PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

National Institute of Mental Health Multimodal Treatment Study of ADHD Follow-up: 24-Month Outcomes of Treatment Strategies for Attention-Deficit/Hyperactivity Disorder

MTA Cooperative Group

Pediatrics 2004;113;754-761
DOI: 10.1542/peds.113.4.754

National Institute of Mental Health Multimodal Treatment Study of ADHD Follow-up: 24-Month Outcomes of Treatment Strategies for Attention-Deficit/Hyperactivity Disorder

MTA Cooperative Group*

ABSTRACT. *Objective.* **In the Multimodal Treatment Study of ADHD (MTA), the effects of medication management (MedMgt) and behavior modification therapy (Beh) and their combination (Comb) and usual community comparison (CC) in the treatment of attention-deficit/hyperactivity disorder (ADHD) differed at the 14-month assessment as a result of superiority of the MTA MedMgt strategy (Comb or MedMgt) over Beh and CC and modest additional benefits of Comb over MedMgt alone. Here we evaluate the persistence of these beneficial effects 10 months beyond the 14 months of intensive intervention.**

Methods. **Of 579 children who entered the study, 540 (93%) participated in the first follow-up 10 months after the end of treatment. Mixed-effects regression models explored possible persisting effects of the MTA medication strategy, the incremental benefits of Comb over MedMgt alone, and the possible superiority of Beh over CC on 5 effectiveness and 4 service use domains.**

Results. **The MTA medication strategy showed persisting significant superiority over Beh and CC for ADHD and oppositional-defiant symptoms at 24 months, although not as great as at 14 months. Significant additional benefits of Comb over MedMgt and of Beh over CC were not found. The groups differed significantly in mean dose (methylphenidate equivalents 30.4, 37.5, 25.7, and 24.0 mg/day, respectively). Continuing medication use partly mediated the persisting superiority of Comb and MedMgt.**

Conclusion. **The benefits of intensive MedMgt for ADHD extend 10 months beyond the intensive treatment phase only in symptom domains and diminish over time.** ***Pediatrics* 2004;113:754–761; *ADHD, attention deficit, hyperactivity, stimulant medication, behavior therapy.***

Received for publication Mar 12, 2003; accepted Jul 30, 2003.

Reprint requests to (P.S.J.) Center for the Advancement of Child & Adolescent Mental Health, Department of Child Psychiatry, Unit 78, New York State Psychiatric Institute/Columbia University, 1051 Riverside Dr, New York, NY 10032. E-mail (L. Eugene Arnold): arnold.6@osu.edu

*The MTA is a cooperative treatment study performed by 6 independent research teams in collaboration with the National Institute of Mental Health and the Office of Special Education Programs of the US Department of Education, Washington, DC.

In the cited methodology papers for this study, the treatment assignments were called medication (MED), psychosocial treatment (PS), combined treatment (CT), and community-treatment assessment and referral (A&R). To reflect more accurately the actual treatments, the terminology is changed for all outcome papers to these more clear and more specific terms: medication management (MedMgt), behavioral treatment (Beh), combined treatment (Comb), and community comparison (CC).

The opinions and assertions contained in this report are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of Health and Human Services or the National Institutes of Health.

ABBREVIATIONS. MTA, Multimodal Treatment Study of ADHD; RCT, randomized clinical trial; ADHD, attention-deficit/hyperactivity disorder; MedMgt, medication management; Beh, behavior therapy; Comb, combined treatment; CC, community comparison; ITT, intention-to-treat; ODD, oppositional defiant disorder.

The 6-site National Institute of Mental Health Multimodal Treatment Study of ADHD (MTA) was designed as a randomized clinical trial (RCT) comparing well-established and widespread treatments for children with attention-deficit/hyperactivity disorder (ADHD). A total of 597 children with ADHD-Combined type, aged 7 to 9.9 years, were randomly assigned to 1 of 4 treatments: medication management (MedMgt), behavior modification (Beh), MedMgt and Beh combined (Comb), or routine community care comparison (CC).[1,2] At the end of the 14-month treatment phase, we found that all 4 groups enjoyed sizable improvement over time, with significant differences among groups in the rate of improvement in some areas. For ADHD symptoms, Comb and MedMgt improved significantly more than Beh or CC. Comb and MedMgt did not differ significantly on direct comparisons, but in several instances (parent-rated oppositional/aggressive symptoms, internalizing symptoms, teacher-rated social skills, parent–child relations, and reading achievement), Comb proved superior to Beh and/or CC, whereas MedMgt did not. The MTA intensive medication strategy (Comb/MedMgt) was superior to CC despite that two thirds of CC-treated participants received similar medication during the study.[1,2]

On the basis of these and other analyses,[3–6] we concluded that for ADHD symptoms, carefully crafted medication management (maintained through 14 months) was superior to Beh alone (faded by 14 months) and to routine CC that included medication. Although Comb showed significant advantages over CC in every domain whereas MedMgt did not, it did not prove significantly superior to MedMgt for individual, specific outcome measures. However, Comb did show modest significant advantages over MedMgt on global or composite outcome indices[3,4] that summed small but consistent advantages across domains. Comb also provided modestly greater benefits than MedMgt for non-ADHD symptom domains and positive functioning, as well as greater levels of parent satisfaction.[1,5,6] This additive effect was more pronounced for dually comorbid children, ie, those

with both internalizing (depression and/or anxiety disorder) and disruptive behavior disorders (conduct or oppositional-defiant disorder).[7,8]

A major question for an RCT is whether treatment effects persist.[9–12] Here we examine the status of the randomly assigned groups 10 months after the end of the treatment phase to address this question. Building on findings from the primary and secondary analyses at the end of treatment, in the follow-up analyses presented here, we use a set of orthogonal comparisons (ie, nonoverlapping analyses that test no more than once for potential differences among any of the specific groups) to help understand any overall treatment effect. Also, we use a parsimonious set of outcome measures across 5 areas of functioning, and, when possible, we combine related domains (eg, symptom ratings of Inattention and Hyperactivity/Impulsivity) and sources (ratings by parent and teacher) to reduce the number of analyses performed.[13–15] For these analyses, we continue to use the RCT gold standard, the intention-to-treat (ITT) approach (which uses all available subjects in the analysis, even if they failed to comply fully with the treatment assignment) to determine whether initial treatment assignment had a persistent impact on outcome at the first follow-up (10 months after the end of treatment, 24 months postrandomization).

METHODS

Sample

Table 1 shows the demographics of the 540 participants who were assessed in the 10-month follow-up, which are not different from the demographic characteristics of the originally randomized 579 MTA participants described elsewhere. The only statistically significant difference among treatment groups was a trivial difference in age (MedMgt was 0.3 years older than Beh). As a check, the first analysis below was repeated with age covaried; it made no difference. In addition to all children meeting criteria for ADHD-Combined type at baseline, parents were interviewed with the Diagnostic Interview Schedule for Children[16] for other child psychiatric disorders. The ADHD Diagnostic Interview Schedule for Children module was completed face-to-face with the child's principal caregiver by trained research interviewers.

Assessments

On the basis of experience from 14-month analyses and statistical advice, we selected 5 measures from conceptually distinct domains that were likely to be affected by treatment: 1) parent- and teacher-rated Swanson, Nolan, and Pelham Scale (SNAP)[17] ADHD symptoms, 2) parent- and teacher-rated oppositional defiant disorder (ODD) symptoms,[17] 3) Wechsler Individual Achievement Test reading score,[18] 4) a "negative/ineffective discipline" factor,[5,6] and 5) parent- and teacher-rated total social skills from the Social Skills Rating System (SSRS).[19]

Analyses

Univariate analyses of the 24-month means were performed with baseline covaried, with 1 main test for each of the 5 symptom and function domains, using a mixed-effects regression mod-

TABLE 1. Demographic Characteristics of MTA Participants in 24-Month Assessments

Variable	Totals Whole Sample (n = 540)	Comb (n = 138)	MedMgt (n = 128)	Beh (n = 139)	CC (n = 135)	Range Across Sites (P Value)
Participant variables						
Age (y; mean [SD])	8.4 (0.8)	8.4 (0.8)	8.6 (0.9)	8.3 (0.8)	8.5 (0.8)	8.4–8.6 (.01)‡
Male (n [%])	430 (80)	107 (78)	105 (82)	110 (79)	108 (80)	72%–88% (.06)
Ethnicity (n [%])						Overall χ^2 (.0001)
White	329 (61)	82 (59)	84 (66)	79 (57)	84 (62)	22%–81%
Black	105 (19)	23 (17)	23 (18)	36 (26)	23 (17)	4%–37%
Hispanic	46 (9)	14 (10)	11 (9)	12 (9)	9 (7)	0%–36%
Grade (n [%])						Overall χ^2 (.01)
First	84 (16)	20 (14)	16 (13)	29 (21)	19 (14)	6%–29%
Second	226 (42)	60 (43)	47 (37)	63 (45)	56 (41)	40%–45%
Third	163 (30)	43 (31)	48 (38)	33 (24)	39 (29)	22%–41%
Fourth	66 (12)	15 (11)	16 (13)	14 (10)	21 (16)	8%–19%
Fifth	1 (0.2)		1 (0.8)			0%–1%
WISC-3 IQ (mean [SD])						
Verbal	100.6 (14.9)	100.6 (15.6)	98.7 (14.0)	100.9 (14.4)	102.2 (15.2)	96.4–104.8 (.01)
Performance	101.4 (15.8)	101.1 (16.0)	100.1 (14.7)	101.5 (15.8)	103.2 (16.5)	95.1–104.9 (.001)
Total	101.0 (14.8)	100.7 (15.2)	99.1 (13.6)	101.1 (14.5)	102.8 (15.7)	95.2–105.1 (.001)
Parent/family variables						
High school graduate (n [%])						
Mother	505/536 (94)	131 (96)	116 (91)	130 (95)	128 (95)	87%–100% (.01)
Father	385/423 (91)	106 (94)	88 (86)	95 (91)	96 (92)	79%–99% (.0001)
Employed (n [%])*						
Mother	384/537 (72%)	93 (68)	86 (67)	105 (77)	100 (74)	62%–79% (.06)
Father	353/418 (84%)	94 (85)	87 (86)	89 (86)	83 (81)	71%–95% (.007)
Welfare (n [%])	100 (19)	28 (20)	24 (19)	21 (15)	27 (20)	10%–40% (.0001)
Income (n [%])						Overall χ^2 (.001)
0-$20K	105 (19%)	26 (19)	23 (18)	30 (22)	26 (19)	10%–34%
20-$50K	224 (41%)	52 (38)	58 (45)	56 (40)	58 (43)	35%–50%
$50K+	200 (37%)	58 (42)	43 (34)	52 (37)	47 (35)	27%–46%
Married (n [%])†	350 (65)	93 (67)	86 (67)	81 (58)	90 (67)	53%–72% (.09)
Family composition						
Parent (n [%])	145 (27%)	34 (25)	28 (22)	48 (35)	35 (26)	21%–37% (.12)

SD indicates standard deviation. Attrition from MedMgt was marginally greater than from Beh ($P < .05$, with Bonferroni-corrected significance level being .008). Most of the MedMgt attrition (13 of 16) occurred immediately after randomization, before beginning treatment.

* Refers to that proportion of sample whose parents hold full- or part-time jobs.

† Means proportion of sample with intact, 2-parent families (married or common-law).

‡ Note that treatment groups differed significantly on only 1 variable (age), whereas sites differ significantly on almost all variables.

el.[20–24] We set the significance level at $P < .01$ for each of the 5 tests to maintain an overall $P < .05$ significance level. Within each domain, we performed orthogonal (nonoverlapping) contrasts that were developed to decompose the overall effects at the end of treatment. Three statistically independent questions were addressed[4]: 1) Do children who were exposed to the MTA's intensive medication strategy (Comb or MedMgt) show persisting superior outcomes over children who were not (Beh and CC)—the "MTA medication algorithm effect"? 2) Do children who were assigned to Comb show superior outcomes over those who were assigned to MedMgt—"the multimodality superiority effect"? 3) Do children who were exposed to intensive Beh show superior outcomes over those in usual CC—"the behavioral substitution effect"?

In addition to these 5 domains of effectiveness, we obtained measures from 4 domains of services used during the follow-up phase (use and dose of medication, use of specialty mental health services, and use of special education services) from a structured interview developed for this purpose, the Services Use by Children and Adolescents–Parent Interview (SCAPI). The SCAPI asks the child's caregivers about any services that the child may have received since the previous assessment, including medication, physician or therapy contacts, and school services. We calculated the percentages of participants in each treatment group who received the respective service between 14 and 24 months and performed χ^2 tests to determine whether these percentages differed across the 4 treatment groups. Last, given the highly significant effects found at the 14-month endpoint as a function of the MTA medication algorithm (Comb and MedMgt groups), we repeated any of the above analyses that showed significant differences between groups at 24 months, this time adjusting for participants' medication use in the 14- to 24-month interim and in the 30-day period before the 24-month outcome assessments. These analyses were done to parse out the extent to which any persisting outcome differences might be fully or partly explained by participants' current medication use, rather than their original group assignment.

RESULTS

Table 2 shows the results of the ITT analyses for 5 effectiveness domains and the 4 services domains. The overall treatment effect was significant for the ADHD and ODD ratings but not for academic achievement, social skills, or negative/ineffective discipline.

In the decomposition of the significant ADHD and ODD effects into nonoverlapping components by the orthogonal contrasts, the MTA medication algorithm contrast was significant but the other 2 contrasts were not. In other words, the average for Comb and MedMgt was significantly superior to the average for Beh and CC, but for the other pairwise contrasts, Comb was not statistically different from MedMgt, and Beh was not statistically different from CC. For the symptom ratings, the main effect of rating source was significant (parents rated their children better than did teachers), but the source-by-treatment interactions were not (ie, these differences were present regardless of treatment, and relative treatment differences were similar regardless of source). For the significant ADHD and ODD outcomes, we calculated the effect size (Cohen d, estimated as the difference in the means for Comb+MedMgt vs Beh+CC, divided by the pooled standard deviation). For ADHD ratings, $d = 0.30$; for ODD ratings, $d = 0.21$, small effects.

Medication use showed differences between groups on both the proportions using medication at any point during the follow-up period (86%, 85%, 44%, and 69% for Comb, MedMgt, Beh, and CC, respectively; $\chi^2 = 77.2$, $df = 3$, $P < .001$) and in mean dosage of the stimulant-medicated participants (methylphenidate equivalents: 30.4, 37.5, 25.7, and 24.0 mg/day, respectively; $F = 13.15$, $df = 3, 363$; $P < .0001$). However, from end of treatment to the first follow-up, the percentage of participants on medication decreased for Comb (87% vs 70%) and MedMgt (93% vs 72%) but increased for Beh (23% to 38%) and CC (55% to 62%). Other services use domains showed no significant group differences. Therefore, only medication use was used as a mediator for subsequent analyses.

Mediator Analyses

To test whether any differences found were attributable to differential medication use during 14 to 24 months, we repeated the 5 univariate analyses with interim medication use covaried. We used 2 approaches to test for the effects of medication use, based on whether the participant took any medication prescribed for ADHD at any time during the 10-month interval or whether the participant was taking any of these medications in the 30 days before the 24-month assessment. (Medications in this analysis included not only the psychostimulants, which constituted ~90% of the medications prescribed across all 4 groups, but also bupropion, tricyclics, and α-agonists.) A third approach was briefly considered but rejected, on the basis of the advice of our statistical consultant: medication dose, in methylphenidate equivalents at the 24-month assessment. Examination of the data distribution showed that this variable was too skewed for valid use.

Table 3 repeats the first 5 analyses from Table 2 but with medication status covaried: whether the participant took medication for ADHD during the 14- to 24-month interim. In addition, these same analyses were conducted using the medication variable whether the participant took medication in the 30 days before the 24-month assessment. There were no substantive differences between these 2 sets of analyses, so only the first set is shown here. These analyses generally reveal a significant main effect of medication (the covariate) on ADHD symptoms and tend toward such an effect on ODD symptoms and social skills. These data suggest that whether participants took medication mediated significant differences in outcomes and that differential medication-taking after the intensive 14-month randomized treatment experience explains at least part of the persisting advantage of the MTA medication algorithm (Comb and MedMgt). Note, however, even after controlling for medication use (both for the 14- to 24-month interim period and during the 30 days before the 24-month assessment), the advantage of Comb/MedMgt over Beh/CC remained significant ($P = .002$) for ADHD symptoms and almost significant ($P = .016$) for ODD symptoms.

Clinical Significance

To measure clinically satisfactory response (near normalization, a "treatment success"), we determined the proportion of children within each treatment group with an item mean ≤1.0 ("just a little") on a composite of parent and teacher SNAP ratings

TABLE 2. 24-Month Outcomes: Symptoms, Functioning, and Services Use

	ADHD Sx (SNAP; Mean [SD]*; n = 526; Lower = Better)		ODD Sx (SNAP; Mean [SD]*; n = 524; Lower = Better)		Social Skills (SSRS Total P&T; Mean [SD]*; n = 522; Higher = Better)		Negative Ineffective Discipline Factor (Mean [SD]; n = 481; More Negative = Better)		Reading (WIAT; Mean [SD]; n = 510; Higher = Better)		Medication Use† (% on Medication)			% Receiving Special Education Services	% Receiving Specialty MH Services	Last Dose (in MPH-Equivalent mg‡; n = 367)
	BL	24 months	*BL*	24 Months	*BL*	24 Months	*BL*	24 Months	*BL*	24 Months	*BL*	14–24 Months	24 Months	14–24 Months	14–24 Months	14–24 Months
Tx group																
Comb	2.01 (.56)	1.17 (.66)	1.34 (.81)	0.83 (.70)	0.94 (.28)	1.17 (.29)	−0.14 (1.68)	−1.45 (1.59)	96.50 (14.61)	98.99 (14.17)	30%	86%	70%	35%	30%	30.43 (14.46)
MedMgt	2.06 (.53)	1.21 (.68)	1.42 (.86)	0.96 (.76)	0.92 (.29)	1.11 (.29)	0.24 (1.81)	−0.86 (1.78)	96.11 (13.82)	99.17 (15.94)	32%	85%	72%	36%	35%	37.47 (17.70)
Beh	2.05 (.56)	1.38 (.69)	1.40 (.79)	1.04 (.81)	0.91 (.26)	1.09 (.28)	0.07 (1.55)	−1.05 (1.64)	95.30 (13.66)	97.84 (13.96)	26%	44%	38%	38%	31%	25.69 (13.03)
CC	2.02 (.58)	1.40 (.68)	1.42 (.80)	1.06 (.79)	0.94 (.28)	1.12 (.28)	−0.07 (1.50)	−0.99 (1.47)	95.15 (14.04)	96.41 (13.29)	34%	69%	62%	41%	34%	24.00 (13.15)
Mixed-effects models or ANCOVAs§ χ^2 or F (P value)												χ^2 Tests Below				ANOVA
Site¶	F = 2.9 (P = .014)		F = 4.35 (P = .0007)		F = 4.08 (P = .001)		F = 0.76 (P = .58)		F = 1.37 (P = .23)							F = 0.73 (P = .60)
Site*Tx	F = 1.44 (P = .13)		F = 1.56 (P = .083)		F = 1.16 (P = .30)		F = 1.56 (P = .08)		F = 1.33 (P = .18)							F = 1.31 (P = .19)
Rater	F = 5.67 (P = .018)		F = 3.08 (P = .08)		F = 9.99 (P = .002)											
Rater means (SD)	Parent 1.08 (.025) Teacher 1.16 (.025)		Parent 1.02 (0.029) Teacher 0.95 (.030)		Parent 1.14 (.012) Teacher 1.09 (.013)											
Rater*Tx	F = 1.38 (P = .25)		F = 0.94 (P = .42)		F = 0.52 (P = .67)						χ^2 Tests‖ Medication Use 14–24 Months			% Receiving Special Education Services	MH Services	
Tx	F = 7.66 (P = .0001)		F = 5.04 (P = .0019)		F = 2.48 (P = .06)		F = 2.60 (P = .05)		F = 1.03 (P = .38)		χ^2 (3df) = 77.35 (P < .0001)			χ^2 = 1.15, (P = .80)	χ^2 = 1.19 (P = .80)	F = 13.15 (P < 0.0001)
Orthogonal contrasts	MTA meds vs not: P = .0001; Comb vs MedMgt: P = .56; Beh vs CC: P = .25		MTA meds vs not: P = .001; Comb vs MedMgt: P = .081; Beh vs CC: P = .41		MTA meds vs not: P = .09; Comb vs MedMgt: P = .05; Beh vs CC: P = .52		MTA meds vs not: P = .17; Comb vs MedMgt; P = .03; Beh vs CC: P = .32				MTA meds vs not: P<.0001; Comb vs MedMgt: P = .56; Beh vs CC: P = .23					MTA meds vs not: P < .0001; Comb vs MedMgt: P = .0013; Beh vs CC: P = .44

Tx indicates treatment; Sx, symptoms; SSRS, Social Skills Rating System; P+T, parent+teacher; WIAT, Wechster Individual Achievement Test; MH, mental health; MPH, methylphenidate; BL, baseline; ANCOVA, analysis of covariance

* The baseline and 24-month item means for the first 3 measures (SNAP and Social Skills) are for average of teacher and parent ratings.

‡ Medication doses last reported during the 14- to 24-month follow-up period for those who took stimulants during that time Comb: 117; MedMgt: 106; Beh: 58; CC: 87), with other stimulants converted to MPH equivalents (eg, 10 mg of d-amphetamine = 20 mg of MPH). Mean doses were not significantly different by site.

‖ Wald χ^2 was used to test the orthogonal contrasts in a logistic regression model for the percentages taking medication 14–24 months and in the 30 days before the 24-month assessment point.

† The medication use percentages on medication show 1) proportion of participants BL on ADHD medication before study entry (not significantly different between groups at BL); 2) proportion of participants with any ADHD medication use between 14 and 24 months (O2 = 77.7; P < .001), and 3) the percentages using medication in the 30 days before the 24-month assessment.

§ Significance level for the mixed-effects regression models and ANCOVAs was set at P = .01 to adjust for 5 analyses. Only the first 3 analyses have dual raters nested within subject. In the absence of dual raters, a standard ANCOVA was performed.

¶ As expected, site differences emerged on 2 measures as a result of differences in local populations. The lack of significant site-by-Tx interaction shows that these did not affect validity of the Tx comparisons. Similarly, the lack of rater-by-Tx interactions shows that rater differences did not affect Tx comparisons. See text for more on rater effects.

Because age at baseline was significantly different between MedMgt and Beh (Table 1), this analysis was repeated with age covaried as a check. It made no practical difference. Covarying age resulted in treatment significance of P = .0003 instead of P = .0001 and rater significance of P = .02, the same as without age covaried. The site effect (P = .13), site x treatment interaction (P = .16), and rater x treatment interaction (P = .32) remained clearly nonsignificant with age covaried.

TABLE 3. Mixed-Effects Models and ANCOVAs of 24-Month Symptomatic and Functional Outcomes With Medication Use Status During 14–24 Months Covaried (See Table 2 for means and SDs)

	ADHD Sx (SNAP)	ODD Sx (SNAP)	Social Skills (SSRS Total P&T)	Negative Ineffective Discipline Factor	Reading (WIAT)
Site	$F = 2.6$ ($P = .025$)	$F = 4.0$ ($P = .001$)†	$F = 3.73$ ($P = .003$)†	$F = 0.74$ (NS)	$F = 1.37$ ($P = .23$)
Site*Tx	$F = 1.48$ ($P = .11$)	$F = 1.57$ ($P = .07$)	$F = 1.19$ ($P = .27$)	$F = 1.56$ ($P = .08$)	$F = 1.32$ ($P = .18$)
Rater	$F = 5.75$ ($P = .017$)	$F = 2.58$ ($P = .109$)	$F = 10.66$ ($P = .001$)		
Rater*Tx	$F = 1.46$ ($P = .23$)	$F = 1.01$ ($P = .38$)	$F = 0.73$ ($P = .54$)		
Tx	$F = 5.13$ ($P = .0017$)	$F = 3.49$ (ns, $P = .016$)	$F = 1.66$ ($P = .18$)	$F = 2.51$ ($P = .06$)	$F = 1.02$ ($P = .38$)
24-mo medication main effect*	$F = 9.98$ ($P = .0017$)	$F = 3.43$ ($P = .065$)	$F = 3.98$ ($P = .047$)	$F = 0.35$ ($P = .56$)	$F = 0.00$ ($P = .99$)
Orthogonal contrasts	MTA meds vs not: $P = .0003$; Comb vs MedMgt: $P = .86$; Beh vs CC: $P = .16$	MTA meds vs not: $P = .008$; Comb vs MedMgt $P = .17$; Beh vs CC: $P = .26$		‡	‡

Significance level for the mixed-effects models and ANCOVAs was set at $P = .01$ to adjust for 5 analyses. NS indicates not significant. The mixed-effects model was used when dual raters were nested within subject; otherwise, a standard ANCOVA was used.
* The interaction of medication status 14–24 mo with Tx group was not significant for any of the 5 clinical outcomes.
† As expected, site differences emerged on 2 measures as a result of differences in local populations. The lack of significant site-by-Tx interaction shows that these did not affect validity of the Tx comparisons. Similarly, the lack of rater-by-Tx interactions shows that the significant rater differences did not affect Tx comparisons. See text for more on rater effects.
‡ Pairwise differences computed only in the presence of a significant Tx main effect

(range: 0–3), a standard achieved by 88% of nonclinical classmates.[4] After consideration of alternative methods of measuring normalization, such as calculating movement toward the normal mean,[25,26] we selected this criterion as most clinically meaningful. Although more stringent than moving half the distance toward the norm, this criterion was attained by 68% of Comb children at 14 months.

The proportion of children with SNAP item means ≤1.0 (near normalization[4] or "excellent responders") at 24 months was 48%, 37%, 32%, and 28%, for Comb, MedMgt, Beh, and CC, respectively. In a logistic regression, the medication algorithm contrast (Comb/MedMgt vs Beh/CC) for these percentages was significant, but the other 2 contrasts were not. More in-depth analyses of normalization will be explored in a subsequent paper.

DISCUSSION

On the 2 symptom measures of the 5 domains assessed, the ITT analyses revealed a persisting superiority for exposure to the conditions that included the MTA MedMgt approach (Comb and MedMgt) over the conditions that did not (Beh and CC), but the effect size was smaller by half at the 24-month than at the 14-month assessment for both the ADHD (effect size = $0.30_{24\text{-month}}$ vs $0.60_{14\text{-month}}$) and ODD (effect size = $0.21_{24\text{-month}}$ vs $0.39_{14\text{-month}}$) outcome measures. The medication algorithm did not significantly affect negative parental discipline, social skills, or academic achievement. Intriguing nonsignificant but consistent trends in numerical superiority of Comb over MedMgt were noted for ODD symptoms, social skills, and parental discipline, as well as in overall rates of normalization, a situation similar to the 14-month outcomes of these domains.

Although the pattern of statistically significant differences across the MTA treatment conditions (Comb∼Med>Beh∼CC) endured, the magnitude of this medication algorithm effect was reduced by approximately half during the 10-month follow-up. The modest size of these persisting group differences on clinical symptoms contrasts sharply with the striking group differences in the use of medication during the follow-up. Most of the Comb and MedMgt participants (85%–86%) continued to receive some form of medication (principally stimulants), whereas fewer Beh and CC participants (44% and 69%, respectively) were medicated during the 14- to 24-month interim. Among those who were medicated, doses were higher in MedMgt and Comb than in Beh and CC (see Table 2). These findings, along with the significant main effect of medication use in the mediator analyses and the loss of significance on ODD symptoms by covarying medication status, suggest that part of the continuing advantage of the MTA Comb and MedMgt treatments is mediated by differential medication use in the 14- to 24-month interim. However, it does not seem entirely mediated by continued medication, because the advantage of the Comb and MedMgt groups for ADHD symptoms withstood statistical control for interim (14–24 months) and endpoint (at the 24-month assessment point) medication use. These analyses suggest that some Comb and MedMgt families may have continued to benefit as a result of their early intensive medication experience, regardless of whether they continued to take medication after 14 months. Findings should be interpreted cautiously, however, because self-selection factors for medication use after 14 months may cause the medication covariate to misstate actual medication effects.[27]

The continued symptom benefit with reduced effect size may have resulted from some participants'

continuing to do well after stopping intensive MedMgt, whereas others deteriorated. An obvious partial explanation for such individual differences in trajectory might come from some participants' stopping medication, a possibility explored in the companion paper. Because medication use does not seem to explain all of the continuing benefit, we should note that not only did the MTA's 0- to 14-month medication approach for the Comb and MedMgt groups result in higher medication doses by 14 months than CC-treated participants (31.2 and 37.7 mg vs 22.8 mg/day in methylphenidate equivalents, respectively),[1,2] but also during its implementation, the medication algorithm included an initial double-blind, placebo-controlled titration (a method to yield optimal symptom reduction and minimal side effects, tailored for each child); performed sequential testing of other medications if the child did not respond well to methylphenidate; provided monthly half-hour office visits with the pharmacotherapist to review caregiver concerns, evaluate progress, and provide advice and support; communicated regularly with the child's teachers via monthly telephone calls by the pharmacotherapist; and readjusted medications if the child was not doing well. Apart from the higher doses produced in Comb and MedMgt participants by the 14-month treatment endpoint, these activities are likely also to have offered supportive benefits to families. Once these nonspecific supports were withdrawn after the MTA treatments ceased, some loss of benefit need not be surprising. Our analyses controlling for 14- to 24-month medication dose do not (and cannot) take these factors into account. Such nonspecific but potentially therapeutic influences can be appreciated visually by inspecting the differences between the 14- and 24-month endpoints (ie, loss of benefit) for Comb and MedMgt participants, a topic of the accompanying report.[28] Nonetheless, research in other fields suggests that such factors can and do function as a type of social support, even providing an incremental part of the therapeutic benefit of an efficacious medication program (via the therapeutic alliance with the physician).[29] In sum, these findings might suggest that some children and families receive maximum medication benefit only when it is accompanied by fairly intensive support and regular contact with their doctor and/or a behavior therapist, whereas others may continue to benefit even after this support is withdrawn.

At follow-up, MedMgt participants' dose levels (in methylphenidate equivalents) were significantly higher than in Comb participants (Table 2). These interesting results suggest the possibility that early Comb interventions might allow reducing overall medication requirements during later periods, consistent with findings that others have reported.[30] This deserves future study as a possible way to keep doses lower to avoid side effects.

Results should also be understood in the context of the actual medications taken by participants in each of the groups over the course of the study. Although level of use of medication differed as a result of the MTA medication algorithm versus no structured approach in the Beh and CC groups, group differences in the types of medications used were minimal. For example, although MedMgt and Comb participants were begun during initial titration on methylphenidate, by 14 months, only 73.4% were being successfully maintained on methylphenidate, 10.4% on dextroamphetamine, 1.4% on pemoline, 1% on imipramine, 0.3% on bupropion, 0.3% on haloperidol, and 3.1% on no medication. Similarly, at 14 months, among those CC participants who were taking medications, some form of stimulant (methylphenidate, pemoline, and amphetamine) composed >89% of medications prescribed, with tricyclics (6%), clonidine/guanfacine (4%), and bupropion (1%) constituting the remainder.[1] At the 24-month follow-up study endpoint, although the groups differed greatly in the proportion taking medication, among those who were taking medication, medication types were fairly evenly distributed across the groups, with stimulants (methylphenidate, amphetamine, and pemoline) composing 88.5% of prescribed medications and tricyclics, bupropion, and clonidine/guanfacine composing 4.5%, 1.2%, and 2.1%, respectively, of other medications prescribed.

Not unexpected, rater differences emerged on 2 of the 3 analyses that had dual informants, with parents rating their children more favorably than teachers on both measures. However, both raters tended to show the same differential rating of the treatments, documented by the absence of significant rater-by-treatment interactions. This supports the validity of the findings across raters, despite differences in raters' perspectives of the child. These findings suggest that it is appropriate and valid to include in the same analysis some participants with parent rating but no teacher rating and others with teacher rating but no parent rating. We propose that this nested mixed-model method be considered for the standard approach to analysis of parent and teacher behavior ratings in ADHD clinical trials. Similar arguments were advanced recently by Kuo et al.[24]

Limitations and Cautions

First, there were slight differences in attrition across groups. Thus, attrition for MedMgt (16 of 144) was marginally ($P < .05$) different from Beh (5 of 144), so the results could be slightly biased by the poorest medication responders dropping out. Such bias is not likely, however, because most of the attrition in MedMgt (13 of 16) occurred immediately after randomization (disappointment in assignment), before drug response could be known. Second, the primary findings are mainly from nonblinded parent and teacher ratings, which could be influenced by biases about treatment. However, any such bias was likely opposite of the findings because Beh fared significantly better than MedMgt on a parent/teacher consumer satisfaction survey at 14 months.[1] Furthermore, the 2 symptom domains with enduring significant effects incorporated ratings from teachers who were not involved with the original treatment, likely unaware of original treatment assignment—essentially blind, although not by design. Third, results should not be interpreted to mean that behavioral therapeutic approaches are ineffective. In fact,

56% of parents of the children in the Beh group left their children unmedicated, suggesting that a substantial proportion perceived their children as functioning well enough to make medication unnecessary (despite recommendations at 14 months for 74% of them to supplement with medication) or at least not worth the perceived trade-offs entailed. Moreover, our orthogonal contrast analyses revealed no significant differences between CC and Beh on any of the 5 outcome domains, even though many more CC participants (69%) than Beh participants (44%) were on medication during the follow-up period, more evidence that concomitant behavioral treatment allows less medication with results at least as good at the group level.

Fourth, failure to find significant incremental effects of the Beh intervention (either by itself or in the Comb intervention) at this first follow-up should not be understood to mean that Beh does not add some advantages under more ideal and ongoing implementation circumstances. Instead, our findings might best be considered in view of the ease with which medication is delivered and continued, in contrast to the greater challenges in maintaining a behavioral intervention. However, examination of the 9-month data relative to 14-month data suggested that the Beh generalization and maintenance procedures with parents had succeeded in maintaining the Beh response after fading of the professional contact at 9 months,[31] and the lack of deterioration in Beh participants from 14 to 24 months suggests that this continued beyond 14 months. If more effective and efficient means to assist families in continuing behavioral interventions become available, then future studies may find even better Beh outcomes.

Last, because the original study was powered only for the .05 significance level and significance in this report was set at .01, the marginally significant (P = .05) persisting differential treatment effect on negative/ineffective discipline, apparently as a result of the multimodality effect (the Comb vs MedMgt orthogonal contrast in Table 2; P = .03) on that measure may be prone to type II error. Consequently, future studies may need to explore further the possible benefits of the Comb over and above intensive MedMgt alone in this and perhaps other areas of functioning.

Clinical Implications

The findings reported here provide new evidence that exposure to the MTA medication algorithm produces long-term beneficial effects on ADHD symptoms, even after families are left to pursue whatever treatments seem to fit them best and the intensive study-delivered treatments have been "handed off" to community physicians. Although definite statistical differences persisted from the original MTA treatment assignment after subsequent medication use was statistically controlled, the clinical significance of the medication algorithm may be modest. Of course, even these modest effects may have substantial public health impact to the extent that widely prevalent chronic conditions may be ameliorated by earlier, intensive interventions. The possibility of larger effects for any of the treated subgroups will be explored in future analyses.

Viewed in the context of our original outcome reports,[1,2] 1 major finding from this report is that some children (particularly those in the MedMgt and Comb groups) lost some of the initial benefits during follow-up. Our companion paper explores this issue, particularly the impact of interim medication use and "switching" between originally assigned treatment groups, as well as relationships between children's subsequent medication use and growth.

ACKNOWLEDGMENTS

This study was supported by UO1 MH50461 (Drs Hinshaw and Elliot), UO1 MH50477 (Drs Conners, Wells, and March), UO1 MH50440 (Drs Swanson, Cantwell, and Wigal), UO1 MH50453 (Drs Abikoff and Hechtman), UO1 MH50454 (Drs Greenhill and Newcorn), and UO1 MH50467 (Drs Pelham and Hoza) from the National Institute of Mental Health (Bethesda, MD).

The National Institute of Mental Health collaborators are Peter S. Jensen, MD (Office of the Director); L. Eugene Arnold, MD, MEd (Department of Psychiatry, Ohio State University); Joanne B. Severe, MS (Biostatistics and Data Management Unit, Division of Services and Intervention Research); Benedetto Vitiello, MD (Child & Adolescent Treatment and Preventive Interventions Research Branch); and Kimberly Hoagwood, PhD (Office of the Director). Principal investigators and coinvestigators from the 6 sites are as follows. University of California, Berkeley/San Francisco: Stephen P. Hinshaw, PhD (Department of Psychology, University of California, Berkeley), Glen R. Elliott, MD, PhD (Department of Psychiatry, University of California, San Francisco); Duke University: C. Keith Conners, PhD, Karen C. Wells, PhD, John March, MD, MPH (Department of Psychiatry & Behavioral Sciences); University of California, Irvine/Los Angeles: James Swanson, PhD (Department of Pediatrics and Cognitive Science, University of California, Irvine), Dennis P. Cantwell, MD, deceased (Department of Psychiatry, Neuropsychiatric Institute, University of California, Los Angeles), Timothy Wigal, PhD (Department of Pediatrics, University of California, Irvine); Long Island Jewish Medical Center/Montreal Children's Hospital: Howard B. Abikoff, PhD (Department of Psychiatry, New York University School of Medicine), Lily Hechtman, MD (Department of Psychiatry, McGill University); New York State Psychiatric Institute/Columbia University/Mount Sinai Medical Center: Laurence L. Greenhill, MD (Department of Psychiatry, Columbia University), Jeffrey H. Newcorn, MD (Department of Psychiatry, Mount Sinai School of Medicine); University of Pittsburgh: William E. Pelham, PhD (Department of Psychology, State University of New York at Buffalo), Betsy Hoza, PhD (Department of Psychological Sciences, Purdue University). Helena C. Kraemer, PhD (Stanford University, Department of Psychiatry & Behavioral Science) is statistical and design consultant. Robert D. Gibbons, PhD (Center for Health Statistics, University of Illinois at Chicago) is statistical consultant for the follow-up. The Office of Special Education Programs/US Department of Education principal collaborator is Ellen Schiller, PhD.

REFERENCES

1. MTA Cooperative Group. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry.* 1999;56:1073–1086
2. MTA Cooperative Group. Moderators and mediators of treatment response for children with attention-deficit /hyperactivity disorder. *Arch Gen Psychiatry.* 1999;56:1088–1096
3. Conners CK, Epstein J, March JS, et al. Multimodal treatment of ADHD (MTA): an overall measure of treatment outcome. *J Am Acad Child Adolesc Psychiatry.* 2001;40:159–167
4. Swanson JM, Kraemer HC, Hinshaw SP, et al. Clinical relevance of the primary findings of the MTA: success rates based on severity of symptoms at the end of treatment. *J Am Acad Child Adolesc Psychiatry.* 2001; 40:168–179
5. Hinshaw SP, Owens EB, Wells KC, et al. Family processes and treatment outcome in the MTA: negative/ineffective parenting practices in

relation to multimodal treatment. *J Abnorm Child Psychol*. 2000;28:555–568
6. Wells KC, Epstein JN, Hinshaw SP, et al. Parenting and family stress treatment outcomes in attention deficit hyperactivity disorder (ADHD): an empirical analysis in the MTA study. *J Abnorm Child Psychol*. 2000;28:543–553
7. Jensen PS, Hinshaw SP, Kraemer HC, et al. ADHD comorbidity findings from the MTA study: comparing comorbid subgroups. *J Am Acad Child Adolesc Psychiatry*. 2001;40:147–158
8. March JS, Swanson JM, Arnold LE, et al. Anxiety as a predictor and outcome variable in the multimodal treatment study of children with ADHD. *J Abnorm Child Psychol*. 2000;28:527–541
9. Paternite CE, Loney J, Salisbury H, Whaley MA. Childhood inattention-overactivity, aggression, and stimulant medication history as predictors of young adult outcomes. *J Child Adolesc Psychopharmacol*. 1999;9:169–184
10. Gillberg C, Melander H, von Knorring AL, et al. Long-term stimulant treatment of children with attention-deficit hyperactivity disorder symptoms. A randomized, double-blind, placebo-controlled trial. *Arch Gen Psychiatry*. 1997;54:857–864
11. Hechtman L, Abikoff H. Multimodal treatment plus stimulants vs. stimulant treatment in ADHD children: results from a two year comparative treatment study. In: *Proceedings of the Annual Meeting of the American Academy of Child and Adolescent Psychiatry*. Washington, DC: American Academy of Child and Adolescent Psychiatry; 1995:63
12. Pelham WE Jr, Fabiano GA. Behavior modification. *Child Adolesc Psychiatr Clin North Am*. 2000;9:671–688
13. Arnold LE, Abikoff HB, Cantwell DP, et al. NIMH collaborative multimodal treatment study of children with ADHD (MTA): design challenges and choices. *Arch Gen Psychiatry*. 1997;54:865–870
14. Hinshaw S, March J, Abikoff H, et al. Comprehensive assessment of childhood ADHD in the context of a multisite, multimodal clinical trial. *J Attention Disord*. 1997;1:217–234
15. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol Bull*. 1987;101:213–232
16. Shaffer D, Fisher P, Dulcan M, et al. The second version of the NIMH Diagnostic Interview Schedule for Children (DISC-2). *J Am Acad Child Adolesc Psychiatry*. 1996;35:865–877
17. Swanson JM. *School-Based Assessments and Interventions for ADD Students*. Irvine, CA: KC Publications; 1992
18. Wechsler D. *Wechsler Individual Achievement Test—Manual*. San Antonio, TX: The Psychological Corporation; 1992
19. Gresham FM, Elliott SN. *Social Skills Rating System—Parent, Teacher, and Child Forms*. Circle Pines, MN: American Guidance Systems; 1989
20. Laird NM, Ware JH. Random effect models for longitudinal data. *Biometrics*. 1982;38:963–974
21. Bryk AS, Raudenbush SW. Toward a more appropriate conceptualization of research on school effects: a three-level hierarchical linear model. *Am J Educ*. 1988;97:68–108
22. Gibbons RD, Hedeker D, Elkin I, et al. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. Application to the NIMH Treatment of Depression Collaborative Research Program dataset. *Arch Gen Psychiatry*. 1993;50:739–750
23. Hedeker D, Gibbons RD, Flay BR. Random-effects regression models for clustered data with an example from smoking prevention research. *J Consult Clin Psychol*. 1994;62:757–765
24. Kuo M, Mohler B, Raudenbush SL, Earls FJ. Assessing exposure to violence using multiple informants: application of hierarchical linear model. *J Child Psychol Psychiatry*. 2000;41:1049–1056
25. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59:12–19
26. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol*. 1999;67:300–307
27. Marcus SM, Gibbons RD. Estimating the efficacy of receiving treatment in randomized clinical trials with noncompliance. *Health Serv Outcomes Res Method*. 2001;2:247–258
28. MTA Cooperative Group. National Institute of Mental Health Multimodal Treatment Study of ADHD follow-up: changes in effectiveness and growth after the end of treatment. *Pediatrics*. 2003;113:762–769
29. Krupnick JL, Sotsky SM, Simmens S, Moyer J. The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol*. 1996;64:532–539
30. Carlson CL, Pelham WE, Milich R, Dixon J. Single and combined effects of methylphenidate and behavior therapy on the classroom performance of children with ADHD. *J Abnorm Child Psychol*. 1992;20:213–231
31. Arnold LE, Chuang S, Davies M, et al. Nine months of multicomponent behavioral treatment for ADHD and effectiveness of MTA fading procedures. *J Abnorm Child Psychol*. 2003;32:39–51

A FAMOUSLY UNNEEDED LESSON

"Many doctors and the drug industry promoted use of hormone therapy a few years ago despite the lack of evidence from clinical trials. But now experts say that experience has taught doctors and society a lesson in the need to exert extreme caution in introducing new therapies."

Altman LK. New study links hormones to cancer risk. *New York Times*. August 8, 2003

Submitted by Student

National Institute of Mental Health Multimodal Treatment Study of ADHD Follow-up: 24-Month Outcomes of Treatment Strategies for Attention-Deficit/Hyperactivity Disorder

MTA Cooperative Group

Pediatrics 2004;113;754-761
DOI: 10.1542/peds.113.4.754