

## The Teaching-Family Model and Post-Treatment Recidivism: A Critical Review of the Conventional Wisdom

David E. Kingsley

### Abstract

Conventional wisdom suggests that the Teaching-Family Model (TFM) approach to treating youthful offenders is not effective in reducing post-treatment recidivism. This article reviews two major studies referenced in support of this widespread perception. Data presented in one widely referenced study are treated with a Cochran-Mantel-Haensel test, which, the author argues, is appropriate for data originally presented in two separate 2 X 2 tables (one for boys and one for girls). The construct and statistical conclusion validity of a major evaluation study presented to the NIMH is critically evaluated and discussed. A revised view of the leading TFM evaluations has implications for public policy regarding juvenile justice. The author suggests that a belief in the lack of post-treatment efficacy associated with community-based residential treatment has resulted in harsher treatment of juveniles and a higher incarceration rate.

Keywords: Teaching Family Model (TFM), Treatment, Juveniles, Recidivism

---

Can youthful offenders be rehabilitated? In the United States during the past thirty years, this question has engendered ongoing debate and disagreement (Glaser, 1980; Lipton, Martinson & Wilks, 1975; Martinson, 1974; Palmer, 2002; Wilson & Herrnstein, 1985). However, a few decades ago, there was considerable optimism regarding the efficacy of treatment for juvenile delinquents. The Teaching-Family Model, in fact, grew out of a 1960s zeitgeist of all things are possible when it comes to reforming society (Wolf, Braukmann, & Ramp, 1987).

The theoretical underpinnings of the Teaching-Family Model (TFM) have been described as radical behaviorism (Morris & Braukmann, 1987). Delinquency, according to the theory, is the result of behavior deficiency rather than psychopathology (Phillips, Phillips, Fixsen, & Wolf, 1973). As applied to the treatment of adjudicated youth, the radical behaviorist approach is characterized by a “token economy system of reinforcement” (Phillips, et al., 1973, page 75). Youth in treatment receive points for compliance and achievement that can be exchanged for privileges. However, as the program developed in the early years, it became obvious that the teaching, social-interaction aspects of the treatment became “the heart of the program” (Phillips, et al., 1973, page 75).

The process of “give-and-take-instruction, demonstration, practice, feedback,” (Phillips, 1973, page 75) is designed to help youth overcome behavior deficiencies and learn prosocial behaviors. Hence, the model is characterized by a small number of youths (eight) in a community-based residential setting managed by a married couple trained in the prescribed techniques (Phillips, Phillips, Fixsen, & Wolf, 1974). The Teaching Family Association developed as an accrediting agency. In general, fidelity to the model necessitates a highly structured program with specific protocols and continuous measures of each youth’s behavior.

Early evaluations of the model by researchers responsible for its development at the University of Kansas suggested phenomenally better results of the TFM compared to institutionalization and probation (Phillips, et al., 1973; Kirigin, Wolf, Braukmann, Fixsen, & Phillips, 1979). However, since the early 1980s, it has been widely perceived as a model that lacks efficacy in reducing recidivism. (Fonagy, Target, Cottrell, Phillips, & Kurtz, 2002; Jones, Weinrott, & Howard, 1981; Kirigin, Braukmann,

Atwater, & Wolf, 1982; Morris & Braukmann, 1987; Quay, 1986; U. S. Department of Health and Human Services, 1999; Wilson, 1983; Wilson & Herrnstein, 1985).

According to Jenkins (2006), the backlash to the 60s spirit started around 1974. Indeed, in criminology the “nothing works movement” led by Martinson and his colleagues appeared on the scene with a publication in the *The Public Interest* (Martinson, 1974).

In spite of evaluations showing strong positive effects of the TFM compared to “no treatment” and “institutional” comparison programs reported in scholarly publications during the 1970s (Phillips, et al., 1973; Kirigin, et al., 1979), influential social scientists James Q. Wilson and Richard Herrnstein in their acclaimed book, *Crime & Human Nature* (1985) indicated that the Teaching Family Model served as evidence of the futility of rehabilitative efforts.<sup>3</sup>

It is ironic and unfortunate that Wilson and Herrnstein based their assessment of the TFM on two reported studies (Wilson & Herrnstein, 1985; Wilson, 1983), one of which was reported by Kirigin, et al. (1982) in the *Journal of Applied Behavioral Analysis*. The other study was an evaluation project pertaining to the TFM reported to the NIMH in 1981 (Jones, et al., 1981). Ironically, Kirigin and her colleagues were members of the core team at the University of Kansas responsible for developing and disseminating the model. The unfortunate aspect of the Kirigin, et al. article and Jones, et al. report (as submitted to the NIMH) is that administrators and scholars accepted them at face value.

Over the past two decades, the conclusions of these studies have been viewed as definitive answers to questions about TFM efficacy pertaining to the treatment of juvenile delinquents. (Fonagy, Target, Cottrell, Phillips, & Kurtz, 2002; Jones, Weinrott, & Howard, 1981; Kirigin, Braukmann, Atwater, & Wolf, 1982; Morris & Braukmann, 1987; Quay, 1986; U. S. Department of Health and Human Services, 1999); Wilson, 1983; Wilson & Herrnstein, 1985). Even a cursory analysis of validity issues in these studies should have given pause to statistically and methodologically sophisticated social scientists referencing them in support of a viewpoint.

This article will attempt to make a case for the importance of revisiting research responsible for the “nothing works” viewpoint in general and the conventional wisdom concerning TFM post-treatment effectiveness in particular. The belief that TFM is effective while youths are in treatment but is no more effective than “treatment as usual” after they leave is in fact the conventional wisdom.

Without doubt, this viewpoint has influenced the 1990s emphasis on a more punitive approach to juvenile delinquency. Predictions of a coming wave of super-predators (Dilulio, Walters, & Bennett, 1996; Wilson & Petersilia, 199) and a widespread belief that treatment for troubled youth is not effective were coincidental with increasingly harsh juvenile justice systems in practically all states (Zimring, 2005).

It will be demonstrated in the following pages that the current state of affairs concerning perceptions of treatment of adjudicated youth is based on faulty analyses and a host of fallacies and methodological errors. Primarily, but not exclusively, the remainder of this article will focus on statistical conclusion validity, and construct validity. However, as in any quantitative research, it is not difficult to uncover a

---

<sup>3</sup> Since the 1970s, Professor Wilson has been one of the most influential criminologists in the United States. As a professor at Harvard and president of the American Political Science Association, along with his role as trustee for powerful policy entities such as the American Enterprise Institute and Manhattan Institute, he has had considerable influence on criminal justice policies at the National level since the Reagan Administration. If anyone should doubt the respect accorded to Professor Wilson, they need only consider his status as a recipient of the Medal of Freedom awarded by President George W. Bush.

tangled web of issues pertaining to design sensitivity in which meaningful effects of a treatment are often overlooked.

Hence, sample size, effect size, measurement error, heterogeneity of subjects, experimental error, and statistical analyses, all of which are factors in the capacity of study to find meaningful effects when they are present, (Lipsey, 1990) have been to some degree or other integral in a misperception concerning the TFM.

### **Jones, Weinrott & Howard Evaluation**

In 1981, Jones, Weinrott, and Howard reported the results of a national evaluation of the Teaching-Family Model to the National Institute of Mental Health. According to Jones, et al., their evaluation, funded by the NIMH, consumed six years – 1975 to 1981. Generally, the authors concluded that the Teaching-Family Programs impacted treated youths, “...at least as well as the state-of-the-art community-based comparison programs, were operating less expensively overall and most cost effectively in the school domain, and evaluated more highly by community consumers” (Jones, et al., page 2). These positive findings aside, the evaluators concluded that the, “...the chronic problem of delinquency continues to evade the efforts of even the better developed programs like the Teaching-Family Model” (Jones, et al., page 2).

Data from the Jones and colleagues (1981), study has been unavailable to this researcher for reanalysis.<sup>4</sup> Nevertheless, the study design is quite problematic and raises doubt about conclusions reported to the NIMH. As will be demonstrated in the next few paragraphs, construct validity of the independent variable, i.e., treatment program (with two levels – TFM and non-TFM) is questionable.

A fair and just evaluation of a treatment model that has achieved nation-wide dissemination must, it seems, include fidelity to the specifications established by its developers, as well as attention to the theoretical framework of treatment techniques (Glaser, 1980). The Teaching-Family Model is based on a set of clearly stated criteria: (1) A married couple with training and certification by the Teaching Family Association, (2) No more than 8 youths in an accredited home, (3) a system of self-governance by the clients, (4) a behavior modification system. The qualifications of staff are established through certification and training.

In a government funded, independent evaluation of a widespread program with comparison of the target model to “treatment as usual” and/or “no treatment” groups, an experimental or (quasi-experimental design), program type would constitute an independent variable with a specific number of levels. For instance, in the Jones and colleagues (1980) evaluation, the independent variable consisted of two levels: (1) TFM and (2) any other group homes available in the area of TFM homes included in the study.

In evaluating youth treatment, construct validity and fidelity to a prescribed model is basically the same thing. If the independent variable is not what it is defined as being, the intended construct is not actually the focus of measurement.

Although Jones, et al. stated that the evaluation was “...designed to compare 26 TFM home and 25 comparison homes ...,” programs were not selected because they met particular criteria (in accordance

---

<sup>4</sup> In an effort to obtain the raw data from the evaluation project, this researcher contacted R. A. Jones, the principle investigator on the project. Dr. Jones indicated that he no longer had the data in his possession. It is possible that the data could be located within the archives of the NIMH. Efforts in that regard will be continued.

with a construct). Rather, teaching-parents were self-selected at three training sites across the United States (Jones, et al., page 40).

The report submitted by Jones, et al. to the NIH clearly indicates that many homes considered TFM in their evaluation did not fit within the prescribed framework:

“The ranges of youth per program in the two samples were 3 to 22 for TFM programs and 5 to 22 for comparison programs. Median numbers of youth per program were 13 (TFM) and 15 (comparison).” (Jones, et al., page 41)

It is troubling to this researcher that at least half of the TFM homes were considerably larger than the criteria for the model. The size of program (number of youth), length of operation (stability), and qualifications of staff most likely impacted within home variance. The evaluation project commenced in 1975 when the TFM was just hitting its stride. According to the evaluators, “No programs were added to the sample during the evaluation study but two were dropped when they ceased to operate” (Jones, et al., page 41). This researcher was somewhat stunned by the evaluators’ admission that data obtained prior to the closure of the programs were, “...retained for analysis, and their youth were continued in the follow-up phase of the study” (Jones, et al., page 41).

Although the TFM was still in an incipient stage of development and dissemination in 1975, the study, in its entirety, focused on impact. The consequences have been summative with scant attention to formative factors.

Selection of homes larger than the model specified and inclusion of a large cadre of non-certified teaching-parents, along with construct validity, should have been considered by those later referencing the study. Furthermore, statistical conclusion validity should have been a concern. Analyses, as reported by Jones, et al., indicated that individual youths were entered as units of analysis without regard for within home variance.

Rather than treating “home” as a random effect, the authors of the study aggregated youth across all 26 TFM homes and 25 comparison homes. The report does not provide a list of homes with home-by-home characteristics such as the qualifications of staff, number of residents, and so forth. Other than two levels of the independent variable, i.e., TFM and non-TFM, no control was exercised for a variety of critical home factors.

Larger homes may have been less effective than homes with the prescribed number of youths. If this were indeed the case, the poor functioning programs would have been weighted more heavily in the analysis. This would hardly be fair to the Teaching Family Model.

Kirigin and colleagues (1982) discussed in depth later in this article), criticized the selection of homes in the Jones and colleagues evaluation. It was pointed out by Kirigin and her colleagues that of the three training sites from which teaching-parents were recruited, two were “... when the study began.” (page 13). According to Kirigin, et al.(1982):

“... one of the sites was never implemented adequately due primarily to insufficient staff. For example, for a significant portion of the study period, no one trained in the model was supervising the site.” (page, 13)

All authors and researchers involved with the Kirigin, et al. 1982 article were associated with the University of Kansas department responsible for developing the TFM. It is apparent from the following

statement from the article that the University of Kansas researchers had taken issue with Jones and his colleagues:

“In the final report, Jones and his colleagues did not present the data analyzed by training site. However, earlier in their research efforts (at a time when approximately 80% of the subjects were in the study), Jones provided us with court-record offense data that were analyzable by site. The court data indicated as of that time, the homes from the Kansas site had during treatment levels of criminal offenses that were about half the levels of their comparison programs. (The pretreatment levels of offenses were comparable for these groups.) These during treatment data are consistent with the findings we have reported here and with those in our more recent self-report data on Kansas homes (page 14).

The Kansas researchers continued their criticism of the Jones, et al. study by reflecting on formative issues in initial attempts to replicate Achievement Place, the original Teaching-Family program. They stated “This failure to find that Teaching-Family programs were better (at least on court measures) than comparisons at these first two replication sites is reminiscent of initial difficulties in replicating the original Achievement Place group home program when we first began working with other group homes in Kansas” (page 14). In spite of their differences with Jones, et al., these researchers associated with the TFM also concluded from their analyses that the TFM homes they evaluated performed no better than group-home treatment as usual.

### **Kirgin, Braukmann, Atwater & Wolf, 1982**

Kirgin, et al. in the 1982 article appearing in the *Journal of Applied Behavior Analysis* concluded that when youth in TFM programs were compared to youth in non-TFM residential programs, a significant “during-treatment” difference was present between the two groups. Nevertheless, the post-treatment differences were not, according the authors, significant for either boys or girls.

It is the opinion of this researcher that the conclusions of the authors were not supported by the data presented in the article. It would appear that weak statistical power and the validity of the analyses with which the data were treated rendered the findings of “no effect” questionable. Reanalysis of the data and statistical power analysis tends to suggest that in comparison to the non-TFM programs, TFM post-treatment effects were likely.

In the following discussion, the data reported by the authors will be presented, followed by an examination of the original analyses. The data and analyses will then be subjected to a power analysis. Finally, results of an analysis of the data with the Cochran-Mantel-Haensel technique will be presented.

### **The Data As Originally Presented:**

The data presented in Figures 1 and 2 is a duplication of the format in which Kirigin et al. presented the data. Based on that presentation, this researcher created 2 X 2 tables for both boys and girls (Tables 1 and 2). A discussion of the data follows Table 2.

#### **BOYS**

**Teaching-Family (n = 102)**

**Non-Teaching-Family (n = 22)**

Effects of Group Home Treatment on Percent of Youths Involved in Offenses

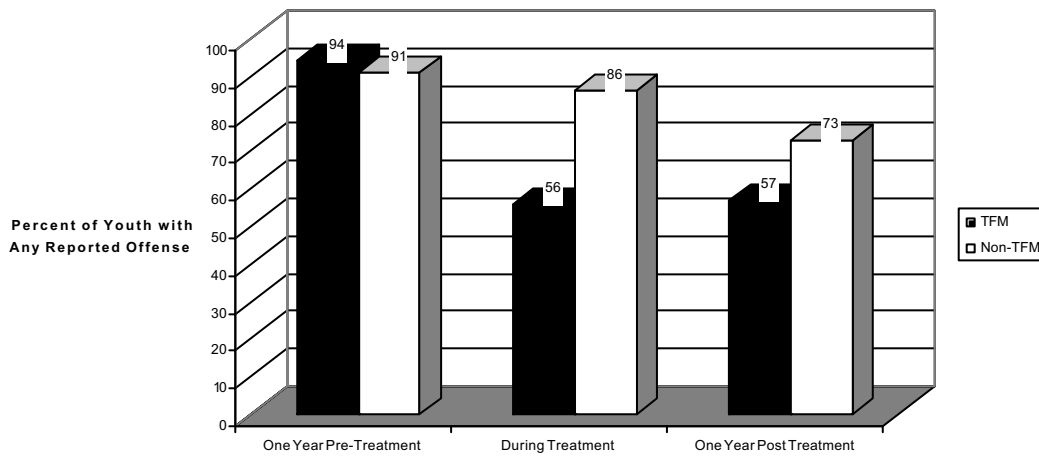


Figure 1

Table 1

Post-Treatment: Boys*			
	Teaching-Family	Non-Teaching-Family	
Involved in Offense	58	16	74
Not Involved in Offense	44	6	50
	102	22	124

\*Of importance to analysis of the data in Table 1:  $\chi^2 = 2.06$

$p = .15$

GIRLS

Teaching-Family (n = 38)

Non-Teaching-Family (n = 30)

Effects of Group Home Treatment on Percent of Youths Involved in Offenses

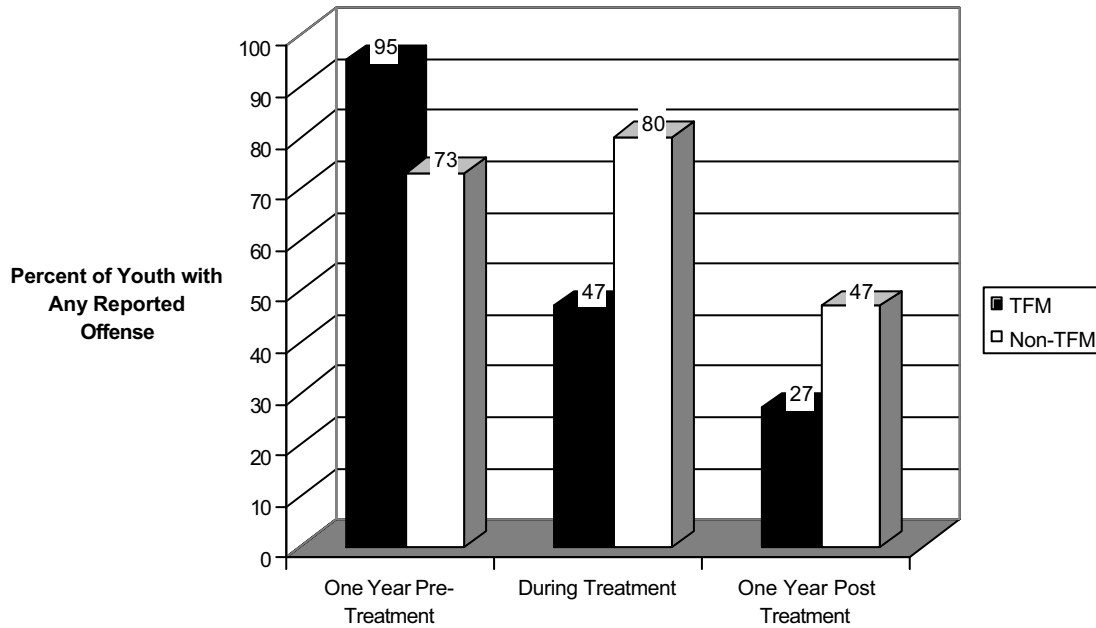


Figure 2

Table 2

Post-Treatment: Girls*			
	Teaching-Family	Non-Teaching-Family	
Involved in Offense	10	14	24
Not Involved in Offense	28	16	44
	38	30	68

\*Of importance to analysis of the data in Table 2:  $\chi^2 = 3.02$

$p = .08$

There are several interesting and important features of the data presented in Figure 1 and Tables 1 and 2:

The sample sizes for both groups are small – the female sample is particularly small.

Given the odds ratios (discussed below), sample sizes, and  $\chi^2$  values, the sensitivity of the research design has likely failed to detect an effect of the TFM treatment.

Given the factors in 2, combining boys and girls into a single analytic technique would reduce the risk of Type II error.

Each of these issues will be discussed below with alternative findings from a Cochran-Mantel-Haensel<sup>1</sup> test.

### Sample Size and $\chi^2$

Due to the influence of sample size, an omnibus  $\chi^2$  statistic is problematic. The  $\chi^2$  value is sensitive to increases or decreases in the cell counts. As Agresti (1996, page 33) states:

“Chi-squared tests of independence, like any significance tests, have serious limitations. They simply indicate the degree of evidence for an association. They are rarely adequate for answering all questions we have about a data set. Rather than relying solely on results of these tests, one should study the nature of the association. It is sensible to decompose chi-squared into components, study residuals, and estimate parameters such as odds ratios that describe the strength of association.”

The odds ratio is a rather good indicator of effect size. Based on the data in Table 1, the odds ratios for boys (using cross products of the cells or  $m_{11} * m_{22} / m_{21} * m_{12}$ ) is:

$$\frac{58 * 6}{44 * 16} = .49$$

Hence the odds are .49 to 1 that TFM boys would be recidivists versus the comparison group. The inverse is 1/.49 or slightly more than a 2 to 1 greater likelihood that comparison group boys would be recidivists versus the TFM boys. These odds ratios suggest a fairly strong effect size.

The odds ratio for TFM girls versus the comparison group is similar to the OR for boys:

$$\frac{10 * 16}{28 * 14} = .41$$

The odds ratio for girls indicates that TFM girls were approximately .4 to 1 as likely as comparison group girls to be involved in an offense post treatment. Conversely, comparison girls were 2.4 times as likely to be involved in an offense post treatment. All else being equal, these odds ratios suggest that TFM treatment had a more positive effect on youth in the study than the comparison programs.

### Statistical Power

Given the sample size, odds ratios and proportions pertaining to post treatment offending, one would have to be wary of conclusions that treatment had no effect – especially when the alpha level was set at .05. The effect size for proportions can be determined by:

$ES_p = \phi_t - \phi_c$  where  $\phi_t$  and  $\phi_c$  refer to the arcsine transformation of the treatment group proportion and the comparison group proportion respectively (Lipsey, 1990, page 90).

The arcsine transformation of proportions is conducted as follows:

$\phi_t = 2\arcsin(\sqrt{p_t})$  where  $\sqrt{p_t}$  is the square root of the treatment proportion (offending post treatment).

$\phi_c = 2\arcsin(\sqrt{p_c})$  where  $\sqrt{p_c}$  is the square root of the comparison proportion (offending post treatment).

Of the TFM boys included in the study, a proportion of .57 (or 57%) had offended after treatment while a comparison group proportion of .73 had offended after treatment. Based on the arcsine transformation,  $ES_p$  is  $1.711 - 2.049 = .34$ . Based on power charts presented in Lipsey (1998, page 91), a .05 alpha level set for a sample size of 120 and an effect size of .34 would yield statistical power of .76. An alpha of .10 would have increased power to .86 while an alpha of .15 would have resulted in power of .90.

Cohen (1988) has suggested that  $1 - \beta$  of .80 meets minimal standards. However, .90 would be desirable (and fair to treatments that are subjected to testing).

The design sensitivity of the comparison of TFM and comparison groups for females was even more problematic than was that for the boys. The  $ES_p$  for the difference in proportions for the female subjects is .42. An effect size of .40 calculated with a sample of 70 subjects – if tested at a .05 alpha level – would yield statistical power of .65. The probability of Type II error, in this case, is unsuitable for a determination that there was no post treatment effect.

### Cochran-Mantel-Haensel Test

The statistical validity of the Kirigin, et al study would have been less questionable with some enhancement of statistical power. This could have been achieved through an increase in alpha or through a larger sample size. However, a more sensitive statistical test could have been employed. Given the 2 X 2 tables for boys and girls (with a chi square test utilized to examine independence separately for each table), a Cochran-Mantel-Haensel test would have been appropriate for combining the two studies. Data in two separate tables to which two separate chi-square tests are applied is a process fraught with problems.

The Cochran-Mantel-Haensel test is suitable for 2 X 2 X K tables with a null hypothesis that X and Y are conditionally independent, controlling for Z (Agresti, 1992). The null hypothesis that the conditional odds ratio  $\theta$  between X and Y equals 1 in each table. The C-M-H is represented by:

$\frac{[\sum_k (n_{11k} - \mu_{11k})]^2}{\sum_k Var(n_{11k})}$  where  $\mu_{11k}$  represents the mean of cell  $n_{11k}$  and  $Var(n_{11k})$  represents the variance for cell  $n_{11k}$

The mean and variance of cell  $n_{11}$  along with marginal (row and column) totals in each 2 X 2 table constitute sufficient statistics for calculation of the C-M-H. The mean and variance of  $n_{11k}$  are:

$$\mu_{11k} = E_{n_{11k}} = \frac{n_{1+k}n_{+1k}}{n_{++k}}$$

where  $\mu_{11k}$  and  $E_{n_{11k}}$  represent the mean or the expected cell count for  $n_{11k}$  and  $n_{1+k}$  is the marginal cell count for column one while  $n_{+1k}$  is the marginal cell count for row one.  $n_{++k}$  is the total across all four cells (the grand total).

$$Var(n_{11k}) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{2k}}{n_{++k}^2(n_{++k} - 1)}$$

**Table 3** displays the data presented in **Tables 1** and **2** plus odds ratios for both genders and mean and variances relevant to calculation of C-M-H.

Table 3

		Recidivism				
Gender	Group	Yes	No	Odds Ratio	$\mu_{11k}$	$Var(n_{11k})$
Male	TFM	58	44	.49	60.9	4.4
	Comp	16	6			
Female	TFM	10	28	.41	13.4	3.9
	Comp	14	16			

The C-M-H statistic for the data in **Table 3** is:

$$\frac{[(58 - 61) + (10 - 13.4)]^2}{4.4 + 3.89} = \frac{40.96}{8.29} = 4.9$$

Which has a large sample chi-square distribution with  $df = 1$ . The critical  $\chi^2$  is 3.841. Hence, the null hypothesis that there is no significant difference between the TFM model and comparison programs can be rejected. The p value for a C-M-H statistic of 4.9 is .027, which is considerable smaller than .05.

The appropriateness of the C-M-H might be questioned as just one means of “fishing” for a statistic that would yield a significant p value and justification for rejection of the null hypothesis. However, the original study was designed with control for gender as a major feature. Otherwise subjects, whether male or female, would have been combined into one crosstab without regard for gender.

Critical analysis of the Jones, et al. and the Kirigin, et al. evaluations indicates the validity problems in claims of “no post-treatment” effect of the TFM in comparison to the usual group home. Nevertheless, the perception of correct confirmation of the null hypothesis is widespread, i.e. there is no post-treatment effect difference between TFM and comparison group homes. The foregoing analysis strongly suggests that the TFM enterprise fell victim to Type II error in its early stages of dissemination.

The Surgeon General’s 1999 report (U.S. Department of Health & Human Services, 1999) covered children with emotional disturbances in a chapter entitled “Children & Mental Health.” The report recognized two major therapeutic group home models: (1) the teaching family model developed at the University of Kansas and then moved to Boys Town in Omaha, Nebraska, and (2) the Charley Model developed at the Menninger Clinic.

In referencing the Kirigin, et al. article, the Surgeon General’s report concluded, “Existing research suggests that therapeutic group home programs produce positive gains in adolescents while they are in the home, but the limited research available reveals that these changes are seldom maintained after discharge” (U.S. Department of Health & Human Services, 1999, page 177). It is unfortunate that researchers responsible for that particular conclusion failed to critically evaluate analyses in studies on which they relied.

### Summary and Implications

Critical analysis of past studies pertaining to post-treatment effectiveness, as presented in this article, illustrates the importance of reviewing reported evaluations that have been highly visible and influential in the public policy arena. When the measure of efficacy is recidivism, the Teaching-Family model, without scientific justification, came to be characterized as a program that adds nothing to “treatment as usual”. Review of relevant research sheds doubt on this characterization.

All programs generated from basic research and disseminated with considerable support and funding from major U.S. administrative entities such as the NIMH, should be adequately evaluated. Program developers, the tax-paying public, and clients needing treatment deserve nothing less. One must wonder how many youth detained and confined in prison-like detention centers would have benefited from treatment in a well-operated Teaching-Family home. Lipsey’s meta-analysis (1992) suggests that at least 300 of every 1000 adjudicated youthful offenders would have been less likely to reoffend in a TFM group home than in other group homes to which the model has been compared.

Along with Lipsey’s meta-analysis, other evaluation research reported since the devastating 1980s evaluation reports has suggested meaningful post-treatment recidivism reduction effects of the Teaching-Family model (Friman, et al., 1992; Larzelere, et al., 2001; Larzelere, et al., 2004; Thompson, et al., 1996; Youngbauer, 1997). It is true that all of these researchers either are or have been associated with programs providing Teaching-Family model treatment.

Nevertheless, this research deserves as much attention, respect, and critical analysis as the Kirigin, et al. (1982) article and the Jones, et al. (1981) evaluation. It was, after all, the Kirigin, et al. (1982) article that was taken as a piece of strong evidence that the Teaching-Family Model lacked post-treatment

effectiveness. Reanalysis of data presented in the 1982 Kirigin, et al. article demonstrates how researchers can understandably make a mistake.

That article, along with other published works by the University of Kansas researchers (Morris & Braukmann, 1987; Wolf, et al., 1987) speaks to the integrity of the model's developers. They reported the results, as they believed them to be, even when they indicated a lack of post-treatment effectiveness.

The real problem here is not that researchers reached conclusions that could be questioned. The history of Teaching-Family-Model-related evaluations illustrates the way journal articles and evaluations reports can be uncritically and superficially referenced in acclaimed books, newsstand issues of major magazines, and even peer reviewed journals. That is the problem

All researchers/evaluators, including Kirigin, et al., Jones, et al., and most certainly this researcher, have reported research/evaluation with flaws and errors. Scientists make mistakes. That is the reason a scientific process should be characterized by doubt and collegial critique. Instead, social scientists along with the Surgeon General have taken early evaluations as summative.

The bigger question becomes: "Would the incarceration of two million Americans have been necessary if sufficient rehabilitation programs had been available?" If indeed treatment interventions with youthful offenders reduce recidivism and cause delinquents to veer from a trajectory toward adult prisons, then the value of rehabilitation will have been established. The necessity of incarcerating youth and adults may have been diminished with sufficient emphasis and resources directed toward community-based residential treatment along the lines of proven programs such as the Teaching-Family Model.

Unfortunately, influential academicians and bureaucrats treated initial evaluations as summative. More weight was accorded to TFM evaluations than they merited. All of the factors in design sensitivity so eloquently explained by Lipsey (1990) were generally ignored. Looking back over these studies, one finds the critical issues related to the sensitivity of a research design to detect a meaningful effect: effect size, sample size, subject heterogeneity, measurement error, experimental error, and statistical technique.

These oversights are not uncommon in the social sciences. Indeed, other than occasional references to Campbell & Cook (1979) and Campbell & Stanley (1963), attendance to statistical conclusion validity and other forms of validity problems are conspicuous by their absence. In addition to the TFM, efforts by the California Youth Authority and other programs have been victimized by early summative evaluations (Palmer, 2002).

However, the TFM has been the focus of this article, and the broader view that "nothing works" in the realm of rehabilitation of offenders is beyond its scope. Nevertheless, because prison populations continue to grow, this is a propitious time for reviewing evaluations across the entire spectrum of offender treatment.

Studies conducted by the Office of Juvenile Justice & Delinquency Prevention (OJJDP) indicate that conditions deleterious to the mental health of youth are widespread in detention centers (Parent, et al., 1994). Furthermore, minority youth have been disproportionately impacted by the trend for higher security institutionalization occurring in the past decade (Hsia, Bridges, & McHale, 2004). Shock incarceration and boot camps have proven ineffective in reducing recidivism (Lipsey, 1995, 1997, 1999; MacKenzie, & Hebert 1996; MacKenzie, Wilson, & Kider, 2001). Given the current nature of the juvenile justice system, a renewed consideration of community-based residential treatment is timely.

## References

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons.
- Bausell, R. B. & Li, Y. (2002). *Power Analysis for Experimental Research*. New York, NY: Cambridge University Press.
- Bennett, W. J., Dilulio, J. J., & Walters, J. P. (1996). *Body Count*. New York, NY: Simon & Schuster.
- Braukmann, C. J., Kirigin Ramp, K. & Wolf, M. M. (1986). Follow-up of Group Home Youths into Young Adulthood. (Competing Renewal Application, Grant MH20030). Lawrence, KS: The University of Kansas Achievement Place Research Project.
- Campbell, D. T. & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin Company.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company.
- Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1973). Achievement Place: Experiments in Self-Government with Pre-Delinquents. *Journal of Applied Behavior Analysis*, 6, No. 1, 31 – 47.
- Fixsen, D. L., Blasé, K. A., Timbers, G. D., & Wolf, M. M. (2001). In Search of Program Implementation: 792 Replications of the Teaching-Family Model. G. A. Bernfield, D. P. Farrington (Eds.), *Offender Rehabilitation in Practice: Implementing & Evaluating Effective Programs*. New York, NY: John Wiley & Sons.
- Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1972). Achievement Place: The Reliability of Self-Reporting and Peer-Reporting and Their Effects on Behavior. *Journal of Applied Behavior Analysis*, 5, No. 1, 19 – 30.
- Fonagy, P., Target, M, Cottrell, D., Phillips, J., Kurtz, Z. (2002). *What Works for Whom? A critical Review of Treatments for Children and Adolescents*. New York, NY: Guilford Press.
- Friman, P. C. Osgood, D. W., Smith, G., Shanahan, D., Thompson, R. W., Larzelere, R., & Daly, D. L. (1996). A Longitudinal Evaluation of Prevalent Negative Beliefs about Residential Placement for Troubled Adolescents. *Journal of Abnormal Child Psychology*, 24, No. 3, 299 – 324.
- Glaser, D. (1980). The Interplay of Theory, Issues, Policy, & Data. In M. W. Klein & K. S. Teilmann (Eds.), *Handbook of Criminal Justice Evaluation*. Beverly Hills, CA: SAGE Publications.
- Hsia, H. M., Bridges, G. S., & McHale, R. (2004). *Disproportionate Minority Confinement 2002 Update*. Washington, D. C.: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice & Delinquency Prevention.

- Jenkins, P. (2006). *Decade of Nightmares*. New York, NY: Oxford University Press.
- Jones, R. R., Weinrott, M. R., & Howard, J. R. (1981). The National Evaluation of the Teaching Family Model. Final Report R01 MH Grants 25631 and MH 31018 awarded by the Center for Studies of Crime and Delinquency, National Institute of Mental Health. Eugene, OR: Evaluation Research Group, Inc.
- Kirigin, K. A., Wolf, M. M., Braukmann, C. J., Fixsen, D. L., & Phillips, E. L. (1979). Achievement Place: A preliminary Outcome Evaluation. In J. S. Stumphauzer (Edt.), *Progress in Behavior Therapy with Delinquents*. Springfield, IL: Charles C. Thomas.
- Kirigin, K. A., Braukmann, C. J., Atwater, J. D., & Wolf, M. M. (1982). An Evaluation of . Teaching-Family (Achievement Place) Group Homes for Juvenile Offenders. *Journal of Applied Behavior Analysis*, 15, 1 – 6
- Klein, M. W. & Teilman, K. S. (1980). *Handbook of Criminal Justice Evaluation*. Beverly Hills, CA: SAGE Publications.
- Larzelere, R. E., Dinges, K., Schmidt, M. D., Spellman, D. F., Criste, T. R., & Connell, P. (2001). Outcomes of Residential Treatment: A Study of the Adolescent Clients of Girls and Boys Town. *Child & Youth Care Forum*, 30, No. 3, 175 – 184.
- Larzelere, R. E., Daly, D. L., Davis, J. L., Chmelka, M. B., & Handwerk, M. L. (2004). Outcome Evaluation of Girls and Boys Town's Family Home Program. *Education and Treatment of Children*, 27, No 2, 130 – 149.
- Lipsey, M. W. (1990), *Design Sensitivity*. Newbury Park, NJ: Sage Publications.
- Lipsey, M. W. (1992). Juvenile Delinquency Treatment: A Meta-Analytic Inquiry into the Variability of Effects. In T. D. Cook, H. Cooper, D. Cordray, H Hartman, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller, (Eds.). *Meta-Analysis for Explanation: A Casebook*. New York, NY: Russell Sage Foundation.
- Lipsey, M. W. (1995). What Do We Learn from 400 Research Studies on the Effectiveness of Treatment with Juvenile Delinquents? In J. McGuire, (Edt.). *What Works? Reducing Offending*. New York, NY: John Wiley
- Lipsey, M. W. & Wilson, D. B. (1998). Effective Intervention for Serious Juvenile Offenders: A Synthesis of the Research. In R. Loeber & D. Farrington, (Edts.). *Serious & Violent Juvenile Offenders*. Thousand Oaks, CA: Sage Publications.
- Lipsey, M. W. (1999). Can Intervention Rehabilitate Serious Delinquents? *Annals, American Academy of Political & Social Sciences*, 564, July, pp. 142 – 167.
- Lipton, D., Martinson, R., & Wilks, J. (1975). *The Effectiveness of Correctional Treatment: A Survey of Evaluation Studies*. New York, NY: Praeger.
- MacKenzie, D. L. & Hebert, E. E. (1996). *Correctional Boot Camps: A Tough Intermediate Sanction*. Washington, D. C: National Institute of Justice.

- MacKenzie, D. L., Wilson, D. B., & Kider, S. B. (2001). Effects of Correctional Boot Camps on Offending. *Annals of the American Academy of Political & Social Sciences*.
- Martinson, R. (1974). What Works? – Questions and Answers about Prison Reform. *Public Interest*.
- Morris, E. K. & Braukmann, C. J. (1987). *Behavioral Approaches to Crime & Delinquency: A Handbook of Application, Research, and Concepts*. New York, NY: Plenum Press.
- Murphy, K. R., & Myers, B. (1998). *Statistical Power Analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Palumbo, D. & Sharp, E. (1980). Process Versus Impact Evaluations of Community Corrections. In D. Nachmias, (Edt.), *The Practice of Policy Evaluation*. New York, NY: St. Martin's Press.
- Parent, D. G., Leiter, V., Kennedy, S., Livens, L. Wentworth, D., & Wilcox, S. (1994). *Conditions of Confinement: Juvenile Detention and Corrections Facilities*. Washington, D. C.: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice & Delinquency Prevention.
- Phillips, E. L. (1968). Achievement Place Token Reinforcement Procedures in a Home-Style Rehabilitation Setting for 'Pre-delinquent' Boys. *Journal of Applied Behavior Analysis, 1, No. 3*, 213 – 223.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1971). Achievement Place: Modification of the Behaviors of Pre-Delinquent Boys with a Token Economy. *Journal of Applied Behavior Analysis, 4, No. 1*, 45 – 59.
- Phillips, E. L., Phillips, E. A. Fixsen, E. L. & Wolf, M. M. (1973). Behavior Shaping Works for Delinquents,” *Psychology Today, June 1973*, pp. 65 – 72.
- Phillips, E. L., Phillips, E. A., Wolf, M. M., & Fixsen, D. L. (1973). Achievement Place: Development of the Elected Manager System. *Journal of Applied Behavior Analysis, 6, No. 4*, 541 – 561.
- Phillips, E. L., Phillips, E.A., Fixsen, D.L., & Wolf, M. M. (1974). *The Teaching Family Handbook*. Lawrence, KS: University of Kansas, Department of Human Development Family Life.
- Quay, H. C. (1986). Residential Treatment. In H. Quay & J. S. Werry (Edts.), *Psychological Disorders of Childhood*. New York, NY: John Wiley & Sons.
- Rogers, P. J. (2000). Program Theory: Not Whether Programs Work But How They Work. In D. L. Stufflebeam, G. F., Madaus, & T. Kellaghan, (Edts.). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston, MA: Kluwer Academic Publishers.
- Roberts, J. V. & Stalans, L. J. (2000). *Public Opinion, Crime, and Criminal Justice*. Boulder, CO: Westview Press.

- Rossi, P. & Williams, W. (1972). *Evaluating Social Programs*. New York, NY: Seminar.
- Scheingold, S. A. (1984). *The Politics of Law & Order*. New York, NY: Longman.
- Snyder, H. N. & Sickmund, M. (2006). *Juvenile Offenders and Victims: 2006 National Report*. Washington, DC: U. S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- Steinberg, M. (1999). Cox Regression Examples. In Marketing Department, SPSS, Inc. (Edt.), *SPSS Advanced Models™ 10.0*. Chicago, IL: SPSS, Inc.
- Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (2000). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston, MA: Kluwer Academic Publishers.
- Stumphauzer, J. S. (1979). *Progress in Behavior Therapy with Delinquents*. Springfield, IL: Charles C. Thomas.
- Suchman, E. A. (1967). *Evaluative Research*. New York, NY: Russell Sage Foundation.
- Thompson, R. W., Smith, G. L., Osgood, D. W., Dowd, T. P., Friman, P. C., & Daly, D. L. (1996). Residential Care: A Study of Short- and Long-term Educational Effects. *Children & Youth Services Review, 18, No. 3*, 221 – 242.
- U. S. Department of Health and Human Services (1999). *Mental Health: A Report of the Surgeon General*. Rockville, MD: U. S. Department of Health & Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.
- Weiss, C. H. (1997). *Evaluation*. Upper Saddle River, NJ: Prentice Hall.
- Wilson, J. Q. (1983). Raising Kids. *Atlantic Monthly, October 1983*.
- Wilson, J. Q. & Herrnstein, R. J. (1985). *Crime & Human Nature*. New York, NY: Simon & Schuster.
- Wilson, J. Q. & Petersilia, J. (1995). *Crime*. San Francisco, CA: Institute for Contemporary Studies Press.
- Wolf, M., Braukmann, C. J., & Ramp, K. A. (1987). Serious Delinquent Behavior as Part of a Significantly Handicapping Condition: Cures and Supportive Environments. *Journal of Applied Behavior Analyses, 20*, pp. 347 – 359.
- Youngbauer, J. G. (1997). *The Teaching-Family Model and Treatment Durability: Assessing Generalization Using Survival Analysis Techniques*. Unpublished Doctoral Dissertation: The University of Kansas.
- Zimring, F. E. (1998). *American Youth Violence*. New York, NY: Oxford University Press.

Zimring, F. E. (2005), *American Juvenile Justice*. New York, NY: Oxford University Press.

**Author Contact Information:**

David E. Kingsley, PhD  
Department of Health Policy & Management  
University of Kansas Medical Center  
Mail Stop 3044  
3901 Rainbow Boulevard  
Kansas City, Kansas 66160  
Ph: 913 588 3763  
Email: [dkingsley@kumc.edu](mailto:dkingsley@kumc.edu)

---

## **Behavior Analyst Online**

[www.Behavior-Analyst-Online.org](http://www.Behavior-Analyst-Online.org)

**The Behavior Analyst Online organization (BAO) develops and deploys new resources for behavior analysts and makes them available on the Internet free of charge to the public. These resources are dedicated to educating the public about behavior analysis as well as serving as a resource for professionals involved in the field of behavior analysis.**

**The BAO organization is responsible to its membership to develop resources that the membership will find useful in everyday research, education, and application of the science of behavior analysis.**

**The BAO organization offers many perks to its members, including a Web Forum and the ABA-PRO Mailing List. In addition, the organization publishes several major free e-journals of interest to the behavior analysis community:**

**The Behavior Analyst Today**

**The Journal of Early and Intensive Behavior Intervention**

**The International Journal of Behavioral Consultation and Therapy**

**The Journal of Speech and Language Pathology - Applied Behavior Analysis**

**Membership in the BAO organization is free. For details, visit our website at**

**[www.behavior-analyst-today.org](http://www.behavior-analyst-today.org)**