Bryn Mawr College
Scholarship, Research, and Creative Work at Bryn Mawr College

Graduate School of Social Work and Social Research Faculty Research and Scholarship

Graduate School of Social Work and Social Research

2006

The Case for Multisystemic Therapy: Evidence or Orthodoxy?

Julia H. Littell
Bryn Mawr College, jlittell@brynmawr.edu

Let us know how access to this document benefits you.

Follow this and additional works at: http://repository.brynmawr.edu/gsswsr_pubs

 Part of the Social Work Commons

Custom Citation

Littell, Julia H. "The Case for Multisystemic Therapy: Evidence or Orthodoxy?" *Children and Youth Services Review* 28, no. 4 (2006): 458-472, doi: 10.1016/j.childyouth.2005.07.002.

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/gsswsr_pubs/34

For more information, please contact repository@brynmawr.edu.

Running head: Evidence or orthodoxy?

The case for Multisystemic Therapy: Evidence or orthodoxy?

In press in *Children and Youth Services Review*

Julia H. Littell

Bryn Mawr College

Author's note: I thank Jim Baumohl, Phoebe Cottingham, and Eileen Gambrill for helpful comments and suggestions on this paper.

Abstract

In this paper, I respond to comments by Henggeler, Schoenwald, Borduin, and Swenson (this issue) on my recent article, “Lessons from a systematic review of Multisystemic Therapy.” I identify factual and logical errors in their response, show how relevant research has been misinterpreted and misrepresented, and suggest constructive new directions for Multisystemic Therapy and the evidence-based practice movement.

Introduction

Sometimes science matters. New evidence or a new look at old evidence can disconfirm even hallowed assumptions. It is thus entirely reasonable to question whether evidence for the effectiveness of Multisystemic Therapy (MST) holds up under close scrutiny. Such an appraisal is warranted especially when rigorous, empirically-based techniques of research synthesis might usefully revise findings of earlier narrative reviews. In "Lessons from a systematic review of Multisystemic Therapy" (*Children and Youth Services Review, 27,* 445-463), I indicated that the effectiveness of MST remains in doubt. The effectiveness of MST was *not* the main subject of that article, however. In the "Lessons" article, I identified wide gaps between the science and practice of research synthesis and argued for enhanced rigor in the conduct, reporting, and synthesis of controlled trials. To illustrate my case, I used examples from my research team's systematic review and meta-analysis of controlled studies on the effects of Multisystemic Therapy, but as its title suggests the "Lessons" article was not intended to be a full account of that systematic review (for a full report, see Littell, Popa, & Forsythe, 2005).

Scott Henggeler, Sonja Schoenwald, Charles Borduin, and Cynthia Swenson chose to respond to selected portions of "Lessons" in "The Littell paper: Methodological critique and meta-analysis as Trojan Horse" (this issue). My purpose in the present paper is to identify and correct the factual and logical errors in their response, and to suggest constructive directions for MST and the evidence-based practice movement.

But it is important to consider first the context of this debate.

Dr. Henggeler and his coauthors are the developers of MST. They also served as principal investigators on the major federal grants used to test this intervention. By 2003, research grants for MST totaled $35 million (Henggeler, 2003). In January 2004, Dr. Henggeler

(2004a) announced the receipt of approximately $20 million in new research grants. Dr. Henggeler and his associates direct and hold stock in MST Services, Inc., a private consulting firm "which has the exclusive licensing agreement… for the dissemination of MST technology and intellectual property" (Rowland & Halliday-Boykins, 2004, p. 4). MST Services Inc. collects licensing, training, and consulting fees of approximately $400 to $550 per youth served (Strengthening Families America, 1999) and serves about 10,000 families per year (Henggeler, 2003).

Quite apart from their hard-won professional pride in their achievements, the developers of MST have a financial interest in this "model program." There is nothing intrinsically wrong with this, but the potential for conflict of interest is obvious and bears watching. As the developing and marketing of pharmaceuticals has taught us, independent trials and disinterested, rigorous, systematic reviews of all credible evidence should be essential elements of policy making for the public good.

I have no personal or financial stake in MST or any other intervention model. I have never met Dr. Henggeler or any of his colleagues. Contrary to their suggestions, if there is a "camp" with some "insidious strategy" to "camouflage the commitment to the status quo" (Henggeler et al., this issue), I have nothing to do with it. I take pride in being a member of the "disputatious community of 'truth-seekers'" that Donald Campbell (1988) envisioned. I have no other agenda, no secret weapons – no Trojan Horse.

Indeed, I share the concerns of Henggeler and his colleagues, and scholars such as Leonard Bickman (2002), about the lack of evidence for the effectiveness of most services for children and families – what Henggeler and colleagues called the "cottage industry" of mental health services. However, it is the "MST industry" that is the subject of *this* debate. Will the

MST industry utilize independent evaluations and evidence-based critiques to advance knowledge and practice on behalf of youth and families, or will it advance a new orthodoxy and demonstrate the "distrust, fear, and behavior protective of the status quo" of which Henggeler and his colleagues are so critical?

I return to these broader issues by way of conclusion. Immediately below, I consider matters of fact and logic.

Factual errors in Henggeler et al.'s response

Contrary to Henggeler et al.'s assertions, I drew no "conclusions" about the effects of MST in the "Lessons" article. Instead, I characterized *preliminary* findings as such and noted that these results might change as new information became available.

The suggestion that my papers were "broadly distributed through the Internet and media" is false. As a professional courtesy, I sent preliminary results of our systematic review to Henggeler and his colleagues, independent MST investigators, and reviewers whose work we cited, noting that this was a draft for comment and discussion only. Like Henggeler, his colleagues, and other scholars, I actively seek input on my work at scholarly and professional conferences and I make conference presentations available on request.

The claim that statements in my paper were not "substantiated with data" is false. I provided Henggeler and colleagues with a full report on our meta-analysis -- and most of the data in it were generated by them.

The claim that our systematic review was not peer-reviewed is false. After preliminary results were vetted by stakeholders and experts, and beginning in early January 2005, the final report was assessed by *ten* anonymous readers for the Cochrane Collaboration and the Campbell Collaboration.

The claim that I "ignored" feedback is false. I provided Henggeler and colleagues with detailed written responses to their criticisms on October 20, 2003, December 2, 2003, and February 11, 2005. I *disagree* with their views of the role of treatment fidelity and site-level variations in explaining results. These issues are addressed in our Cochrane review (Littell, Popa, & Forsythe, 2005) and are taken up below. To disagree with others is not to ignore them.

I did not suggest that the MST trials "were not credible" because of variable follow-up periods (or for any other reason), nor did I imply that the length of follow-ups was "arbitrary" or "manipulated" (Henggeler et al., this issue). If the studies were not credible, why would we include them in our review?

I did not ignore "the efficacy and effectiveness results generated in numerous…MST trials" (Henggeler et al., this issue). These data are included in our systematic review.

My research team did not use a study quality rating system to weight results of our meta-analysis. We did not privilege or elevate results of any study.

Logical errors

Equating the purposes of primary and secondary studies

Henggeler et al. claim that I "did not consider distinctions among purposes" of various MST trials (this issue). However, they did not recognize that, like any secondary analysis of data, a systematic review and/or meta-analysis[1] may be conducted for purposes other than those of the primary studies under review. For example, one can examine gender differences using data from studies that were not originally intended to address this issue. The purposes of my team's systematic review were twofold: We aimed to provide unbiased estimates of the *overall* (i.e.,

[1] Here the term "systematic review" means a review of available data that uses explicit inclusion and exclusion criteria, a systematic strategy to locate all potentially-relevant studies, inter-rater agreement on study inclusion decisions, and systematic coding and inter-rater agreement on key features of included studies. Meta-analysis refers to quantitative synthesis of results of multiple studies. Meta-analysis may or may not be part of a systematic review.

average) effects of MST in controlled studies of outcomes for youth and families and, to the extent possible, identify sources of heterogeneity (i.e., moderators) of these effects across studies. For reasons discussed below (and in our systematic review), available data do not support systematic moderator analysis; thus, it is not possible at present to determine why some studies reported greater effects than others. All of the studies in our review assessed effects of MST, and that is sufficient for our primary purpose, which mirrors proponents' claims that MST is effective across several service populations and settings (Henggeler, Schoenwald, Rowland, & Cunningham, 2002).

Post hoc explanations for heterogeneous effects

Any explanation for variations in results across MST studies is purely speculative at this point. The MST trials differ in terms of their sample characteristics, comparison conditions, perceived fidelity to MST, investigator independence, and the extent to which they support intent-to-treat analysis. Any of these factors can be used to explain the fact that results of early studies conducted by Henggeler and his associates were not replicated in the large, independent, multi-site Ontario trial (Leschied & Cunningham, 2002); yet we cannot know *which* factors were *responsible* for this difference because plausible sources of heterogeneity are confounded. Nevertheless, Henggeler and colleagues advanced several post hoc explanations for these differences, which I consider next.

Efficacy, effectiveness, and transportability

Henggeler (2004b) suggested that results of early MST efficacy trials were more promising than those of later studies of effectiveness, and MST developers took me to task for not recognizing this "difference." However, the distinctions between efficacy and effectiveness research are not at all clear in the original MST studies. Referring to the original MST trials,

Schoenwald and colleagues said the designs "could be considered hybrids of 'efficacy' and 'effectiveness' research" (Schoenwald, Sheidow, Letourneau, & Liao, 2003, p. 234). Post-hoc distinctions on these matters are debatable. Henggeler and colleagues refer to the Ontario trial as a transportability study, but according to its investigators, this study was designed to assess effectiveness and efficiency and it did *not* include plans for a systematic study of factors that affect transportability (see Leschied & Cunningham, 2002).

Fidelity

The MST developers claim that their model is implemented more or less faithfully across sites and therapists, and that greater fidelity to the model predicts more positive outcomes. They state that "demonstrating treatment fidelity is essential for establishing internal validity, and internal validity is essential for determining whether the study provides a fair test of the intervention's effectiveness" (Henggeler et al., this issue). These claims seem reasonable. The MST developers use their own Treatment Adherence Measure (TAM) to assess fidelity to their model. Although the TAM predicts outcomes in *some* studies, it does not appear to be a valid measure of fidelity *to MST* because it taps constructs that are considered essential to *any* therapeutic intervention, constructs such as client engagement, therapeutic alliance, and client satisfaction (sample items are: "The sessions were lively and energetic," "My family and the therapist worked together effectively," "We got much accomplished during the therapy sessions," and "The therapist recommended that family members do specific things to solve our problems;" see www.mstinstitute.org/TAM.htm). To my knowledge, the TAM's ability to discriminate between MST and any other intervention has never been tested (TAM scores are not available for control or comparison cases). Further, factor analyses of TAM scores have produced different results. Schoenwald, Sheidow, Letourneau, and Liao (2003) found that only

15 of the 26 TAM items loaded on a single factor. "This factor, labeled 'Therapist Adherence,' indexes the mutual engagement of the family and therapist in goal-setting, assessment, and intervention activities…[and] as such the 15-item factor reflects the distinct, but inter-related principles of MST" (Schoenwald et al., 2003, p. 228). However, it is unclear whether these principles are truly distinct from those that guide other interventions; hence, it is unclear whether the TAM measures fidelity to MST or more general practice principles, and it is unclear whether studies that use the TAM to measure fidelity to MST truly provide "a fair test of the intervention's effectiveness."

Ontario paid at least $250,000 to MST Services, Inc. for licensing, training, and consultation to support implementation of MST and achieve fidelity to the model (Leshied & Cunningham, 2002). *After* Ontario outcome data were available and in the absence of hard data on site-level differences in fidelity (TAM scores were not associated with outcomes in Ontario), MST developers used their own judgments about site-level differences in fidelity in Ontario (judgments contrary to those of on-site investigators) to argue that fidelity mattered (they said, "although the quality and quantity of adherence data are largely unknown, the site with apparently the worst adherence had the worst outcomes," Henggeler et al., this issue).

Program maturity

Henggeler and colleagues point to evidence of program maturity effects in the Ontario study and claim that I "ignored" the finding that "MST outcomes improved following the first year of program operation" (Henggeler et al., this issue). The data show that outcomes for the first 50 MST cases in the Ontario study were more negative than outcomes for the next 50 cases (Cunningham, 2002). However, as the Ontario investigators noted, random assignment is less likely to equalize differences between small groups than between larger ones; cases initially

assigned to MST had more prior arrests than cases initially assigned to usual services; then, as more cases were added to the experiment, the MST and usual services groups became equivalent on baseline measures. Further, the same graph that compares outcomes for the first and second groups of 50 cases also includes outcomes for a *third* group of 50 MST cases that was more like the first group than the second, showing that MST outcomes *did not* improve consistently with program maturity (Cunningham, 2002). Finally, there were no changes over time in TAM scores. Thus, "program maturity effects" can only be inferred if the evidence is used selectively.

Site-level variations

MST developers charge that I failed to investigate site-level variations in effects. As an anonymous reader observed, "Examining 'site effects' seems counter to the notion of a meta-analysis, which is designed to aggregate the findings of all available research." Still, if there is a sufficient number of sites and systematic variation on factors that may explain differences in effect sizes, then subgroup and moderator analyses can be conducted to explore possible sources of heterogeneity of effects. Of the eight experiments in our systematic review, three were conducted in multiple sites. Two of the three multi-site studies did not report site-level data. The Ontario study reported interim results by site; however, contrary to the MST developers' claim that one of the four Ontario sites "demonstrated clearly favorable outcomes for MST" (Henggeler et al., this issue), *there were no significant between-group differences on any outcome measure in any site* in Ontario. MST cases performed better on about half of the outcome measures and usual services cases performed better on the other half, but none of these differences were statistically significant (Leshied & Cunningham, 2002).

The Ontario site-level data that Henggeler and colleagues referred to (this issue) reflect (1) improvements over time (pre-post differences) in MST cases on certain outcomes and (2)

between-site differences in MST cases. Both comparisons are misleading and are based on faulty inferences. First, while MST cases tend to improve over time, so do comparison cases; in some sites, some pre-post differences are significant for MST cases but not controls. This could be due to statistical regression (MST cases had lower baseline scores on these measures) or other factors. In any case, it is certainly not correct to interpret within-group changes over time as evidence of intervention effects. Second, regarding cross-site comparisons, it is true that recidivism rates for MST cases were lower in some sites than in others, but the same pattern holds for comparison cases in those sites. The difference in recidivism rates across sites is not evidence of intervention effects.

The use of this kind of evidence to "provide a more balanced view of the Canadian findings" (Henggeler et al., this issue) is quite troubling. It ignores the proper contrasts (between treatment and comparison conditions) that were made possible by a randomized design and are necessary to support causal inferences (cf. Shadish, Cook, & Campbell, 2002).

Selective use of evidence

Henggeler and colleagues claim that "across sites, MST programs averaged a 10% reduction in convictions" (this issue) in Ontario. This appears to be based on interim 6-month follow-up data, which appear in the same table with other interim data that show that MST was associated with *increases* in the proportion of youth convicted of any offense at the one-year, two-year, and three-year follow-ups (none of the differences were statistically significant; Cunningham, 2002, Table 6). Moreover, data on the full sample (n = 409) confirm the interim finding that MST was associated with an *increase* in the proportion of youth convicted of an offense at the one-year follow-up (convictions occurred in 47% of MST cases and 42% of comparison cases; again, the difference was not statistically significant; see Littell et al., 2005).

Integrity of the MST Trials

Inconsistent reports and missing cases

The "Lessons" article provides an example of inconsistent reporting on the number of cases in MST trials. Due to a word processing error (which I alerted editors to before the article went to press), a key reference to a 1991 article was missing and is supplied below. As reported in my article, the first published report on the Missouri Delinquency Project indicated that the study included 210 families:

> A total of 210 families of juvenile offenders agreed to participate in the assessment and treatment components of the study. Following the initial assessment session, each family was randomly assigned to either multisystemic therapy or the alternative treatment group. Approximately 84% (n=88) of the families in multisystemic therapy and 65% (n=68) of the families assigned to alternative therapy completed treatment (Borduin & Henggeler, 1990, p. 76).

According to the next report (the missing 1991 reference),

> Following a pretreatment assessment session, adolescent offenders were randomly assigned to either MST (n = 100) or IC (n = 100)…Twenty-four (12%) of the families subsequently refused to enter treatment (Henggeler, Borduin, Melton, Mann, Smith, Hall et al., 1991, p. 45).

Then in 1995 and following reports, authors reported that 200 cases were assessed, but only 176 were randomly assigned (Borduin, Mann, Cone, Henggeler, Fucci, Blaske, et al., 1995, p. 570; Henggeler, Cunningham, Pickrel, Schoenwald & Brondino, 1996, p. 52). Henggeler et al.'s claim that I misread their 1995 paper is inaccurate (they said that I "failed to recognize that families who refused to participate in treatment (n = 24) were never randomized…," Henggeler et al., this

issue). This doesn't change (or even address) the fact that they provided *three* different accounts of the number of cases that were randomized in this study in three published papers. The distinction between consent to treatment and consent to participate in research (to which they appeal for explanation) is not the issue, nor is my reading comprehension at fault; the problem is that their accounts are inconsistent. If, as they claim, the first figure (210) was an "error," was the second number (200) wrong as well? It would be useful to know how this happened (as one Cochrane reader asked, how do investigators lose or miscount 16% (34/210) of the cases in an experiment?). Moreover, this is *not* the only MST trial in which this happened.

In a 1992 report on the Family and Neighborhood Services (FANS) study (also known as the Simpsonville project), the authors said,

> Ninety-six juvenile offenders were referred to the FANS project…Twelve youths were excluded from the research portion of this study for the following reasons: The youth did not have a felony arrest (n = 2), MST was never implemented (n = 6; families moved or refused to participate), random assignment was violated (n = 2; MST was court ordered), or recidivism data were not available on the state's computerized system (n = 2) (Henggeler, Melton, & Smith, 1992, p. 954).

Thus, at least two and perhaps as many as twelve cases were excluded after random assignment. However, subsequent reports described this study as a randomized experiment with 84 cases with *no mention* of the 12 excluded cases (e.g., "84 serious juvenile offenders…were referred by the Department of Youth Services…Youths were randomly assigned to usual services (n=41)…or MST (n=43);" Henggeler, Cunningham, Pickrel, Schoenwald, & Brondino, 1996, p. 50; see also Henggeler, Melton, Smith, Schoenwald, & Hanley, 1993, pp. 286-287). The problem is that we

never get a full and consistent account of how many cases were randomly assigned to groups. Similar problems occurred in two other MST trials (for details, see Littell et al., 2005).

MST trials are not the only studies in which investigators have eliminated certain cases, a dubious practice which Dennis Gorman termed "*post hoc* sample refinement" (Gorman, 2005).

Unyoked designs

In at least one of the MST trials, cases were referred in pairs that were randomly assigned to MST or a comparison group ("Eligible youths were referred by the DYS in yoked pairs, with one youth randomly assigned by the Department of Mental Health to receive MST and the other to receive the usual services," Henggeler, Melton, & Smith, 1992 p. 954). In other trials, "yoked" pairs appear to have been formed after random assignment. In both situations, whenever one case was lost during the study, its mate was retained in the analysis. We consulted two eminent scholars on the integrity of such "unyoked" designs. One expressed the view that this practice undermined the original randomized design, and such studies should be dropped from our analysis. The other thought the retention of unyoked cases made studies vulnerable to the "invidious bias" that drop-outs may produce, and indicated that it would be prudent for the investigators to conduct and report sensitivity analysis to determine whether dropping one or both members of a yoked pair mattered. On September 19, 2003, I asked Dr. Henggeler if it would be possible (for his colleagues or my team) to extract the unyoked cases and re-run analyses of program effects. He responded by saying that he did not think unyoked designs posed a threat to internal validity (Scott Henggeler, personal communication, September 19, 2003). He may be correct, but a sensitivity analysis would have settled the question and might have allowed us to have greater confidence in the results of these studies.

Unclear randomization procedures

As documented in our Cochrane review, several MST studies had unclear randomization procedures. I gave one example (from the MST Diffusion study) in the "Lessons" paper; another is provided in our systematic review. Again, the published descriptions of trials were not clear. On October 20, 2003, I asked Dr. Henggeler about the status of 9 MST cases in the Diffusion study and requested re-analysis of the study data without these cases, if possible, to enhance confidence in the findings. He did not answer this request.

Unstandardized observation periods

Variable follow-up lengths are quite common in clinical and field trials, and don't affect the credibility of a study *if* the data are analyzed properly. The experiments I've conducted all had variable follow-up lengths, which we handled with survival analysis and by identifying outcomes within fixed time intervals (e.g., at one-year post random assignment, including only those cases followed for the entire one-year period). On September 19, 2003, I asked Dr. Henggeler if his colleagues could provide us with data that would yield fixed-time-interval observation periods; they did not do this. Again, this would have increased our confidence in the results of several MST trials.

Subjective definitions of treatment completion

I agree with Henggeler and colleagues' view that it rarely makes sense to define treatment completion in terms of an arbitrary number of sessions. That is not the issue. The question is whether one includes treatment completers *and drop-outs* in analyses of outcomes. The inclusion of those who drop out and those who refuse treatment (after random assignment) is necessary for full intent-to-treat analysis (cf. Shadish, Cook & Campbell, 2002). If drop-outs and refusers are excluded, results are likely to be biased by differential attrition and, as I said before, "this is not an experimental comparison" (Littell, 2005, p. 456). Indeed, studies have shown that

MST drop-outs are *systematically* different from program completers, in that the former have more negative outcomes (Borduin et al., 1995; Leshied & Cunningham, 2002). The concern I expressed is that, when one defines "drop-out" as not completing treatment *satisfactorily*, the bias may be even more pronounced. In the MDP study, post-treatment assessments of psycho-social functioning are only available for program completers and "an MST completer...was judged by the therapist and clinical supervisor to have completed treatment successfully" (Schaeffer, 2000, p. 36).

"Flaws" in the Ontario study

Every study has flaws. The "flaws" that the Ontario investigators acknowledged were *characteristic of virtually all the MST trials* in our analysis, including the favorable studies conducted by the MST developers: Randomization procedures were not well-specified, were not centralized, and could have been subverted. Post-intervention assessments were not blinded and could have been biased. Some outcome data were collected by MST therapists. Other outcomes could have been biased by decision-makers' knowledge of participants' involvement in MST or usual services. Further, the Ontario investigators noted that random assignment does not control for all possible threats to internal validity (e.g., diffusion, compensatory rivalry). Acknowledgement of these problems -- common to MST trials -- does not reflect poorly on the Ontario investigators or the relative quality of their study.

Integrity of the Systematic Review

One of the advantages of a Cochrane or Campbell systematic review is that its procedures are transparent and extensive efforts are made to minimize bias of any type (including publication bias, selection bias, allegiance effects, and conflict of interest). The protocol for our review was published before the review was completed (contrary to MST developers' claims,

there were no "foregone conclusions"). Using the Cochrane Collaboration's RevMan software to generate meta-analyses and forest plots ensures that the data and analysis are completely transparent. In the forest plots, readers can literally see each piece of raw data (group means, standard deviations, valid *n*s) from each study in the analysis, alongside study-level effect sizes, the weights that were used and their effects, and the statistical models (random or fixed effects) used. We compiled additional documentation of the sources of each piece of raw data used in the meta-analysis.

"Elevating" an unpublished study

I believe the Ontario MST study is unpublished because of its null findings, not its methods. Publication bias (i.e., the tendency to publish studies that report significant, positive effects and not those with null or negative findings) affects not just reviewers' and editors' decisions, but also authors' decisions to submit articles for publication (for a thorough discussion of publication bias in meta-analysis, see Rothstein, Sutton, & Bornstein, in press). The Ontario MST trial had the most complete intent-to-treat (ITT) analysis of any study in our review -- and it found no significant differences between MST and usual juvenile justice services in outcomes in four sites.

Henggeler et al. (this issue) confuse ITT analysis with efficacy and suggest that we interpreted the Ontario trial as if it were an "efficacy study of the highest order." That is entirely incorrect. Nothing was done to privilege data from the Ontario trial in our analysis, nor were efficacy studies treated differently than effectiveness studies (as indicated above, we consider the Ontario study an effectiveness trial).

Although we rated MST trials in terms of their relative quality, contrary to Henggeler et al.'s claim, we did not *weight* our meta-analysis by study quality ratings. As stated clearly in the

"Lessons" article, we used inverse variance weights (a standard procedure in meta-analysis; see Lipsey & Wilson, 2001). Our quality rating scheme was used for interpretive purposes only.

Further, Ontario data were included in only 4 of the 21 analyses in our review; and all remaining analyses had overall effects that were not significantly different from zero. How, then, could Ontario be the "lynchpin" in our analysis, as Henggeler et al. claim?

"Idiosyncratic" study quality ratings

There are many tools for rating the methodological quality of controlled trials (Moher, Jadad, Nichol, Penman, Tugwell, & Walsh, 1995), but there is no consensus on the proper way to do this. We followed the Cochrane Collaboration's guidelines, which suggest that reviewers rate study qualities that (a) capture important variations in the set of studies under review and (b) may affect inferences that can be drawn from those studies (especially concerning internal validity of treatment outcome studies; see Alderson, Green & Higgins, 2004). Note that these guidelines suggest that rating methods *should be* idiosyncratic (I prefer the term "tailored") to fit the nature of the studies under review. If there had been variations in blinded assessment among the studies in our review, we would have included this feature in our rating schema. If we had included quasi-experiments, our rating scheme would have been much more complex in order to account for variations in study design.

"Stacking the deck"

An anonymous external reader suggested that we apply "a discount factor" for studies conducted by the MST program developers, assuming that allegiance bias or conflict of interest may have prejudiced those studies. Another anonymous reader suggested that we drop studies that did not provide transparent accounts of random assignment for all cases (as required in the CONSORT statement; see www.consort-statement.org). The fact that we did not take these

suggestions could be construed as a bias in *favor* of MST. The developers' claim that we were biased against MST is simply unfounded.

Generalizability

Although MST has spawned an industry, it is by no means "a proxy for evidence-supported practices in general" (Henggeler et al., this issue). If it turns out that MST is not consistently effective, this does not imply that other practices are ineffective, nor is "the entire field" discredited. Results of MST trials are simply not generalizable to other practices -- nor are they generalizable to MST programs that have not been studied.

War metaphors, ad hominem arguments, and false dichotomies

The response by Henggeler and colleagues takes an alarmist tone by invoking war metaphors ("camps," "camouflage," "attacks," "strategies," a "Trojan horse"), indicating that there is an "insidious strategy" to "discredit" MST and warning that "if MST…is discredited; the entire field is discredited" (Henggeler et al., this issue). They raise the specter of police-state strategies ("boot camps, incarceration, residential treatment, zero-tolerance policies, and scared straight programs") that may ensue should MST fall. This sets up a fearful, "us against them" mentality – a false dichotomy. The notion that people are either for or against evidence-based practice is another false dichotomy.

The MST developers make a number of unwarranted assumptions about other people's motives (especially my own) and they introduce ad hominem arguments that have no place in a serious, scholarly discussion. Ad hominem arguments violate a fundamental principle of science: Ideas should be judged on their logic and empirical support, not on their sources.

Henggeler and colleagues state that the implications of my "contentions" are two-fold: the use of empirically supported practices "should be eliminated" and the peer review process

"should be eliminated." This is nonsense. The fact that there are flaws in the evidence available to inform practice and defects in the peer-review process does not logically lead to the conclusion that we should abandon either. As stated clearly in my article, I believe these essential elements of knowledge development should be *strengthened*, not abandoned.

Programs don't simply "work" or "not work." Unfortunately, this false dichotomy underlies many current attempts to categorize human services. The purpose of a Cochrane or Campbell systematic review is to determine how effective one intervention is compared with others and (if possible) what factors account for variations in effectiveness. Such reviews can also inform us about the evidentiary status of interventions and tell us what we know and what we don't know in a particular field.

Unscientific appeals to authority

Henggeler and colleagues (this issue) reiterate points that I made about the wide acceptance of MST in government and professional organizations. Indeed, they supply an impressive list of official admirers. However, appeals to authority (to status and tradition) do not settle empirical questions. Because authorities are fallible, the norms of science are explicitly anti-authoritarian (Campbell, 1988). As I pointed out in "Lessons," the endorsements of MST by current authorities derive from less-than-rigorous scrutiny of the evidence. Some endorsements may flow from the political necessity to give the appearance of sponsoring programs that "work."

Premature closure of inquiry

Scientific knowledge is tentative and open to revision. It would be wonderful to find an intervention that had strong evidence of lasting, beneficial effects for children, youth, and families. In my view, it is premature to draw firm conclusions about the effectiveness of MST –

one way or the other. MST developers disagree. In a personal communication, Schoenwald questioned the value of my team's review of a treatment that has been "shown to work in efficacy or small-scale effectiveness trials" (Sonja Schoenwald, personal communication, November 21, 2003). Perhaps her faith will in the end be consistent with the evidence from independent trials, but that day has yet to come. Until then, if MST is truly science-based, its developers should seek falsification of their views by subjecting their predictions to critical tests, instead of trying to prove that MST "works" and invoking ad hoc explanations for disconfirming evidence.

Donald Campbell warned that when government asked them what to do, scientists would respond with assurances far beyond the state of the science and out-of-keeping with the tentative nature of scientific knowledge. Thus, according to Campbell, science requires a "disputatious community of 'truth-seekers'" to keep researchers from falling into the "over advocacy" trap (Campbell, 1988). In his view, nowhere is this more important than in the field of evaluation research, where political, personal, financial, and scientific interests can collide.

Summing up

I want to emphasize that the previous discussion does not imply that MST developers are engaged in dishonest or unethical practices. Their work is part of a culture that promotes programs that have *some* evidence of effectiveness. However, this is not consistent with the properly skeptical social science, envisioned by Donald Campbell.

The assertion that my work is a veiled "attempt to discredit" MST is entirely unfounded. There is no Trojan Horse here. The issue is: What is the foundation for MST? Is it built on scientific evidence that is free of allegiance bias? Has it withstood critical-rational assessments? Is it replicable in independent evaluations? Is it built on selective use of evidence by people who

developed and profit from one of today's leading "model programs"? Or has there been some mixture of science and sleight of hand in the making and successful marketing of MST? Until results of several independent, controlled trials are available, it is too soon to tell.

Looking back

Ten years ago, I argued that the case for intensive family preservation services (IFPS) was based largely on assertions and selective use of evidence; that program developers relied on their own studies and ignored results of larger, more rigorous, independent experiments (Littell, 1995). IFPS advocates decried this argument. Later, an independent, multi-site, federally-funded, controlled trial found no evidence that IFPS was significantly more effective than usual child welfare services (Westat, 2002). Unlike IFPS, MST was studied with randomized experiments at an early phase in its development. This is a significant step forward, and other program developers should follow the lead of Henggeler and his colleagues in using RCTs to test and improve interventions. Like IFPS, however, the early MST trials were not as rigorous or credible as larger, independent trials. In both cases, marketing, widespread use, and heightened expectations occurred well in advance of dispassionate, critical assessments of all the evidence, perhaps because these interventions promised a relatively "quick fix" for heretofore unsolved problems.

Moving forward

The fate of the current evidence-based practice movement may hinge on whether it becomes a new orthodoxy, founded on selective use of evidence and premature closure of inquiry, or remains open to critical assessments of evidence, incorporating new evidence, and revising empirically-based knowledge as needed. As indicated in the "Lessons" article, the Campbell Collaboration and the Cochrane Collaboration provide a platform for unbiased

synthesis and ongoing updates of research on intervention effects. If science matters, the MST developers and anyone interested in using evidence to guide practice and policy should embrace these efforts.

Before investigators try to "transport" off-the-shelf programs like MST, funders and consumers should insist on independent evaluations of their effectiveness. Absent this, the rapid proliferation of interventions, based on a handful of nonindependent trials, may eclipse initiatives to find even more effective approaches. It is also important to understand mechanisms of effective intervention (e.g., which components of MST are responsible for variations in outcomes?), using treatment dismantling strategies.

Finally, as my colleagues and I have said many times before, there may be real limitations to what can be accomplished with short-term interventions aimed at children and families, even when we engage them on their own terms and in their own milieus. Absent adequate and ongoing economic and social supports for families, resources such as affordable housing and substance abuse treatment, what happens to families when MST goes away? It is time to develop and test truly ecological intervention models, using place randomized trials (Boruch, 2005). These interventions can target families, neighborhoods, communities, and the public policies that affect them. And they can be subject to rigorous scrutiny.

References

Alderson, P., Green, S., & Higgins, J. P. T. (Eds.) (2004). Cochrane Reviewers' Handbook 4.2.2. In *The Cochrane Library, Issue 1, 2004* [updated March 2004]. Chichester, UK: John Wiley & Sons, Ltd.

Bickman, L. (2002). The death of treatment as usual: An excellent first step on a long road. *Clinical Psychology: Science and Practice, 9,* 195-199.

Borduin, C. M., & Henggeler, S. W. (1990). A multisystemic approach to the treatment of serious delinquent behavior. In R. J. McMahon & R. D. Peters (Eds.), *Behavior disorders of adolescence: Research, intervention, and policy in clinical and school settings.* (pp. 63-80). New York: Plenum Press.

Borduin, C. M., Mann, B. J., Cone, L. T., Henggeler, S. W., Fucci, B. R., Blaske, D. M., et al. (1995). Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology, 63,* 569-578.

Boruch, R. (2005). Preface: Better evaluation for evidence-based policy: Place randomized trials in education, criminology, welfare, and health. *The ANNALS of the American Academy of Political and Social Science, 599,* 6-18.

Campbell, D. T. (1988). The experimenting society. In E. S. Overman (Ed.), *Methodology and epistemology for social science: Selected papers* (pp. 290-314). Chicago: University of Chicago Press.

Cunningham, A. (2002). One step forward: Lessons learned from a randomized study of multisystemic therapy in Canada. PRAXIS: Research from the Centre for Children & Families in the Justice System. London, Ontario: Centre for Children and Families in the Justice System. Accessed June 2, 2005 at: http://www.lfcc.on.ca/One_Step_Forward.pdf.

Gorman, D. M. (2005). Drug and violence prevention: Rediscovering the critical rational dimension of evaluation research. *Journal of Experimental Criminology, 1,* 1-23.

Henggeler, S. W. (2003). Multisystemic therapy: An overview. Data, dissemination, and direction. Presentation at the National Association of State Mental Health Program Directors (NASMHPD) Research Institute Conference, February 2003.

Henggeler, S. W. (2004a). MST research update. Presentation at the Third International MST Conference, Charleston, SC, January 26-27, 2004. PowerPoint presentation accessed June 2, 2005 at www.mstconference.com/presentations/Henggeler.pdf.

Henggeler, S. W. (2004b). Decreasing effect sizes for effectiveness studies - Implications for the transport of evidence-based treatments: Comment on Curtis, Ronan, and Borduin (2004). *Journal of Family Psychology, 18,* 420-423.

Henggeler, S. W., Borduin, C. M., Melton, G. B., Mann, B. J., Smith, L. A., Hall, J. A., Cone, L. & Fucci, B. R. (1991). Effects of multisystemic therapy on drug use and abuse in serious juvenile offenders: A progress report from two outcome studies. *Family Dynamics of Addiction Quarterly, 1,* 40-51.

Henggeler, S. W., Cunningham, P. B., Pickrel, S. G., Schoenwald, S. K., & Brondino, M. J. (1996). Multisystemic therapy: An effective violent prevention approach for serious juvenile offenders. *Journal of Adolescence, 19*, 47-61.

Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using Multisystemic Therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology, 60*, 953-961.

Henggeler, S. W., Melton, B.B., Smith, L. A., Schoenwald, S. K. & Hanley, J. H. (1993). Family preservation using multisystemic treatment: Long tem follow-up to a clinical trial with serious juvenile offenders. *Journal of Child and Family Studies, 2,* 283-293.

Henngeler, S. W., Schoenwald, S. K., Borduin, C. M., & Swenson, C. C. (this issue). The Littell paper: Methodological critique and meta-analysis as Trojan horse. *Children and Youth Services Review,* doi: 10.1016/j.childyouth.2005.07.001.

Henggeler, S. W., Schoenwald, S. K., Rowland, M. D., & Cunningham, P. B. (2002). *Serious emotional disturbance in children and adolescents: Multisystemic Therapy.* New York: Guildford Press.

Leschied, A. W., & Cunningham, A. (2002). Seeking effective interventions for young offenders: Interim results of a four-year randomized study of Multisystemic Therapy in Ontario, Canada. London, Ontario: Centre for Children and Families in the Justice System. Accessed June 2, 2006 at: http://www.lfcc.on.ca/seeking.html.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Littell, J. H. (2005). Lessons from a systematic review of Multisystemic Therapy. *Children and Youth Services Review, 27,* 445-463.

Littell, J. H. (1995). Evidence or assertions? The outcomes of family preservation services. *Social Service Review, 69,* 338-351.

Littell, J. H., Popa, M., & Forsythe, B. (2005). Multisystemic Therapy for social, emotional, and behavior problems in youth age 10-17. *Cochrane Library, Issue 3, 2005*. Chichester, UK: John Wiley & Sons, Ltd.

Moher, D., Jadad, A., Nichol, G., Penman, M., Tugwell, T., & Walsh, S. (1995). Assessing the quality of randomized trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials, 16,* 62-73.

Rothstein, H. R., Sutton, A. J., & Bornstein, M. (in press). *Publication bias in meta-analysis: Prevention, Assessment and Adjustments.* Chichester, UK: John Wiley & Sons, Ltd.

Rowland, M. D., & Halliday-Boykins, C. (2004). Follow-up of Multisystemic Therapy (MST) as an alternative to hospitalization. Presentation at the 17th Annual RTC Conference, Tampa, FL, accessed June 2, 2005 at: http://rtckids.fmhi.usf.edu/rtcconference/17thconference/agenda/17th_handouts/pdf/Session%2019/Rowland.pdf.

Schaeffer, C. M. (2000). Moderators and mediators of therapeutic change in multisystemic treatment of serious juvenile offenders. Unpublished dissertation. University of Missouri, Columbia.

Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of multisystemic therapy: Evidence for multilevel influences. *Mental Health Services Research, 5,* 223-239.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Strengthening Families America (1999). Multisystemic Therapy program. Accessed June 2, 2005 at: http://www.strengtheningfamilies.org/html/programs_1999/04_MST.html.

Westat, Inc. (2002). *Evaluation of family preservation and reunification programs: Final report.* Washington, DC: US Department of Health and Human Services' Assistant Secretary for Planning and Evaluation. Available at: http://aspe.os.dhhs.gov/hsp/fampres94/index.htm.