Is Stacking Intervention Components Cost-Effective? An Analysis of the Incredible Years Program

E. MICHAEL FOSTER, PH.D., ALLISON E. OLCHOWSKI, PH.D.,
AND CAROLYN H. WEBSTER-STRATTON, PH.D., F.A.A.N.

ABSTRACT

Objective: Research demonstrates that interventions targeting multiple settings within a child's life are more effective in treating or preventing conduct disorder. One such program is the Incredible Years Series, which comprises three treatment components, each focused on a different context and type of daily social interaction that a child encounters. This article explores the cost-effectiveness of stacking multiple intervention components versus delivering single intervention components. **Method:** The data involved 459 children, ages 3 to 8, who participated in clinical trials of the Incredible Years Series. Children randomized to one of six treatment conditions received one or more of the three following program components: a child-based program, a parent training program, and a teacher-based program instructing teachers in classroom management and in the delivery of a classroom-based social skills curriculum. **Results:** Per-child treatment costs and child behavior outcomes (observer and teacher reported) were used to generate cost-effectiveness acceptability curves; results suggest that stacking intervention components is likely cost-effective, at least for willingness to pay above $3,000 per child treated. **Conclusions:** Economic data may be used to compare competing intervention formats. In the case of this program, providing multiple intervention components was cost-effective. *J. Am. Acad. Child Adolesc. Psychiatry,* 2007;46(11):1414–1424. **Key Words:** cost-effectiveness, conduct disorder, intervention.

Conduct disorder (CD) is among the most prevalent and costly of the emotional and behavioral disorders affecting youth in the United States; recent estimates suggest a lifetime prevalence of roughly 10% (12% among males; 7% among females; Kessler et al., 2005; Nock et al., 2006). The median age-of-onset for CD is approximately 11 years (Kessler et al., 2005; Nock et al., 2006); however, a smaller group of "early starters" begin to show CD symptoms as early as preschool (Moffit, 1993). Childhood-onset CD (as opposed to adolescent-onset CD) is associated with higher rates of below-median income, below-median parental education, and comorbid ADHD (McCabe et al., 2004). Furthermore, early starters are more likely to display symptoms consistent with comparatively more serious and pervasive CD subtypes (Nock et al., 2006). If early starters do not receive behavioral intervention, they face bleaker developmental trajectories than do youth who develop CD in adolescence (Moffit, 1993).

Mastering social and self-regulation skills is often a challenge for children and adolescents with CD; as a result, these youth face an increased risk of peer rejection and peer isolation from an early age (Kaiser and Hester, 1997; Miller-Johnson et al., 2002). Young people with CD are three times more likely to develop anxiety disorders, including panic disorder, generalized anxiety disorder, and social phobia (Nock et al., 2006). In addition, children with CD often behave in ways that harm themselves and society (Institute of Medicine, 1989; Webster-Stratton and Reid, 2003). Youth with CD are more likely to drop out of school, use weapons, abuse substances, and become pregnant (Bardone et al., 1998; Nock et al., 2006; Robins and Price, 1991; Scott, 1998).

Accepted June 28, 2007.

Dr. Foster is with the Department of Maternal and Child Health, School of Public Health, University of North Carolina, Chapel Hill; Dr. Olchowski is affiliated with the Department of Biobehavioral Health, The Pennsylvania State University; and Dr. Webster-Stratton is with the University of Washington School of Nursing and is the founder of the University of Washington Parenting Clinic.

Correspondence to Dr. E. Michael Foster, Ph.D., School of Public Health, University of North Carolina, Rosenau Hall, Campus Box #7445, Chapel Hill, NC 27599-7445; e-mail: emfoster@unc.edu.

0890-8567/07/4611-1414©2007 by the American Academy of Child and Adolescent Psychiatry.

DOI: 10.1097/chi.0b013e3181514c8a

When youth become involved with crime and violence, the resulting costs of court processing, incarceration, and losses by victims can be enormous. Recent estimates suggest that one life of crime can cost society between $1.5 and $1.7 million (2003 dollars; Cohen, 1998). As a result, the willingness of policymakers and society to pay for effective prevention programs should be high.

The costs of CD speak to the necessity of early identification and prevention. Targeted toward very young children, early intervention approaches may be most effective because patterns of negative behavior may become entrenched over time (Kaiser and Hester, 1997; Keenan and Wakschlag, 2000; Webster-Stratton and Reid, 2003). Effective CD behavioral treatments typically rely on multiple components to address children's behavior across several important contexts, such as school and family (Kaiser and Hester, 1997; Kazdin, 2000). These multicomponent interventions target parent, teacher, and peer communication skills and other mediators.

Past research has investigated the effectiveness of multicomponent behavioral interventions; however, only this article and a related study (Olchowski et al., 2007) have examined whether multicomponent treatment approaches are cost-effective vis-à-vis those same components when delivered separately. This article is the first to apply cost-effectiveness acceptability curve methodology to this issue.

The article examines this question using 20 years of data from the Incredible Years (IY) Parents, Teachers, and Children Training Series, an evidence-based multicomponent intervention created to treat young children with early-onset conduct problems (e.g., oppositional defiant disorder, conduct disorder). An independent American Psychological Association review committee identified the IY Series as one of only two evidence-based multicomponent treatments shown to reduce conduct problems in young children (Brestan and Eyberg, 1998). In addition, the Office of Juvenile Justice and Delinquency Prevention recognizes the IY Series as an effective CD treatment and prevention program for young children (Webster-Stratton, 2000).

METHOD

The IY program, developed by one of the present authors (C.W.S.), and evaluated by colleagues at the University of Washington's Parenting Clinic, seeks to reduce delinquency, drug abuse, and violent behaviors among children with conduct problems. Treatment focuses on the reduction of conduct problems in children; the improvement of children's social, emotional, and academic abilities; the enhancement of parental competence; the encouragement of positive parenting techniques; and the promotion of teachers' classroom management skills (Webster-Stratton, 2000). In addition to serving as an intervention for children with CD, an adapted version of IY is used by agencies and schools as a low-cost, community-based prevention program for at-risk children. Noteworthy strengths of the IY Series include its cultural sensitivity, user-friendliness, and comprehensive nature. The IY Series promotes collaboration among parents, teachers, and IY facilitators so that progress made during the intervention is maintained following program completion (Webster-Stratton et al., 2001a).

The multicomponent nature of the IY Series stems from the three single treatment components that comprise the intervention. These treatment components include a child-based program (Child Training, or CT), a parent-based program (Parent Training, or PT), and a teacher-based program (Teacher Training, or TT). Each program component focuses on improving children's behavior in a unique setting through the promotion of socially appropriate interaction skills. Detailed descriptions of treatment component goals, curriculum, and implementation methods have been published (Webster-Stratton, 2000; Webster-Stratton et al., 2001a).

Multiple randomized control group studies conducted by the developer (Webster-Stratton, 1990; Webster-Stratton and Hammond, 1997; Webster-Stratton et al., 2001b; Webster-Stratton et al., 2001c; Webster-Stratton et al., 2004) and independent investigators (Barrera et al., 2002; Miller and Rojas-Flores, 1999; Scott et al., 2001; Taylor et al., 1998) have provided convincing evidence that the IY Series consistently improves child behavior across a range of contexts and indicators. In addition, when an independent American Psychological Association review committee reviewed findings from more than 82 studies of CD interventions, the IY Series and interventions based on the behavioral theories presented in the manual *Living With Children* (Patterson and Gullion, 1968) were identified as the only two behavioral interventions that met the criteria for well-established efficacious CD treatments (Brestan and Eyberg, 1998). (Refer to Brestan and Eyberg [1998] for a thorough review and comparison of existing child-based CD interventions.)

Since its development more than 20 years ago, the IY Series has been disseminated across multiple types of service agencies (mental health agencies, child welfare agencies, and schools) in the United States and abroad (e.g., Canada, United Kingdom, Norway). Agencies seeking to adopt the IY Series are responsible for purchasing treatment materials and for providing training for CT, PT, and TT facilitators from certified IY trainers as well as ongoing consultation. After the initial purchase of materials and start-up fees, IY Series components (i.e., CT, PT, and TT) may be regularly offered to participants at modest cost to the service agency.

The IY Series has been implemented using the three single treatment components either alone or stacked in various combinations. Program stacking refers to treatment designs in which two or more components are delivered simultaneously (e.g., CT plus the TT program and/or with the PT program).

Participant Characteristics

Outcome data from 21 separate cohorts of children taking part in six IY Series clinical trials were combined (Webster-Stratton, 1982; Webster-Stratton, 1984; Webster-Stratton, 1994; Webster-Stratton and Hammond, 1997; Webster-Stratton et al., 1989; Webster-Stratton et al., 2004). All of the clinical trials were performed in Seattle; participants resided within a 60-mile radius of the Seattle metropolitan area. Data collection took place over a span of nearly 20 years (between the early 1980s and late 1990s). The final combined sample involved 459 children between the ages of 3 and 8, who had been randomized to treatment and control groups. Outcome data from separate cohorts could be combined because of common program implementation and data collection procedures.

The following criteria were required for entry into the IY trials: the child was between 3 and 8 years of age; the child had no debilitating physical impairment, intellectual impairment, or history of psychosis and was not already receiving psychological treatment; the primary clinic referral reason was for conduct problems such as noncompliance, aggression, and oppositional behavior that continued for more than 6 months; parent-report symptoms on the Eyberg Child Behavior Inventory were >2 SD above the mean; and the child met criteria for oppositional defiant disorder and/or CD according to either *DSM-III-R* or *DSM-IV* (American Psychological Association, 1994) depending on the child's study entry date (Webster-Stratton and Reid, 2003).

Families were assigned to one of seven treatment conditions: Child Training only (CT); Parent Training only (PT); Child Training and Parent Training (CT + PT); Parent Training and Teacher Training (PT + TT); Child Training and Teacher Training (CT + TT); Child Training, Parent Training, and Teacher Training (CT + PT + TT); and a control condition (CON). A more detailed summary of participant characteristics may be found in Table 1.

Estimating Treatment Costs

Financial estimates provided by the developer of the IY Series were used to calculate per-child costs for each of the six treatment categories. Per-child costs reflect the payor (or agency) perspective and include all of the fees and expenses for which the agency adopting the IY Series is responsible. These fees and expenses

TABLE 1
Participant Summary Statistics by IY Treatment Category

Treatment Category	*N*	Child Ethnicity	Average Child's Age at Intake, mo	Average Mother's Age at Child Intake, y
Child training only (CT)	54 Boy: 43 Girl: 11	White: 48 Hispanic: 0 Black: 4 Other: 2	72.3	36.1
Parent training only (PT)	292 Boy: 215 Girl: 77	White: 265 Hispanic: 3 Black: 4 Other: 20	59.6	34
Both child and parent training (CT + PT)	38 Boy: 27 Girl: 11	White: 31 Hispanic: 1 Black: 2 Other: 4	72.4	35.4
Both parent and teacher training (PT + TT)	24 Boy: 22 Girl: 2	White: 21 Hispanic: 0 Black: 1 Other: 2	67.4	38.3
Both child and teacher training (CT + TT)	11 Boy: 9 Girl: 2	White: 7 Hispanic: 1 Black: 0 Other: 3	74.3	35.6
All three (CT + PT + TT)	19 Boy: 17 Girl: 2	White: 16 Hispanic: 1 Black: 0 Other: 2	71	39.9
Control	21 Boy: 19 Girl: 2	White: 18 Hispanic: 0 Black: 0 Other: 3	68.9	36.1
Total	459 Boy: 352 Girl: 107	White: 406 Hispanic: 6 Black: 11 Other: 36	69.4	36.5

IY = Incredible Years.

include the following: training and ongoing supervision of CT and PT group leaders and trained teachers; group leader salary (and fringes) including time for peer review, self-study, and preparation; costs of participant materials; and any additional program implementation costs (e.g., childcare costs, taxicab vouchers, snacks, dinners). Salary estimates are estimates based on the developer's experience in hiring personnel in a range of settings. Estimates presented here exclude costs associated with providing space for treatment sessions; agencies seeking to adopt the IY Series generally are able to provide space necessary for treatment implementation. The fees associated with space rental should be included in per-child costs for agencies unavailable to provide unused space. Financial estimates were inflated into 2003 dollars.

Per-child treatment costs were generated for the three single IY treatment components (CT, PT, and TT). These estimates were then summed to generate total per-child costs for each of the four stacked treatment combinations (CT + PT, PT + TT, CT + TT, and CT + PT + TT; i.e., total per-child cost of CT + PT = total per-child cost of CT + total per-child cost of PT). The CT + PT condition requires the purchase of only one set of parent manuals at the cost of \$179.40/12 parents; this fee was not duplicated when summing total per child costs for combined CT and PT. For that reason, the cost of the combination of treatments is less than the combined costs of each treatment when delivered separately. Table 2 lists per-person cost estimates for each of the IY treatment categories excluding the control condition.

Changes in Program Costs Across Cohorts

The above estimates assume that each new CT group leader, PT group leader, and TT trained teacher is retrained for each cohort. In real-world settings, however, leaders may treat multiple cohorts before being retrained. Leaders complete training only before the first IY sequence; therefore, training costs drop as the number of children participating in IY increases. Moreover, the time needed for CT and PT group leaders to review and prepare program materials diminishes with each sequence that they lead. Similarly, after the first cohort of IY is complete, one-time purchases of materials are unnecessary for additional cohorts. Therefore, with each additional cohort of participants, total per-child costs per treatment category decrease. Because our tabulations assume annual training, these cost figures are conservative.

Calculating Treatment Outcomes

Analyses presented here involve pre-test and post-test scores for two outcome measures: a combined 5-item Negative Child Behavior Score measured by the Dyadic Parent-Child Interactive Coding System- Revised (DPICS-R) and a teacher-reported Total Problem Behavior Score measured by the Behar Preschool Behavior Questionnaire (PBQ). Developed by Robinson and Eyberg (1981) and revised by Webster-Stratton (1989), the DPICS-R records behaviors of children and their parents in a home setting (Reid et al., 2004). Trained observers code 39 parental and 8 child behavioral categories during 30-minute in-home observations (Reid et al., 2004). The Negative Child Behavior Score is a composite outcome variable, representing five separate negative child behaviors: negative physical actions, destructive behaviors, yell/cry/whine, "smart talk," and overall behavior valence (Beauchaine et al., 2005). The Total Problem Behavior score was generated using outcome data from the PBQ; the PBQ was developed for use as a screening instrument by mental health professionals. This instrument has been used to assess children ages 3 to 6 by teachers who rate a series of behaviors on a scale from 0 to 2. Test-retest values and interrater reliabilities are acceptable (Behar, 1977; Behar and Stringfield, 1974).

Parent-reported child behavior outcome measures (not presented here) have also been used to evaluate the effectiveness of the IY Series over time. These measures are adapted from questionnaire-based assessments such as the Eyberg Child Behavior Inventory (Robinson et al., 1980), the Child Behavior Checklist (Achenbach, 1991), and the Parenting Stress Index (Abidin, 1983). Examples of specific outcome measures include the following: Total Problem Score, Total Internalizing Score, Total Externalizing Score, Total Anxious-Depressed Score, and Child Demandingness Score (all mother reported).

Using data from the Negative Child Behavior Score and Total Problem Behavior Score, mean difference scores were created by subtracting each child's posttest score from their pretest score. Children were age 5 on average at the time of the intervention; in most instances, they were age 7.5 on average at the time of the postintervention assessment. (For one of the studies providing data on PT only, data were collected at age 6. Rather than discard these cases, we included the earlier data. Among those receiving parent training only, scores did not differ depending on when the follow-up data were collected [$p = .96$]. The inclusion of these data also did not affect other key tests, such as comparisons of change scores across treatment groups.) Difference scores for each treatment category were standardized and divided by the standard deviation of the pretest PT-only group score (the largest treatment group). Negative behaviors are coded highly using the two outcome measures considered here; therefore, lower posttest scores indicates that the treatment reduced negative child behavior.

Cost-Effectiveness Acceptability Curves

Traditional methods of cost-effectiveness analysis focus on incremental cost-effectiveness ratios (ICERs; Drummond and McGuire, 2001; Drummond et al., 1997). The ICER represents the additional or incremental expenditures required to improve the outcome measure by one unit. In a study of an intervention for cardiovascular disease, for example, an ICER could be calculated as dollars expended per myocardial infarction avoided.

To estimate an incremental cost-effectiveness ratio, individual per-person costs from two separate treatments—Treatment New (N) and Treatment Old (O)—are subtracted from one another and divided by the difference in per-person effectiveness:

$$\widehat{ICER} = (\hat{C}_N - \hat{C}_O)/(\hat{E}_N - \hat{E}_O) < \lambda$$

C and E represent the costs and the effects of the two treatments. The ICER represents the incremental costs divided by the incremental effectiveness. The "hat" reminds the reader that both the numerator and the denominator are estimates based on the experiences of samples of individuals participating in clinical trials. As a result, the estimated ICER has a confidence interval.

The key question in cost-effectiveness analysis is whether the ICER of a new treatment relative to an existing one exceeds a policymaker's or society's willingness to pay (λ) for the outcome of interest. For these analyses, λ can be interpreted as society's (or a policymaker's) willingness to pay for a 1-SD improvement in the measures of behavior problems used as outcomes. If the ICER is less than λ, the new technology or treatment is desirable (or preferred to treatment O, the old treatment).

Because the ICER is an estimate, however, inequality (1) is probabilistic—that is, it cannot be said for certain whether the

TABLE 2
Total Per-Child Cost by IY Treatment Category

Treatment Combination	CT: Child Training Only	PT: Parent Training Only	CT + PT: Both Child and Parent Training	PT + TT: Both Parent and Teacher Training	CT + TT: Both Child and Teacher Training	CT + PT + TT: All Three
Training fees						
CT leader training by certified IY trainer (three 8-hour days)	$3,600		$3,600		$3,600	$3,600
PT leader training by certified IY trainer (three 8-hour days)		$3,600	$3,600	$3,600		$3,600
TT leader training by certified IY trainer (four 8-hour days)				$4,800	$4,800	$4,800
CT leader-in-training's time (three 8-hour days)	$480		$480		$480	$480
PT leader-in-training's time (three 8-hour days)		$480	$480	$480		$480
TT teacher-in-training's time (four 8-hour days)				$400	$400	$400
Material fees (training and small group session)						
CT small group session materials	$1,244		$1,244		$1,244	$1,244
CT leader lesson plans	$150		$150		$150	$150
CT small group session handouts (for 6 children)	$14		$14		$14	$14
PT training materials		$15	$15	$15		$15
PT small group session materials		$1,390	$1,390	$1,390		$1,390
CT and PT parent manuals (for 12 parents)	$179	$179	$179	$179	$179	$179
TT teacher handbook and classroom handouts				$35	$35	$35
Additional fees (staff time)						
CT leader's time in sessions	$880		$880		$880	$880
CT weekly supervision	$440		$440		$440	$440
PT leader's time in sessions		$720	$720	$720		$720
PT leader's additional time		$480	$480	$480		$480
TT consultation costs				$390	$390	$390
Additional fees (IY implementation)						
PT small group session meals		$1,000	$1,000	$1,000		$1,000
PT small group session babysitting fees		$1,080	$1,080	$1,080		$1,080
PT small group session taxicab vouchers		$240	$240	$240		$240
PT small group session day care costs		$288	$288	$288		$288
TT training session snacks				$160	$160	$160
Total cost (TC)	$6,987	$9,472	$16,280	$15,257	$12,772	$22,065
Total per-child cost (TC/no. of children served)*	$1,164	$1,579	$2,713	$1,868	$1,454	$3,003

Note: Total per-child costs were calculated under the assumption that CT and PT small group sessions served 6 children (e.g., total per-child cost of CT ($1,164) = ($6,987)/6). Total per-child costs for stacked treatment categories involving TT (i.e., PT + TT, CT + TT, and CT + TT + PT) were computed using two assumptions regarding the number of children served. For each teacher participating in TT, it was assumed that on average 20 classroom students were exposed to the program. Thus, the total cost of TT ($5,785) was divided by 20 to yield a total per-child TT cost of $289; this per-child cost was added to per-child costs of PT and CT to form the total per-child costs for the three stacked treatment categories involving TT; thus, total per-child costs for PT + TT, CT + TT, and CT + TT + PT were calculated using two different treatment numbers.

ICER is larger than λ. Rather, the best that can be done is to estimate the probability that the inequality is true. Calculating this probability is complicated by the unique statistical properties of ratios. A variety of solutions have been proposed involving the delta method, Feiler's method, and bootstrapping, among others (Briggs, 2000; Briggs and Fenn, 1998; O'Brien and Briggs, 2002; Sendi and Briggs, 2001).

None of these methods have proven to be entirely satisfactory. Even if appropriate confidence intervals were developed for ICERs, conceptual problems remain. Specifically, negative regions of the confidence interval are ambiguous. A negative ICER may indicate either that a new treatment is more costly and less effective or that the new treatment is less costly and more effective. Under these two scenarios, decision makers would make completely opposite decisions; however, a negative ICER value offers no insight into which of the two decisions is appropriate.

As a result, until recently, analysts skirted these issues by not reporting confidence intervals. This solution is hardly satisfactory, and for that reason, two related alternatives have been proposed: net benefits and the closely related net health benefits. Although both involve manipulating the familiar cost-effectiveness ratio, we use the former here (Hoch et al., 2002; Lothgren and Zethraues, 2000; Sendi and Briggs, 2001). This technique involves rearranging equation (1):

$$NB = \lambda(\hat{E}_N - \hat{E}_O) - (\hat{C}_N - \hat{C}_O) > 0 \qquad (2)$$

NB values >0 indicate that the new treatment is preferred relative to the old treatment. Policymakers and researchers are able to calculate the probability of a positive NB value (i.e., $P(NB > 0)$) across a range of values of λ; to generate cost-effectiveness acceptability curves (CEACs), these probability values are plotted. The CEAC is a useful tool for policymakers because different stakeholders may have varying tolerance for uncertainty. As a result, the CEAC provides policymakers with the ability to make informed decisions under uncertainty (Claxton et al., 2000).

RESULTS

The data provided in Table 1 indicate that children were not yet 6 years old at study entry. The vast majority of the children were white, and consistent with the population prevalence of CD, the majority of participants were male (American Psychological Association, 2000). The average maternal age was 36.5 years, suggesting that the youth comprising the sample may not represent first-born children.

Table 3 reports the change scores across waves for each treatment category as measured by the combined Negative Child Behavior score and the Total Problem Behavior score. (As discussed, for the purposes of these cost-effectiveness analyses, change scores were standardized using the standard deviation of the pretest score for the largest group, PT only.) For the Negative Problem Behavior score (observer reported), children's behavior improved over time for all treated groups, with the exception of those children who received CT only. The behavior of children randomized to the no-treatment condition also deteriorated over time.

For the Total Problem Behavior score (teacher reported), children's behavior improved over time for

TABLE 3
Summary of Changes in Outcomes, By Outcome Measure and IY Treatment Category

	Per-Child Effectiveness[a]							
	Reductions (Improvement) in Total Problem Behavior Score (PBQ; teacher reported)				Reductions (Improvement) in Negative Child Behavior Score (DPICS-R; observer reported)			
	Obs	Mean	SD	p^b	Obs	Mean	SD	p^b
CT	48	−2.24	6.13	0.01	54	0.36	2.38	0.27
PT	214	−1.80	6.87	0.00	292	−0.06	2.94	0.72
CT + PT	32	−3.13	7.23	0.02	38	−0.84	3.45	0.14
PT + TT	23	−5.17	5.64	0.00	24	−0.48	3.54	0.51
CT + TT	11	−2.25	7.16	0.32	11	−0.58	1.51	0.23
CT + PT + TT	16	1.50	7.28	0.42	19	−2.51	10.28	0.30
No treatment (control)	21	NA[c]	NA	NA	21	1.80	2.97	0.01
p^d		.07				.01		

Note: PBQ = Behar Preschool Behavior Questionnaire; DPICS-R = Dyadic Parent-Child Interactive Coding System-Revised; CT = child training only; PT = parent training only; CT + PT = both child and parent training; PT + TT = both parent and teacher training; CT + TT = both child and teacher training; CT + PT + TT = all three.

[a] Both outcome measures code negative child behavior highly; thus, a successful treatment produced lower posttest scores compared to pretest scores. A negative mean change score indicates that the IY treatment category reduced the frequency of negative child behavior.

[b] *p* Value pertains to the null hypothesis that change over time is 0.

[c] As discussed in the text, we assumed that the no-treatment group showed no improvement on this measure.

[d] *p* Value pertains to the null hypothesis that there is no variation across groups.

all treated groups, with the exception of those children who received all three treatments combined (CT + PT + TT). Lack of teacher-reported improvement for the CT + PT + TT treatment group was most likely a function of small sample size (*N* = 16); posttest difference scores generated using observer- and parent-reported measures indicated that children who received CT + PT + TT demonstrated significant improvements in behavior. We lack follow-up data for this measure for children randomized to the no-treatment group. We assume their condition remained unchanged. If their condition deteriorated (as the DPICS measure indicates), then these analyses understate the cost-effectiveness of treatment relative to no treatment. Because the no-treatment group serves as the reference group for all of the treated groups, this assumption does not influence the position of the treatment combinations relative to each other.

Table 3 reports two levels of significance for children's behavior outcome scores. The first (located in the last row of the table), reports the significance level for variation in change over time across the treatment groups. The second (located in the last column of each pane) reports the significance level for the change over time for each treatment group. As one may expect, given the small sample sizes, many of these differences were not significant. For that reason, it is essential that the cost-effectiveness analyses reflect this uncertainty.

For each treatment, we calculated net benefits as described in Equation 2. For these analyses, willingness to pay represented the willingness of a policymaker to pay for a 1-SD improvement in the behavior score. The "old" treatment or reference category was represented by the control condition (no treatment). We estimated the probability that a given treatment was cost-effective by estimating the probability that it produced the highest net benefits. That probability was estimated using 500 bootstrapped samples for each level of λ. (Net benefits for the no-treatment group equal zero. "No treatment" was the best option when the net benefits of all of the other options were negative.)

For each level of λ, the CEAC identifies the treatment that is most likely cost-effective. Ideally, at every willingness to pay, one of the lines would have approached 100%; if this were the case, then the policymaker would have been virtually certain as to the best treatment option. However, it can be seen that neither CEAC rose above 80%. This uncertainty reflected the level of power in the underlying studies. (To reduce cluttering, we removed from our figures those treatment choices for which the CEAC never rose above 20%. These choices generally involved a treatment that was more expensive and less effective than an alternative treatment included in the figure.)

Figure 1 presents the CEAC for teacher-reported behavior problems (PBQ). At low levels of λ, the option

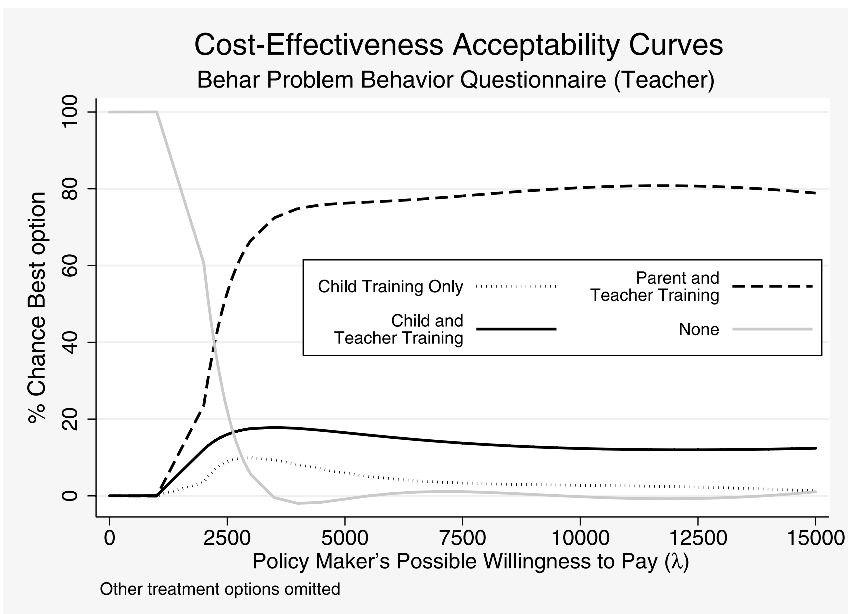

Fig. 1 Cost-effectiveness acceptability curves: PBQ (teacher reported). PBQ = Behar Preschool Behavior Questionnaire.

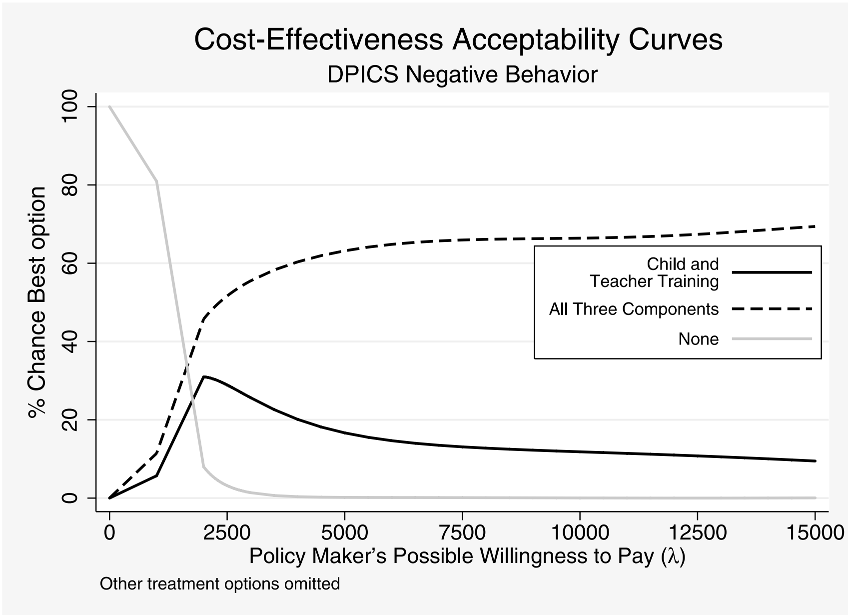

Fig. 2 Cost-effectiveness acceptability curves: DPICS-R (observer reported). DPICS-R = Dyadic Parent-Child Interactive Coding System-Revised.

most likely to be cost-effective was no treatment. Of course, this finding makes sense—if no value is placed on the outcome, the best option was the least costly. Once λ rose to a modest level of roughly $3,000, PT + TT was most likely to be cost-effective.

Figure 2 presents the CEAC curve for the DPICS observational data on behavior problems. At low levels of willingness to pay, the best treatment option again was no treatment. For reasonable values for λ, however, the likelihood that all three components combined (CT + PT + TT) was cost-effective reaches roughly 70% and exceeded other treatment options. It is important to note that this line would never exceed 70% no matter how high willingness to pay. As willingness to pay approaches infinity, the probability that a treatment option is cost-effective simply equals the probability that it is effective. That probability is determined by the power of the study as originally designed.

DISCUSSION

This article examines the differential cost-effectiveness of delivering multiple treatment components in combination versus the delivery of a single component within the context of the Incredible Years program. A sophisticated (and appropriate) cost-effectiveness methodology—cost-effectiveness acceptability curves—was applied to the data.

Cost-effectiveness acceptability curves represent an important tool for comparing competing treatments or treatment formats. The degree of certainty is a function of sample size; however, given the costs associated with large clinical trials of behavioral interventions for children, it is likely that data from several cohorts will be combined to generate larger samples. For this reason, it is important that researchers perform randomization and use an identical set of outcome measures longitudinally.

These analyses support stacking intervention components as a cost-effective strategy for treating and preventing behavior problems in young children. Except at low levels of willingness to pay for reductions in behavior problems, combinations of treatment components are cost-effective. At a modest level of willingness to pay ($3,000), the treatment most likely cost-effective involves a combination of treatments. A policymaker's willingness to pay is likely far greater than this figure. Focusing on public costs alone, Foster and colleagues find that the costs of conduct disorder are enormous and exceed $70,000 over a 7-year period (Foster et al., 2005). Clearly not all children at risk will go on to develop CD whether or not they receive early intervention. Even a 10% reduction in the chance of developing conduct disorder would imply a willingness to pay of $7,000. Furthermore, these costs of illness figures are conservative; they do not capture the broader costs borne by other members of society (e.g., victims of crime).

The specific nature of the components combined did matter; simply providing multiple components is not necessarily cost-effective. To some extent, the cost-effective combination depends on the outcome chosen. For problems at school, PT + TT is cost-effective; for problems at home, the combined treatment condition, CT + PT + TT, is cost-effective. The main (additional) benefits of child training may accrue at home. For example, it is possible that in the classroom context, the benefits of TT (i.e., providing teachers with methods for motivating students, promoting social competence, reducing disruptive behavior, and encouraging appropriate problem-solving techniques [Webster-Stratton et al., 2001a]) improve the learning environment to such an extent that children have less need to rely on the peer interaction skills promoted by CT. In the home environment, children interact with siblings and friends outside the structure provided by the classroom; in this context, children may have more opportunities to implement the prosocial skills learned in CT.

Limitations

The data used in this study reflect methodological limitations of the original studies, such as fairly small sample sizes. As a result, a substantial amount of uncertainty remains concerning the choice of a cost-effective program. Another key limitation of the study involves its generalizability. All of the participants were clinic-referred, suggesting that their behavior problems were relatively serious; in addition, nearly all of the study's participants were white. It is important to note, however, that the IY Series has effectively improved child behavior within minority samples as well as in low-income samples. As an example, when PT was implemented among Head Start children, treatment children demonstrated significantly lower levels of conduct problems and higher prosocial behavior compared to control peers (Reid et al., 2004). Among Head Start children, the effectiveness of the IY programs did not vary across samples of African American, Asian American, white, and Hispanic American children (Reid et al., 2001). Other studies have found that the IY Series significantly reduced problem behavior within ethnically diverse samples of low-income toddlers (Gross et al., 2003). Therefore, although the effectiveness of the IY Series has been substantiated within minority and low-income populations, its cost-effectiveness has not been explored for diverse samples. Whether and how the cost-effective choice of treatments would differ for non–clinic-referred and minority youth is an area for future research.

Another limitation involves the relatively short time horizon for the study. As noted above, the cost-effectiveness of the program likely depends on the degree to which young children with behavior problems develop CD. To date, no long-term data on the impact of the Incredible Years are available. As noted above, however, because of the program's relatively low cost, only a modest reduction in adolescent problems are required to make the program cost-effective. Whether and how such reductions occur is an important area for future research.

Clinical Implications

If society's resources for treating behavior problems in young children are limited, then those resources may best be focused on offering more services to fewer children. It is important to remember that the samples analyzed here were clinic referred and met diagnostic criteria for CD. A single-component strategy may prove cost-effective for children and youth with less severe problems. In a population context, a mixed strategy may prove most cost-effective with the most severe cases receiving the most intensive services.

Even within the samples considered, variation may exist in the cost-effectiveness of multiple components or in the particular components combined. Such variation has been found in other studies. In analyses of the cost-effectiveness of the treatments offered in the Multimodal Treatment for Attention Deficit Disorder study, the combination of medication management and an intensive psychosocial treatment was cost-effective, but only for cases with comorbid CD (Foster et al., 2006). The relatively small size of the samples included here prevent our exploring this issue further, but the issue remains an important one for future research.

Disclosure: Dr. Webster-Stratton provides training and supplemental instructional materials for the programs described in this article, and therefore stands to gain financially. This interest has previously been disclosed to the University of Washington and is being managed consistent with federal and university policy. The other authors have no financial relationships to disclose.

REFERENCES

Abidin RR (1983), *Parenting Stress Index Manual*. Charlottesville, VA: Pediatric Psychology Press

Achenbach TM (1991), *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: Department of Psychology, University of Vermont

American Psychological Association (2000), *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. Washington, DC: American Psychological Association

Bardone AM, Moffit TE, Caspi A, Dickson N, Stanton WR, Silva PA (1998), Adult physical health outcomes of adolescent girls with conduct disorder, depression, and anxiety. *J Am Acad Child Adolesc Psychiatry* 37:594–601

Barrera M Jr, Biglan A, Taylor TK et al. (2002), Early elementary school intervention to reduce conduct problems: a randomized trial with Hispanic and non-Hispanic children. *Prev Sci* 3:83–94

Beauchaine TP, Webster-Stratton C, Reid MJ (2005), Mediators, moderators, and predictors of one-year outcomes among children treated for early-onset conduct problems: a latent growth curve analysis. *J Consult Clin Psychol* 73:371–388

Behar L (1977), The Preschool Behavior Questionnaire. *J Abnorm Child Psychol* 5:265–275

Behar L, Stringfield SA (1974), A behavior rating scale for the preschool child. *Dev Psychol* 10:601–610

Brestan EV, Eyberg SM (1998), Effective psychosocial treatment of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *J Clin Psychol* 27:180–189

Briggs AH (2000), Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 17:479–500

Briggs AH, Fenn P (1998), Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ* 7:723–740

Claxton K, Lacey LF, Walkter SG (2000), Selecting treatments: a decision theoretic approach. *J R Stat Soc Ser A Stat Soc* 163:211–225

Cohen MA (1998), The monetary value of saving a high risk youth. *J Quant Criminol* 4:5–33

Drummond MF, McGuire A (2001), *Economic Evaluation in Health Care: Merging Theory with Practice*. New York: Oxford University Press

Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW (1997), *Methods for the Economic Evaluation of Health Care Programmes*, 2nd ed. New York: Oxford University Press

Foster EM, Jensen PS, Schlander M et al. (2006), Treatment for ADHD: is more complex treatment cost-effective for more complex cases? *Health Serv Res* 42:165–182

Foster EM, Jones DE, Bierman KL, Coie JD, Conduct Problems Prevention Research Group (2005), The high costs of aggression: public expenditures resulting from conduct disorder. *Am J Public Health* 95:1767–1772

Gross D, Fogg L, Webster-Stratton C, Garvey C, Julion W, Grady J (2003), Parent training of toddlers in day care in low-income urban communities. *J Consult Clin Psychol* 71:261–278

Hoch JS, Briggs AH, Willan AR (2002), Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ* 11:415–430

Institute of Medicine (1989), *Research on Children and Adolescents with Mental, Behavioral, and Developmental Disorders. Mobilizing a National Initiative. Report of a Study by a Committee of the Institute of Medicine, Division of Mental Health and Behavioral Medicine*. Washington, DC: National Academy Press

Kaiser AP, Hester PP (1997), Prevention of conduct disorder through early intervention: a social-communicative perspective. *Behav Disord* 22:117–130

Kazdin AE (2000), Treatment of conduct disorders. In: *Conduct Disorders in Childhood and Adolescence*, Hill J, Maughan B, eds. Cambridge, UK: Cambridge University Press

Keenan K, Wakschlag LS (2000), More than the terrible twos: the nature and severity of behavior problems in clinic-referred preschool children. *J Abnorm Child Psychol* 28:33–46

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005), Lifetime prevalence and age-of-onset distributions of *DSM-IV* disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 62:593–602

Lothgren M, Zethraues N (2000), Definition, interpretation, and calculation of cost-effectiveness acceptability curves. *Health Econ* 9:623–630

McCabe KM, Rodgers C, Yeh M, Hough R (2004), Gender differences in childhood onset conduct disorder. *Dev Psychopathol* 16:179–192

Miller-Johnson S, Coie JD, Maumary-Gremaud A, Bierman K, Conduct Problems Prevention Research Group (2002), Peer rejection and aggression and early starter models of conduct disorder. *J Abnorm Child Psychol* 30:217–230

Miller LS, Rojas-Flores L (1999), *Preventing Conduct Problems in Urban, Latino Preschoolers Through Parent Training: A Pilot Study*. New York: New York University Child Study Center

Moffit TE (1993), Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev* 100:674–701

Nock MK, Kazdin AE, Hiripi E, Kessler RC (2006), Prevalence, subtypes, and correlates of *DSM-IV* conduct disorder in the National Comorbidity Survey Replication. *Psychol Med* 36:699–710

O'Brien BJ, Briggs AH (2002), Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Stat Methods Med Res* 11:455–468

Olchowski AE, Foster EM, Webster-Stratton C (2007), Examining the differential cost-effectiveness of behavioral health interventions: an assessment of the Incredible Years Series. *J Early Intensive Behav Intervention* 4:284–304

Patterson GR, Gullion ME (1968), *Living with Children: New Methods for Parents and Teachers*. Champaign, IL: Research Press

Reid MJ, Webster-Stratton C, Baydar N (2004), Halting the development of conduct problems in Head Start children: the effects of parent training. *J Clin Child Adolesc Psychol* 33:279–291

Reid MJ, Webster-Stratton C, Beauchaine TP (2001), Parent training in Head Start: a comparison of program response among African American, Asian American, Caucasian, and Hispanic mothers. *Prev Sci* 2:209–227

Robins LN, Price RK (1991), Adult disorders predicted by childhood conduct problems: results from the NIMH Epidemiologic Catchment Area Project. *Psychiatry* 54:116–132

Robinson EA, Eyberg SM (1981), The dyadic parent-child interaction coding system: standardization and validation. *J Consult Clin Psychol* 49:245–250

Robinson EA, Eyberg SM, Ross AW (1980), That standardization of an inventory of child conduct problem behaviors. *J Clin Child Psychol* 9:22–28

Scott S (1998), Aggressive behaviour in childhood. *BMJ* 316:202–206

Scott S Spender Q, Doolan M, Jacobs B, Aspland H (2001), Multicentre controlled trial of parenting groups for child antisocial behaviour in clinical practice. *BMJ* 323:194–198

Sendi PP, Briggs AH (2001), Affordability and cost-effectiveness: decision-making on the cost-effectiveness plan. *Health Econ* 10:675–680

Taylor TK, Schmidt F, Pepler D, Hodgins H (1998), A comparison of eclectic treatment with Webster-Stratton's Parents and Children Series in a children's mental health center. *Behav Ther* 29:221–240

Webster-Stratton C (1982), The long-term effects of a videotape modeling parent-training program: comparison of immediate and one-year follow-up results. *Behav Ther* 13:702–704

Webster-Stratton C (1984), A randomized trial of two parent-training programs for families with conduct-disordered children. *J Consult Clin Psychol* 52:666–678

Webster-Stratton C (1989), The relationship of marital support, conflict, and divorce to parent perceptions, behaviors, and childhood conduct problems. *J Marriage Fam* 51:417–430

Webster-Stratton C (1990), Enhancing the effectiveness of self-administered videotape parent training for families with conduct-problem children. *J Abnorm Child Psychol* 18:479–492

Webster-Stratton C (1994), Advancing videotape parent training: a comparison study. *J Consult Clin Psychol* 62:583–593

Webster-Stratton C (2000), *The Incredible Years Training Series*. Juvenile

Justice Bulletin 1-23. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention
Webster-Stratton C, Hammond M (1997), Treating children with early onset-conduct problems: a comparison of child and parent training interventions. *J Consult Clin Psychol* 65:93–109
Webster-Stratton C, Hollinsworth T, Kolpacoff M (1989), The long-term effectiveness and clinical significance of three cost-effective training programs for families with conduct-problem children. *J Consult Clin Psychol* 57:550–553
Webster-Stratton C, Mihalic S, Fagan A, Arnold D, Taylor T, Tingley C (2001a), *Blueprints for Violence, Book Eleven: The Incredible Years Parent, Teacher, and Training Series*. Boulder, CO: Center for the Study and Prevention of Violence
Webster-Stratton C, Reid MJ (2003), Treating conduct problems and strengthening social and emotional competence in young children: the Dina Dinosaur Treatment Program. *J Emotional Behav Disord* 11:130–143
Webster-Stratton C, Reid MJ, Hammond M (2001b), Preventing conduct problems, promoting social competence: a parent and teacher training partnership in Head Start. *J Clin Child Psychol* 30:283–302
Webster-Stratton C, Reid MJ, Hammond M (2001c), Social skills and problem-solving training for children with early-onset conduct problems: who benefits? *J Child Psychol Psychiatry* 42:943–952
Webster-Stratton C, Reid MJ, Hammond M (2004), Treating children with early-onset conduct problems: intervention outcomes for parent, child, and teacher training. *J Clin Child Adolesc Psychol* 33:105–124

Need for and Use of Family Leave Among Parents of Children With Special Health Care Needs Paul J. Chung, MD, MS, Craig F. Garfield, MD, MAPP, Marc N. Elliott, PhD, Colleen Carey, BA, Carl Eriksson, MD, MPH, Mark A. Schuster, MD, PhD

Objectives: Parents of children with special health care needs are especially vulnerable to work–family conflicts that family leave benefits might help resolve. We examined leave-taking among full-time–employed parents of children with special health care needs. *Methods:* We identified all children with special health care needs in 2 large inpatient/outpatient systems in Chicago, Illinois, and Los Angeles, California, and randomly selected 800 per site. From November 2003 to January 2004, we conducted telephone interviews with 1105 (87% of eligible and successfully contacted) parents. Among the sample's 574 full-time–employed parents, we examined whether leave benefits predicted missing any work for child illness, missing >4 weeks for child illness, and ability to miss work whenever their child needed them. *Results:* Forty-eight percent of full-time–employed parents qualified for federal Family and Medical Leave Act benefits; 30% reported employer-provided leave benefits (not including sick leave/vacation). In the previous year, their children averaged 20 missed school/child care days, 12 doctor/emergency department visits, and 1.7 hospitalizations. Although 81% of parents missed work for child illness, 41% reported not always missing work when their child needed them, and 40% of leave-takers reported returning to work too soon. In multivariate regressions, parents who were eligible for Family and Medical Leave Act benefits and aware of their eligibility had 3.0 times greater odds of missing work for child illness than ineligible parents. Parents with >4 weeks of employer-provided leave benefits had 4.7 times greater odds of missing >4 weeks than parents without benefits. Parents with paid leave benefits had 2.8 times greater odds than other parents of missing work whenever their child needed them. *Conclusions:* Full-time–employed parents of children with special health care needs experience severe work–family conflicts. Although most have leave benefits, many report unmet need for leave. Access to Family and Medical Leave Act benefits and employer-provided leave may greatly affect leave-taking. **Pediatrics** 2007;119:e1047–e1055.