipt of special education services (yes–no), youth's prior legal charges (yes–no). The stepwise procedure identified nine conditioning variables for the logistic regression (see Table 1).

A composite outcome variable using three categories (desirable, mixed, and undesirable) was created based on available indicators of the behavioral and functional status of youth. The categories were derived from the original 1-year outcome variables that included living with family, educational progress, trouble with the law, and out-of-home placement during the follow-up period. These 1-year outcome variables are relevant for both IIHT and RC youth because they reflect domains that are targeted by treatments in both settings for emotionally and behaviorally challenging youth.

Based on the identified clinical goals of both IIHT and RC treatment, 1-year postdischarge outcomes have been classified as desirable, mixed, or undesirable. The desirable category consisted of those cases with indication for all of the following criteria: living with family, progress in school, no trouble with the law, and no out-of-home placements in the follow-up period. The mixed category consisted of those cases in which the child was living with the family at the time of follow-up but met at least one of the following additional criteria: no progress in school, trouble with the law, or an out-of-home placement during the follow-up period. The undesirable category consisted of those cases with indication of the child not living with family, regardless of the status of other outcome indicators. Because the outcome was an ordinal variable, an ordinal logistic regression to predict the probability of falling into each of the three outcome categories was conducted.

Evaluation of Bivariate Differences (Step One)

Bivariate differences between IIHT and RC youth on a wide variety of 25 demographic and psychosocial history variables were evaluated. The chi-square test for categorical variables was used. The purpose of this preliminary examination was to confirm the need for PSM by assessing the extent of selection bias.

Selection of Conditioning Variables in the Logistic Regression Model Predicting Propensity Score (Step 2)

The next step was to test the significance of these variables after controlling for their joint dependency by using logistic regression. With so many bivariate significant differences and a relatively small sample of 1,369, inclusion of all these variables in one logistic regression model was not possible. A stepwise procedure to identify the variables that may be included in the logistic regression was used. In this process, a significance level of .30 was required to allow a variable to enter into the model, and a significance level of .35 was required to stay in the model. Prior studies on propensity score analysis all indicate the importance of having a correct set of matching variables in the logistic regression predicting propensity scores (e.g., Smith & Todd, 2005). We have taken this issue seriously in this study, having used two different sets of independent variables to predict propensity scores. Results show that the predicted propensity scores are not sensitive to the different sets of independent variables. Therefore, the propensity scores we used for matching came from one of the two models.

One-to-One Nearest Neighbor Matching Within Caliper (Step 3)

A one-to-one nearest neighbor matching within caliper was then conducted to identify a group of RC youth who were most identical to IIHT youth in terms of propensity for receiving IIHT. Following conventions of the field, two calipers were specified: .10 (i.e., a narrowed or more accurate one) and one a quarter of the standard deviation of the propensity score (i.e., a wider one to help to reduce the number of participants who cannot find matches). Results of matching within these calipers were exactly the same; that is, the size of the calipers did not influence the results of the match. Therefore, the results that follow are presented on the use of matching with the narrower caliper of .10.

Outcome Analysis (Step 4)

Based on the comparison groups resulting from the matching, a final analysis on outcome differences between IIHT and RC was conducted. For each resample, two ordinal logistic regression analyses were carried out; one contained fewer covariates than the other. The purpose of running two sets of ordinal logistic regression models was to check whether the treatment difference in outcome is sensitive to different specification of the ordinal logistic regression, a threat to the validity of propensity score analysis. This approach produced parallel analyses, which include some key variables and differing additional variables to better understand the robustness of these variables when used differently. For each parallel analysis, the ordinal logistic regression model analyzing the final outcome contains exactly the same set of variables identified as conditioning variables in the prior propensity score model.

Parallel Analysis 1. The logistic regression model predicting the propensity score contains the following variables: race, gender, age, number of mental health problems, committed of a status

offense, committed a sex offense, siblings in out-of-home care, and past treatment.

Parallel Analysis 2. The logistic regression model predicting the propensity score contains the following variables: race, gender, age, presenting problem of delinquency, number of mental health problems, commission of a status offense, other criminal behavior, committed a sex offense, and receipt of special education services.

Results

Bivariate results showed that the two groups differed on many variables to a statistically significant degree (see Table 1), confirming that these youth were not randomly assigned to treatment programs. At the same time, there was considerable overlap between their characteristics, indicating that they may have been assigned, in part, based on bureaucratic selection and not on a precise matching of their unique characteristics with the characteristics of the treatment. Therefore, selection bias was an extraneous factor that needed to be controlled in the evaluation.

Table 1 shows that, after matching, there were no significant differences between the groups for 19 variables, for either Scheme 1 or 2 (38 comparisons in all). Table 1 also shows that, after matching, five bivariate differences remained statistically significant: race, age, occurrence of mental health problems (for Scheme 1 only), number of mental health problems, and commission of a status offense (a total of nine comparisons in all). In such instances when differences remain, indicating that the covariate distributions for the two groups being compared did not overlap sufficiently, subsequent analyses should adjust for these covariates (Rubin & Thomas, 1996). By using the variables that remained statistically significant as further control variables in the ordinal logistic regression, the issue of nonoverlapping distributions is sufficiently addressed by the follow-up analysis.

Table 2 shows the results of the logistic regression model using a stepwise procedure. These results also indicate that the two groups were not identical before matching; the RC group was composed of more African Americans and males, older youth, more youth with a high number of mental health problems, more youth who had committed a status offense, more youth who had previously committed a sex offense, and more youth whose siblings were in out-of-home care compared with the IIHT group. Results of the stepwise logistic regression based on the second set of entry variables showed a similar pattern as those in Table 2.

Using Schemes 1 and 2, two outcome analyses using an ordinal logistic regression model were conducted. Results of the differences on probability of falling into each of the three outcomes between IIHT and RC are presented in Table 3. Comparing results of Scheme 1 with those of Scheme 2, both produced the same pattern with respect to the effects of treatment; therefore, the results of Scheme 1 are the focus of the discussion.

Scheme 1 results show that IIHT youth are similar to RC youth in the likelihood of having a desirable outcome (.615 vis-à-vis .558), and an undesirable outcome (.202 vis-à-vis .243), and a mixed outcome (.183 vs. .199). Taken together, these data suggest that IIHT has a more positive impact on youth's outcomes than RC. However, neither of the two schemes produced a difference between the two treatment groups that was statistically significant at $p < .05$, because the p values associated with better overall outcomes for the intensive-in-home group variable were .061 for Scheme 1 and .095 for Scheme 2. The consistency of treatment effects across the two schemes indicates robustness of the result; the finding of a trend toward a positive impact of IIHT compared

Table 2
Logistic Regression Model Predicting Propensity Scores: IIHT vs. RC

Predictor	*N* (%)	% of IIHT clients	p-value of χ^2 test	Logistic Regression of the Propensity Score Model Odds Ratio (95% C.I.)
All	1369 (100%)	68.4%		
Race (other)	1006 (73.5%)	72.8%	<.0001	.521 (.400, .678)
African American	363 (26.5%)	56.5%		
Gender (female)	389 (28.4%)	79.2%	<.0001	.446 (.333, .598)
Male	980 (71.6%)	64.2%		
Age group[a]			<.0001	.669 (.560, .798)
0–11 years	368 (26.9%)	80.7%		
12–15 years	707 (51.6%)	62.5%		
16 years+	294 (21.5%)	67.4%		
Number of mental health problems[a]			<.0001	.829 (.703, .978)
0	457 (33.4%)	76.8%		
1–2	423 (30.9%)	64.5%		
3+	489 (35.7%)	64.0%		
Committed of a status offense (No)	1190 (86.9%)	70.2%	.0004	.746 (.526, 1.057)
Yes	179 (13.1%)	57.0%		
Committed a sex offense (No)	1189 (86.9%)	70.3%	.0001	.816 (.576, 1.156)
Yes	180 (13.2%)	56.1%		
Siblings in out-of-home care (No)	1336 (97.6%)	69.1%	.001	.534 (.260, 1.097)
Yes	33 (2.4%)	42.4%		

Reference group for categorical variables is shown next to the variable name.
[a] This variable was treated as an ordinal variable in the logistic regression analysis. As such, the odds ratio indicates the relative change in the odds of using IIHT when the variable increases by one unit.

Table 3
Predicted Probabilities of Having Each of the Three Outcomes by Treatment Group

Analytic Scheme and Group	Probability of Falling into Each Category of Outcome		
	Desirable	Mixed	Undesirable
Based on Scheme 1 ($p > .05$)			
IIHT	0.615	0.183	0.202
RC	0.558	0.199	0.243
Based on Scheme 2 ($p > .05$)			
IIHT	0.604	0.192	0.203
RC	0.553	0.207	0.240

with RC with a statistical trend is not sensitive to (a) the conditioning variables included in the logistic regression or (b) the covariates included in the outcome analysis.

Table 4 presents the estimated odds ratios based on the ordinal logistic regression of Scheme 1. Predictors that have a significant relationship to the three-level outcome (i.e., those with an odds ratio whose 95% confidence interval does not cover a zero) include the following: (a) siblings in out-of-home care (the likelihood of having a more desirable outcome for those who had siblings in out-of-home care is 64.4% less than those who did not); and (b) past mental health treatment (the likelihood of having a more desirable outcome for those who had past treatment is 33.8% less than those who did not).

Discussion

This study used a rigorous method to evaluate the effectiveness of IIHT vis-à-vis RC because of the inherent limitations of the quasi-experimental framework, including selection bias, warped counters, and failing factuals (Guo et al., 2006). The following specific measures were adopted to increase the rigor of the evaluation: (a) use of conditioning variables to predict propensity scores of receiving IIHT, even for youth who, in fact, received RC; (b) matching of the RC youth with the IIHT youth on propensity scores; (c) employment of multivariate analysis on the matched sample; and (d) use of various sensitivity analyses, including different sets of conditioning variables, different caliper sizes for matching, and different sets of covariates for the outcome analysis.

After adopting these measures, clients who received IIHT had a consistently higher probability of having a desirable composite outcome and a lower probability of having an undesirable composite outcome compared with clients in RC. Although these very meaningful findings are the result of a rigorous analytical approach, caution is warranted. First, the selection of outcome variables was based on availability through case record data, which limited the range of outcomes available and the sensitivity of measurements of them. Second, some preexisting differences did exist after matching and may not have been adequately controlled in the final multivariate analyses. Also, although the differences between IIHT and RC are consistent, they only reach the level of a statistical trend, not statistical significance. Thus, these findings do not provide strong support for the argument that IIHT is more effective than RC, only that it is highly unlikely that RC is better than IIHT. Randomized clinical trials would offer a stronger basis for conclusions, and we encourage their future use.

The two differences that emerged as contributors to overall outcomes bear some discussion because they may offer alternative views of the results of this study. Children with siblings in out-of-home care may have done worse because of greater family dysfunction or greater family caregiver burden or because they were endeavoring to be reunited (there is some evidence to this effect in the literature on reasons why children run away; Courtney et al., 2005). Alternatively, it is possible that these children are more likely to be assigned to group care, as shown in Table 1, and that involvement in group care is associated with worse outcomes. The finding of poorer outcomes for youth who have had prior mental health treatment may indicate the greater chronicity of their difficulties: These children were evenly matched between the groups as a result of the PSM.

Conclusions

Debate continues within the children's services community about the most appropriate services and setting for children with significant mental health care needs. Many children and youth continue to enter RC before other alternatives have been tried. James et al. (2006) indicate that among youth under child welfare services supervision who enter out-of-home care, 25% experience an intensive or restrictive setting during their first out-of-home care episode. In addition, 48.9% of the youth with such placements are placed into intensive or restrictive settings as a first placement during their very first out-of-home episode, bypassing foster care and treatment foster care. Lyons and McCulloch (2006) argue that intensive in-home approaches like MST and wrap-around services could reduce the need for RC, but that reducing the use of RC

Table 4
Estimated Odds Ratios Based on the Ordinal Logistic Regression of Scheme 1

Variable	Odds Ratio (95% C.I.)
Treatment (RC)	
IIHT	1.311 (.987, 1.740)
Race (other)	
African American	.819 (.611, 1.099)
Gender (female)	
Male	.821 (.580, 1.161)
Age group[a]	1.003 (.812, 1.238)
Number of mental health problems[a]	.949 (.786, 1.146)
Committed of a status offense (No)	
Yes	.968 (.679, 1.380)
Committed a sex offense (No)	
Yes	.923 (.634, 1.344)
Siblings in out-of-home care (No)	
Yes	.356 (.167, .756)*
Past mental health treatment (No)	
Yes	.662 (.470, .931)*

Reference group for categorical variable is shown next to the variable name.
[a] This variable was treated as an ordinal variable in the logistic regression analysis. As such, the odds ratio indicates the relative change in the odds of having a desirable outcome when the variable increases by one unit.
$^{*}p < .05$

should not precede the development of these alternatives. They argue that such a move could increase the rate of hospitalization as well as juvenile detention. This might especially be the case for children in specialty populations, such as those who need sex offender treatment, who are often placed into RC as a matter of community safety. Our data provide no evidence about what would happen in the absence of RC, but they persuasively show that the current direction of greater use of intensive in-home services is unlikely to result in worse outcomes, and that the continued use of RC might do just that.

The seemingly universal belief that children should be served in the least restrictive environment in which they can be safely treated provides an impetus to increase the availability of IIHT services. The cost of IIHT compared with residential treatment is another factor that enhances the attractiveness of this approach. At the agency that provided data for this study, the cost of providing IIHT is between 22.6% and 24.6% that for RC (G. Gregory, personal communication, April 14, 2007). The cost of IIHT represents provision of the in-home service by bachelor's- and master's-level therapists with caseloads of four to six families, with a robust training, supervision, and consultation structure (four therapists per supervisor; weekly team supervision and consultations; intensive initial and quarterly training coupled with weekly therapist development plans). Although IIHT services are not inexpensive (often between $80 and $120/day), they are considerably less expensive than the residential treatment centers, juvenile justice facilities, or psychiatric hospitals that currently serve youth who could benefit from these services.

IIHT is less restrictive and less expensive than most other treatment options that are widely available. This study demonstrates that there is a reasonable likelihood that IIHT is more successful in achieving positive long-term outcomes for youth. Such evidence may be of interest to policymakers and to purchasers of mental health services for children. There has been substantial acquisition activity among for-profit corporations providing residential treatment services. In addition, these companies are increasingly marketing their services to child welfare systems (P. Lawler, personal communication, April 12, 2007). Officials charged with providing for the mental health needs of children in the system may feel pressure from a variety of sources to use more restrictive settings for youth, partly because of the increasing availability of such beds and the lack of access to intensive in-home services. As evidence mounts concerning less expensive alternatives to residential treatments that are also at least as effective, policymakers may consider shifting their priorities to emphasize treatment in the least restrictive, most effective setting for each child. Such a move could provide attractive options for stakeholders who are looking to improve services for the children and families in their care.

The findings are not inconsistent with reviews of the general literature indicating that placement in group care often results in worse outcomes, mitigated only when those programs have a family focus (e.g., Barth, 2005). Yet these findings are not fully supportive of this conclusion either. One explanation is that the model used for RC is a family- and community-focused model and is one of the few RC programs that does have evidence of providing benefit (Hooper et al., 2000). More generally, family-focused RC programs obtain the strongest results. This would tend to mitigate differences between the outcomes of the two family-focused approaches compared here.

References

Allison, P. D. (2001). *Missing data (quantitative applications in the social sciences).* Thousand Oaks, CA: Sage

Baker, A. J. L., Wulczyn, F., & Dale, N. (2005). Covariates of length of stay in residential treatment. *Child Welfare, 84*(3), 363–386.

Barth, R. P. (2005). Residential care: From here to eternity. *International Journal of Social Welfare, 14,* 158–162.

Barth, R. P., & Chintapalli, L. K. (in press). Impermanence for children and youth in out-of-home care. In B. Kerman & A. N. Maluccio & M. Freundlich. (Eds.). *Achieving permanence for older children and youth in foster care*. New York: Columbia University Press.

Burns, B. J., & Hoagwood, K. (2002). *Community treatment for youth: Evidence-based interventions for severe emotional and behavior disorders.* New York: Oxford University Press.

Chamberlain, P., & Reid, J. B. (1998). Comparison of two community alternatives to incarceration for chronic juvenile offenders. *Journal of Consulting and Clinical Psychology, 66*(4), 624–633.

Courtney, M., Skyles, A., Miranda, G., Zinn, A., Howard, E., & Goerge, R. (2005). *Youth who run away from substitute care.* Chicago: Chapin Hall Center for Children at the University of Chicago.

Dishion, T. J., & Dodge, K. A. (2005). Peer contagion in interventions for children and adolescents: Moving towards an understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology, 33*(3), 395–400.

Eddy, J. M., Whaley, R. B., & Chamberlain, P. (2004). The prevention of violent behavior by chronic and serious male juvenile offenders: A 2-year follow-up of a randomized clinical trial. *Journal of Emotional and Behavioral Disorders, 12*(1), 2–8.

Friedman, R. M., Pinto, A., Behar, L., Bush, N., Chirolla, A., Epstein, M., et al. (2006). Unlicensed residential programs: The next challenge in protecting youth. *American Journal of Othopsychiatry, 76* (3), 295–303.

Gorske, T. T., Srebalus, D. J., & Walls, R. T. (2003). Adolescents in residential centres: Characteristics and treatment outcome. *Children and Youth Services Review, 25,* 317–326.

Guo, S., Barth, R., & Gibbons, C. (2006). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children and Youth Services Review, 28,* 357–383.

Hair, H. J. (2005). Outcomes for children and adolescents after residential treatment: A review of research from 1993 to 2003. *Journal of Child and Family Studies, 14*(4), 551–575.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equations system. *Econometrica, 46,* 931–960.

Heckman, J. J. (1979). Selection bias as a specification error. *Econometrica, 47,* 153–161.

Heckman, J. J., Ichimura, H., & Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies, 64,* 605–654.

Henggeler, S. W. (2006). Methodological critique and meta-analysis as Trojan horse. *Children & Youth Services Review, 28,* 447–457.

Henggeler, S. W., Rowland, M. D., Halliday-Boykins, C., Sheidow, A. J., Ward, D. M., Randall, J., et al. (2003). One-year follow-up of multisystemic therapy as an alternative to the hospitalization of youths in psychiatric crisis. *Journal of the American Academy of Child & Adolescent Psychiatry, 42,* 543–551.

Hoagwood, K., Burns, B. J., Kiser, L., Ringeisen, H., & Schoenwald, S. K. (2001). Evidence-based practice in child and adolescent mental health services, *Psychiatric Services 52*, 1179–1189.

Hobbs, N. (1982). *The troubled and troubling child.* San Francisco: Jossey-Bass.

Hooper, S., Murphy, J., Devaney, A., & Hultman, T. (2000). Ecological

outcomes of adolescents in a psychoeducational residential treatment facility. *American Journal of Orthopsychiatry, 70*(4), 491–500.

James, S., Leslie, L. K., Hurlburt, M. S., Slymen, D. J., Landsverk, J., Davis, I., et al. (2006). Children in out-of-home care: Entry into intensive or restrictive mental health and residential care placements. *Journal of Emotional and Behavioral Disorders, 14*(4), 196–208.

Leichtman, M. (2006). Residential treatment of children and adolescents: Past, present, and future. *American Journal of Orthopsychiatry, 76*(3), 285–294.

Leve, L. D., & Chamberlain, P. (2005). Association with delinquent peers: Intervention effects for youth in the juvenile justice system. *Journal of Abnormal Child Psychology, 33*(3), 339–347.

Littell, J. H. (2006). The case for multisystemic therapy: Evidence or orthodoxy? *Children & Youth Services Review, 28,* 458–472.

Lyons, J. S., & McCulloch, J. R. (2006). Monitoring and managing outcomes in residential treatment: Practice-based evidence in search of evidence-based practice. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(2), 247–251.

Peterson, M., & Scanlan, M. (2002). Diagnosis and placement variables affecting the outcome of adolescents with behavioral disorders. *Residential Treatment for Children and Youth, 20,* 15–23.

Pumariega, A. J. (2006). Residential treatment for youth: Introduction and a cautionary tale. *American Journal of Orthopsychiatry, 76*(3), 281–284.

Quay, H. C. (1979). Residential treatment. In H. C. Quay & J. S. Werry (Eds.), *Psychological disorders of childhood* (2nd ed., pp. 387–410). New York: John Wiley & Sons.

Quay, H. C. (1986). Residential treatment. In H. C. Quay & J. S. Werry (Eds.), *Psychological disorders of childhood* (2nd ed., pp. 558–580). New York: John Wiley & Sons.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics, 52,* 249–264.

Smith, J. A., & Todd, P. E. (2005). Does matching address Lalonde's critique of nonexperimental estimators? *Journal of Econometrics, 125,* 305–353.

Stage, S. A. (1999). Predicting adolescents' discharge status following residential treatment. *Residential Treatment for Children and Youth, 16,* 37–56.

Wall, A. E., & Barth, R. P. (2005). Aggressive and delinquent behavior of maltreated adolescents: Risk factors and gender differences. *Stress, Trauma, and Crisis: An International Journal, 8,* 1–24.

Whittaker, J., & Pecora, P. (1984). A research agenda for residential care. In T. Philpot (Ed.), *Group care practice*: The challenge of the next decade (pp. 71-86). Surrey, England: Business Press International.

Wilson, D. R., & Lyman, R. D., (1983). Residential treatment of emotionally disturbed children. In C. E. Walker & M. C. Roberts (Eds.), Handbook of child clinical psychology (pp. 1096-1088). New York: John Wiley & Sons.

Webster, D. (1999). *California's group care population: A longitudinal analysis of placement dynamics and outcomes.* Berkeley: University of California at Berkeley.

Wolins, M., & Piliavin, I. (1964). *Institution and foster family: A century of debate.* New York: Child Welfare League of America.

Wulczyn, F., Barth, R. P., Yuan, Y. Y., Jones Harden, B., & Landsverk, J. (2005). *Beyond common sense: Evidence for child welfare policy reform.* New York: Transaction De Gruyter.

Wulczyn, F. (in press). Foster youth in context. In B. Kerman, A. N. Maluccio, & M. Freundlich, (Eds.). *Achieving permanence for older children and youth in foster care*. New York: Columbia University Press.

Received May 7, 2007
Revision received September 21, 2007
Accepted September 21, 2007 ■