Secondary Evaluations of MTA 36-Month Outcomes: Propensity Score and Growth Mixture Model Analyses

JAMES M. SWANSON, PH.D., STEPHEN P. HINSHAW, PH.D., L. EUGENE ARNOLD, M.D., ROBERT D. GIBBONS, PH.D., SUE MARCUS, PH.D., KWAN HUR, PH.D., PETER S. JENSEN, M.D., BENEDETTO VITIELLO, M.D., HOWARD B. ABIKOFF, PH.D., LAURENCE L. GREENHILL, M.D., LILY HECHTMAN, M.D., WILLIAM E. PELHAM, PH.D., KAREN C. WELLS, PH.D., C. KEITH CONNERS, PH.D., JOHN S. MARCH, M.D., M.P.H., GLEN R. ELLIOTT, PH.D., M.D., JEFFERY N. EPSTEIN, PH.D., KIMBERLY HOAGWOOD, PH.D., BETSY HOZA, PH.D., BROOKE S.G. MOLINA, PH.D., JEFFREY H. NEWCORN, M.D., JOANNE B. SEVERE, M.S., AND TIMOTHY WIGAL, PH.D.

ABSTRACT

Objective: To evaluate two hypotheses: that self-selection bias contributed to lack of medication advantage at the 36-month assessment of the Multimodal Treatment Study of Children With ADHD (MTA) and that overall improvement over time obscured treatment effects in subgroups with different outcome trajectories. **Method:** Propensity score analyses, using baseline characteristics and severity of attention-deficit/hyperactivity disorder symptoms at follow-up, established five subgroups (quintiles) based on tendency to take medication at the 36-month assessment. Growth mixture model (GMM) analyses were performed to identify subgroups (classes) with different patterns of outcome over time. **Results:** All five propensity subgroups showed initial advantage of medication that disappeared by the 36-month assessment. GMM analyses identified heterogeneity of trajectories over time and three classes: class 1 (34% of the MTA sample) with initial small improvement followed by gradual improvement that produced significant medication effects; class 2 (52%) with initial large improvement maintained for 3 years and overrepresentation of cases treated with the MTA Medication Algorithm; and class 3 (14%) with initial large improvement followed by deterioration. **Conclusions:** We failed to confirm the self-selection hypothesis. We found suggestive evidence of residual but not current benefits of assigned medication in class 2 and small current benefits of actual treatment with medication in class 1. *J. Am. Acad. Child Adolesc. Psychiatry*. 2007;46(8):1003–1014. **Key Words:** attention-deficit/hyperactivity disorder, clinical trial, stimulant, behavior therapy, multimodal treatment.

In previous analyses of the initial 14-month outcome (MTA Cooperative Group, 1999a) and persisting 24-month outcome (MTA Cooperative Group, 2004a), we reported a relative advantage of the Multimodal Treatment Study of Children With ADHD (MTA) Medication Algorithm, shown by a greater reduction of

Accepted December 19, 2006.

Please see end of text for author affiliations.

The work reported was supported by cooperative agreement grants and contracts from the National Institute of Mental Health to the following: University of California, Berkeley: U01 MH50461 and N01MH12009; Duke University: U01 MH50477 and N01MH12012; University of California, Irvine: U01 MH50440 and N01MH 12011; Research Foundation for Mental Hygiene (New York State Psychiatric Institute/Columbia University): U01 MH50467 and N01 MH12007; Long Island-Jewish Medical Center U01 MH50453; New York University: N01MH 12004; University of Pittsburgh: U01 MH50467 and N01 MH 12010; and McGill University N01MH12008. The Office of Special Education Programs of the U.S. Department of Education and the Office of Juvenile Justice and Delinquency Prevention of the Justice Department also participated in funding.

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 National Institute of Mental Health, the National Institutes of Health, or the Department of Health and Human Services.

Reprint requests to Dr. James M. Swanson, UCI Child Development Center, 19722 MacArthur Blvd., Irvine, CA 92612; e-mail: jmswanso@uci.edu.

0890-8567/07/4608-1003©2007 by the American Academy of Child and Adolescent Psychiatry.

DOI: 10.1097/CHI.0b013e3180686d63

ADHD symptoms with the medication management (MedMgt) and combined (Comb) conditions than with the behavioral (Beh) and community comparison (CC) conditions [(Comb+Med)-(Beh+CC)]. However, this Medication Algorithm advantage declined over time, and we stated (MTA Cooperative Group, 2004b), "If the apparent differential deterioration…continues, then the outcome of the four randomly assigned MTA groups (Comb, MedMgt, Beh, and CC) will converge over time."

Intent-to-treat (ITT) analyses of the 36-month outcomes confirm this prediction (Jensen et al., 2007) and revealed that current medication use at the 36-month assessment was associated with a slight disadvantage rather than a relative advantage. This led to a self-selection hypothesis that proposes the following: cases with higher compared to lower severity of psychopathology at entry into the MTA or during the MTA follow-up phases would be more likely to have adverse outcomes, the same individuals also would be more likely to receive medication after the initial 14-month intervention period, and the association of severity and long-term medication use would result in selective long-term treatment of the most severe cases, potentially masking beneficial long-term effects of medication.

The ITT analyses also showed a general improvement over time across all randomly assigned treatment groups. A secondary hypothesis is that this trajectory of outcome was heterogeneous within the overall group and proposes that homogeneous subgroups (latent classes) could be identified and evaluated for differential effects of initial treatment assignment and actual use of medication.

The tests of these two hypotheses presented here were intended to clarify and qualify the findings of nonsignificant differences between randomly assigned MTA treatments by the 36-month follow-up. A discussion of the relationship of these findings to the general literature on long-term effects of treatment is provided in the companion paper by Jensen et al. (2007), so it is not repeated here.

The propensity score and growth mixture model (GMM) analytic methods go beyond the ITT methods used in the companion paper by Jensen et al. (2007), and these methods allow the evaluation of continuation of assigned treatment and actual treatment. Of course, these are secondary analyses in the framework on a randomized clinical trial design, which characterized the initial phase of the MTA. These adaptations are necessary as the randomized trial evolved into a naturalistic follow-up study over time. Some limitations of these methods are considered in the final section of this article.

METHOD

Overall Approach

Propensity score analysis is an accepted means of adjusting for selection factors; it was originally described by Rosenbaum and Rubin (1983) (see Marcus and Gibbons, 2001, for review). We used this method to evaluate whether baseline characteristics (including initial symptom severity and continuing severity) of ADHD symptoms during follow-up affected decisions to stop, start, or continue the use of stimulant medication over time, thus masking beneficial medication effects at the 36-month assessment. As in previous analyses, we defined the overall severity of symptoms as the summary or average rating per item (0–3) of the 18 *DSM-IV* symptoms on the Swanson, Nolan, and Pelham (SNAP) rating scale, obtained by averaging across domains (the nine inattention items and the nine hyperactivity-impulsivity items) and across sources (parent and teacher). Using a linear combination of the potentially confounding variables, we formed five subgroups (quintiles) based solely on the likelihood of taking medication at the 36-month assessment. Then, we estimated the relative outcomes of the separate quintiles as well as medication effects separately within each quintile.

Growth mixture model (GMM) analysis is an accepted means for identifying heterogeneity of longitudinal response trajectories (i.e., growth curves or patterns of response over time) to identify subgroups or latent classes. This general method was also used in a companion paper in this issue of the *Journal* (Jensen et al., 2007) to evaluate the 36-month outcome reflected by other measures related to delinquency and substance use. The traditional mixed-effects regression model originally described by Laird and Ware (1982) (see Hedeker and Gibbons, 2006, for a general overview) is based on the null hypothesis that a single longitudinal response process exists and that all of the members of the population systematically deviate from this common temporal response pattern (or growth curve). The alternative hypothesis is that there are two or more different longitudinal response processes in the population, with individual subjects deviating from one of these and outcome better represented by a mixture of temporal response patterns (growth curves) rather than a single growth curve. We used GMM analysis to identify homogeneous subgroups and then evaluated the effects of treatment and other fixed effects within each subgroup (latent class) to determine whether latent class membership was associated with baseline factors, including severity of symptoms and treatment assignment.

Design

To introduce these analytical methods for the evaluation of the long-term observational follow-up phases of the MTA, we focus here on one outcome measure based on a summary of ADHD symptoms based on the SNAP (Swanson, 1992) ratings averaged across domains and sources. This strategy was also used in previous reports (MTA Cooperative Group, 2004a, b). To obtain the