Catching Up on Health Outcomes: The Texas Medication Algorithm Project

T. Michael Kashner, Thomas J. Carmody, Trisha Suppes, A. John Rush, M. Lynn Crismon, Alexander L. Miller, Marcia Toprac, and Madhukar Trivedi

Objective. To develop a statistic measuring the impact of algorithm-driven disease management programs on outcomes for patients with chronic mental illness that allowed for treatment-as-usual controls to "catch up" to early gains of treated patients.
Data Sources/Study Setting. Statistical power was estimated from simulated samples representing effect sizes that grew, remained constant, or declined following an initial improvement. Estimates were based on the Texas Medication Algorithm Project on adult patients (age ≥ 18) with bipolar disorder ($n = 267$) who received care between 1998 and 2000 at 1 of 11 clinics across Texas.
Study Design. Study patients were assessed at baseline and three-month follow-up for a minimum of one year. Program tracks were assigned by clinic.
Data Collection/Extraction Methods. Hierarchical linear modeling was modified to account for declining-effects. Outcomes were based on 30-item Inventory for Depression Symptomatology—Clinician Version.
Principal Findings. Declining-effect analyses had significantly greater power detecting program differences than traditional growth models in constant and declining-effects cases. Bipolar patients with severe depressive symptoms in an algorithm-driven, disease management program reported fewer symptoms after three months, with treatment-as-usual controls "catching up" within one year.
Conclusions. In addition to psychometric properties, data collection design, and power, investigators should consider how outcomes unfold over time when selecting an appropriate statistic to evaluate service interventions. Declining-effect analyses may be applicable to a wide range of treatment and intervention trials.

Key Words. Program evaluation, treatment algorithm, disease management systems, severe mental illness.

In this paper, we developed a new approach, called the Declining-Effects Model, to analyze longitudinal data evaluating a disease management program (DMP) for patients with chronic illness, including mental illness. This approach takes into account how health outcomes may unfold over time by comparing the course of illness between patients assigned to new treatment programs with controls who receive treatment as usual (TAU). The model was

tested using data from the Texas Medication Algorithm Project (TMAP), a DMP for severe mental illness that included consensus-based medication-algorithms, as well as patient education, uniform clinical reports, expert consultation, and clinical coordinators overseeing algorithm adherence (Rush et al. 1999).

Investigators often evaluate DMPs by assigning patients to treatment tracks and repeatedly assessing their outcomes over time, beginning at baseline when treatment begins. Disease management programs are considered effective if the outcomes among treated patients are better than outcomes experienced among controls. Statisticians often summarize these differences by calculating an effect-statistic. While conceptual factors underlying program rationale often influence the choice of a primary outcome measure (McDowell and Newell 1996), the choice of an appropriate effect-statistic typically depends on: (1) the psychometric properties of selected

---

This project was supported, in part, by the Department of Veterans Affairs Health Services Research and Development Research Career Scientist Award (RCS# 92-403) and the Texas Medication Algorithm Project (TMAP) through grants from the Robert Wood Johnson Foundation (#031023), Meadows Foundation (#97040055), the National Institute of Mental Health (#5-R24-MH53799), Mental Health Connections and the Texas Department of Mental Health and Mental Retardation, Nanny Hogan Boyd Charitable Trust, Betty Jo Hay Distinguished Chair in Mental Health, the Rosewood Corporation Chair in Biomedical Science, and the Department of Psychiatry of the University of Texas Southwestern Medical Center at Dallas, and Astra-Zeneca Pharmaceuticals, Abbott Laboratories, Bristol-Myers Squibb Company, Eli Lilly and Company, Forest Laboratories, Glaxo Wellcome, Inc., Janssen Pharmaceutica, Novartis Pharmaceuticals Corporation, Organon Inc., Pfizer, Inc., SmithKline Beecham, U.S. Pharmacopoeia, and Wyeth-Ayerst Laboratories. Opinions expressed herein, however, are those of the authors and do not necessarily reflect the views of supporting agencies and sponsors.

Address correspondence to T. Michael Kashner, Ph.D., J.D., M.P.H., Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9086. Dr. Kashner is Research Career Scientist and Director, Dallas VA Center for Health Services Research, Department of Veterans Affairs, North Texas Health Care System, and Professor, Department of Psychiatry, the University of Texas Southwestern Medical Center at Dallas. Thomas J. Carmody, Ph.D., is an Instructor, Academic Computing Services, the University of Texas Southwestern Medical School at Dallas. Trisha Suppes, M.D., Ph.D., and Madhukar Trivedi, M.D., are Associate Professors, Department of Psychiatry, the University of Texas Southwestern Medical School at Dallas. A. John Rush, M.D., is Professor and Vice-Chair for Research, Betty Jo Hay Distinguished Chair in Mental Health and Rosewood Corporation Chair in Biomedical Sciences, Department of Psychiatry, the University of Texas Southwestern Medical School at Dallas. M. Lynn Crismon, Pharm.D., is Professor and Southwestern Drug Corporation Centennial Fellow in Pharmacy, College of Pharmacy, the University of Texas at Austin. Alexander L. Miller, M.D., is Professor, Department of Psychiatry, the University of Texas Health Science Center at San Antonio. Marcia Toprac, Ph.D., is Director of Research, Texas Department of Mental Health and Mental Retardation, Austin.

outcome measures, (2) the research design, and (3) properties of the statistic itself. These properties include the power of the statistic to avoid falsely detecting effects that do not exist (false positives, or type I error) and failing to find effects that do exist (false negatives, or type II error) (Siegel 1956). Investigators often report the latter as statistical *power*, representing the chance that a statistic would significantly detect an actual effect. All things equal, the most desirable statistic is one with the greatest power for a given type I error.

Not all effect-statistics require working assumptions about how outcomes unfold over time (Lavori 1990). However, to summarize program effectiveness into a single estimate, researchers often borrowed from the efficacy trials literature to select statistics powered to detect effects that *grow* with time. Under a *growth* hypothesis, the outcomes of patients receiving efficacious therapies are expected to improve with time, while their untreated counterparts would remain the same, or get worse. Thus, when differences in outcomes between program tracks grow with time, we say outcomes exhibit an *increasing-effects* pattern.

In this paper, we assert that outcomes of algorithm-driven DMPs for chronic mental illness may be more complex. Specifically, we postulate that the size of an effect may increase by a lump-sum amount that accrues during an initial period following baseline. After such an initial advantage, differences may either remain constant, or decline as DMP versus TAU differences become negligible with time. We call this initial rise, then fall, of a DMP advantage a *declining-effects* pattern.

We thus (1) formulated an effect-statistic that could detect declining-effects patterns; (2) compared the power of both declining-effects and traditional growth statistics to detect an initial effect that either grows, remains constant, or declines with time; and (3) applied the statistic to evaluate an algorithm-driven disease-management program for outpatients with bipolar disorder.

RATIONALE

For Algorithm-Driven Disease Management Programs

The need for algorithm-driven DMPs for chronic mental illness is underscored by the proliferation of medical knowledge and cost-conscious practitioners who often lack time to explore the scientific literature and apply its latest discoveries. Also known as preferred-practices, evidenced-based care, clinical-pathways, or best-practices, clinical algorithms are often presented

as flowcharts designed to help practitioners improve outcomes (Suppes et al. 2001; Field and Lohr 1990; American Psychiatric Association 1995; Jobson and Potter 1995) and contain costs (Lubarsky et al. 1997; La Ruche, Lorougnon, and Digbeu 1995; McFadden et al. 1995) by organizing strategic (what treatments) and tactical (how to treat) decisions into sequential stages (Rush and Prien 1995). Having been applied elsewhere (e.g., Department of Veterans Affairs VHA Directive 96-053, 1996), algorithm-driven DMPs may satisfy concerns first expressed by the late Avedis Donabedian that structured process interventions may be required before results from outcome studies can begin to influence clinical practice (Donabedian 1976).

For Declining-Effects Patterns

There are several reasons why DMP outcomes may follow a declining-effects pattern.

Treatment-as-Usual (TAU). Sometimes ethically required, TAU, rather than no treatment, comparison groups are often used by investigators to answer policy questions raised when new modalities are being considered to replace current practices. If algorithm-driven DMPs help practitioners find the service mix that optimizes outcomes, one may hypothesize that TAU physicians will eventually find the optimum mix, allowing TAU patient outcomes to *catch up* to their DMP counterparts. That is, DMP enhances the speed of an ultimate recovery.

Chronic Iillness. Patients with chronic illness might relapse after therapies improve health, as treatments merely delayed an *inevitable* deterioration in health, or the treatment effects *wore-off* with time.

Derived Outcome. DMPs are intended to improve patient health outcomes by influencing provider behaviors. However, impact on practitioner behaviors may be short lived, or TAU practitioners may adopt the targeted behaviors, causing patient outcomes from the two program tracks to *blend.* The concept of such a *derived* outcome is similar to Grossman's concept of *derived* demand in which use of care springs from consumer wants for health (Grossman 1972).

Texas Medication Algorithm Project (TMAP)

Study data came from the Texas Medication Algorithm Project (TMAP) evaluation of the cost and outcome of a DMP consisting of consensus-based medication algorithms (Crismon et al. 1999), physician training and continued

consultation, standardized chart forms, on-site clinical coordinators provided physician feedback on algorithm adherence and patient progress, and patient education about mental illness and its treatment. The study enrolled 1,421 evaluable patients who received medical treatment in 19 clinics of the Texas Department of Mental Health and Mental Retardation between March 1998 and April 1999 (Suppes et al. 2001; Gilbert et al. 1998; Rush et al. 1999; Rush et al. 1998; Crismon et al. 1999; Chiles et al. 1999; Kashner, Rush, and Altshuler 1999). Patients were long-term sufferers with severe mental illness (major depressive, schizophrenic, or bipolar disorder) in need of a medication change or required medical attention for severe psychiatric symptoms or side effects. Patients were assessed every three months for at least one year for psychiatric symptoms, health functioning, side-effect burden, patient satisfaction, quality of life, and health care costs. The DMP or TAU assignments were by clinic to avoid treatment blending (same doctor treats both intervention and control patients) and water-cooler effects (doctors from the same clinic confer about treating patients).

Program Effects

Growth-Effect Statistics

When outcomes exhibit an increasing-effects pattern, the health of both DMP and TAU patients are assumed to change with time. Outcomes are thus described in terms of a growth-rate. If the rate of change is constant for each patient, the difference in growth-rates between DMP and TAU patients is called the DMP's growth-effect. A DMP is considered effective if it results in significant and favorable growth-effects.

Growth-effect statistics can be calculated from simple growth models, including Pearson's r, Spearman's r, linear slope, and changed scores, with Kendall's τ displaying the greatest power without type I error inflation (Arndt et al. 2000). Kendall's τ is the difference between the numbers of improved and worsened outcomes, divided by either the number of possible comparisons, or comparisons reporting change. A nonparametric statistic applicable to ordinal data, larger τ values indicate growth-effects.

Growth-effects can also be estimated from random regression models that include hierarchical designs such as growth curves (Stanek and Diehl 1988; Willett, Ayoub, and Robinson 1991), random-effects (Laird and Ware 1982; Ware 1985), random regression (Gibbons et al. 1993; Jennrich and Schluchter 1986), empirical Bayesian (Maritz 1970), general mixed linear

(Goldstein 1986), and hierarchical linear (Byrk and Raudenbush 1992) models. These methods are desirable because they do not depend on fixed intervals between observations, allow for missing observations, account for repeated observations nested within patient, permit more flexible covariance structures for a better model fit (Bryk and Raudenbush 1992), such as regression to the mean (Berndt et al. 1998), heteroscedastic, and autocorrelated level-1 covariances (Louis and Spiro 1984; Hedeker 1989; Chi and Reinsel 1989), and has been modified to handle ordinal data (Hedeker and Gibbons 1994).

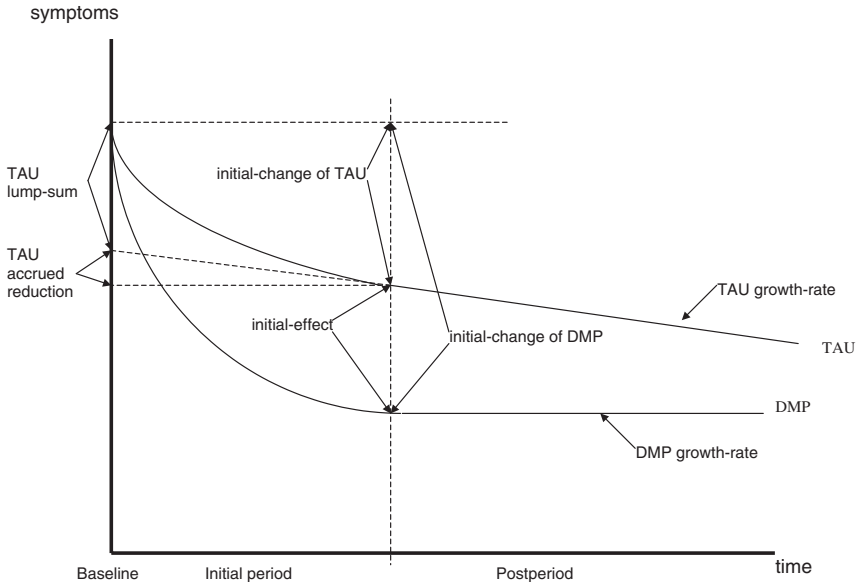
Described in the Appendix, growth-effects estimated from First-Order Growth models are constant whenever differences between program tracks during follow-up grow at a constant rate. With Second-Order Growth models, patient outcomes are expected to grow, but at a diminishing rate with time. Thus, Second-Order Growth models can be used to estimate both a growth-effect, and the rate the growth-effect diminishes with time.

Declining-Effects Statistics

The declining-effects statistics are mathematically derived in the Appendix and displayed in Figure 1 where the *Y*-axis represents symptoms (smaller values indicate better health), and the *X*-axis represents time beginning at baseline and partitioned into initial (first three months) and post (remaining nine months) periods. Under a declining-effects pattern, the health of both DMP and TAU patients are assumed to improve with time at a constant *growth-rate*, with an additional, one-time *lump-sum* improvement accrued during the initial period. We combined the lump-sum improvement and the accumulated growth during the initial period to determine for each patient an *initial-change* in health. The initial-change is calculated by taking the difference in health determined at the end of the initial period and at baseline. Outcomes are thus described in terms of initial-changes and post-period growth-rates.

The impact of a DMP on patient outcomes is assessed by two parameters. The *initial-effect* is the mean difference in initial-changes between DMP and TAU patients. A favorable initial-effect means that DMP patients experienced a more favorable improvement in health between baseline and the end of the initial period than their TAU counterparts. The *growth-effect* is the mean difference in growth-rates between DMP and TAU patients during the postperiod. An unfavorable growth-effect means that the outcomes of TAU patients improve at a faster rate than their DMP counterparts during the postperiod. A favorable initial-effect followed by an unfavorable growth-effect

Figure 1: Declining-Effects Condition



Note: "Y" axis reflects symptom scale with lower values reflecting better health.

indicates a declining-effects pattern of outcomes. Thus, if DMP physicians, aided by algorithms, identify the optimal service mix before their TAU counterparts can determine an effective mix by trial and error, then DMP patient outcomes should be initially better, with TAU patient outcomes catching up with time.

One drawback is the size of the initial-effect would depend on the length of the initial period, requiring investigators to consider how outcomes unfold over time when setting assessment schedules for longitudinal studies. Setting longer initial periods will lead to smaller estimates of initial effects under a "catching-up" TAU.

STATISTICAL POWER

We constructed simulated databases to compare the power of effect-statistics based on Kendall's τ, First and Second-Order Growth Models, and Declining-Effects Models under increasing, constant, and declining-effects conditions.

Health was assessed using the 30-item Inventory for Depression Symptomatology—Clinician Version (IDSC) (Rush et al. 1986; Rush et al. 1996), an interval scale of depressive symptom severity, with well-known psychometric properties, and with larger scores indicating more depressive symptoms and poorer health. The desired statistic has greatest power (likelihood of detecting differences when they existed) without inflating type I error (likelihood of detecting differences when none existed), with power calculated as the proportion of samples detecting significant effects to the total created. To compare models by examining power alone, test statistics were recalibrated to achieve a strict 2.5 percent type I error.

Databases representing each of five test patterns were simulated to include a baseline and four quarterly values for five thousand samples of three hundred subjects each, with a 7 percent random drop-out per quarter, beginning with the second assessment. Data values for each of five patterns were generated, with values deviating from the pattern by independently and normally distributed random variates with zero mean and constant variance of 9.5 IDSC units. Test statistics were based on one-tailed tests, $\alpha = .025$, with standard errors estimated from each drawn sample. Type I errors were determined from a simulated "no effects" and constant-effects cases. See Appendix for equations.

A DMP is considered effective if investigators find a significant growth-effect (Kendal's τ, First and Second-Ordered Growth Models) or a significant initial-effect (Declining-Effects Model). Outcomes follow a declining-effects pattern if the sample reveals a significant initial effect and significant but unfavorable growth-effect (indicating TAU patient outcomes catch up to DMP). Second-Order Growth Model can be adopted to test for declining-effect patterns. See Appendix.

Table 1 contains a description of five test patterns and power estimates, by model, effect, and test pattern. All five test patterns had the same initial-change in outcomes for DMP (-6 IDSC), for TAU (-2 IDSC), and thus the same initial-effect (–4 IDSC). Pattern 1 represents an increasing effect size (-2 IDSC/qtr), resulting from a DMP growth-rate (-3 IDSC/qtr) that was more favorable than TAU (-1 IDSC/qtr). Pattern 2 represents an increasing effect size (-2 IDSC/qtr) that diminishes each quarter (+.50 IDSC/qtr). Pattern 3 represents a constant effect size (0 IDSC/qtr). Patterns 4 and 5 represent declining-effects of 1/qtr and 1.33/qtr, respectively. Technical details are provided in the Appendix.

Results are reported in Table 1. Detecting if DMP had any impact on outcomes of care, effect-statistics derived from First-Order Growth Models

Table 1: Power for Selected Statistics for Selected Conditions Based on One-Tailed t-tests of 300 Subjects

	PATTERN 1 Increasing Effects plus Lump-sum		*PATTERN 2 Increasing Effects that Diminish*		*PATTERN 3 Constant Effects*		*PATTERN 4 Declining Effects (Small)*		*PATTERN 5 Declining Effects (Large)*	
Pattern Characteristics										
	DMP	TAU	DMP	TAU	DMP	TAU	DMP	TAU	DMP	TAU
Growth-rate (IDSC/qtr)	−3.00	−1.00	−6.75	−2.25	−1.00	−1.00	0.00	−1.00	0.33	−1.00
Diminish growth-rate (IDSC/qtr^2)	0.00	0.00	0.75	0.25	0.00	0.00	0.00	0.00	0.00	0.00
Lump-sum (IDSC)	−3.00	−1.00	0.00	0.00	−5.00	−1.00	−6.00	−1.00	−6.33	−1.00
Initial-change (IDSC)	−6.00	−2.00	−6.00	−2.00	−6.00	−2.00	−6.00	−2.00	−6.00	−2.00
Growth-effect (IDSC/qtr)	−2.00		−4.50		0.00		1.00		1.33	
Initial-effect (IDSC)	−4.00		−4.00		−4.00		−4.00		−4.00	
Effect Statistic										
Kendal's τ										
Growth-effect	81.92%		84.56%		2.18%		0.00%		0.00%	
First-Order Growth Model										
Growth-effect	100.00%		100.00%		86.56%		33.92%		18.40%	
Second-Order Growth Model										
Growth-effect	77.76%		92.50%		67.80%		60.46%		59.26%	
Diminishing growth-effect	16.62%		36.10%		50.10%		64.98%		71.50%	
Declining-effect					(2.50%)		15.72%		25.86%	
Declining-Effects Model										
Initial-effect	76.98%		83.56%		77.12%		75.84%		76.48%	
Negative growth-effect	0.00%		0.00%		2.64%		48.06%		72.86%	
Declining-effect					(2.50%)		42.62%		62.26%	

Technical details are provided in the Appendix.

(growth-effects) were superior under an increasing-effects pattern (1–2), while Declining-Effects Models (initial-effects) exhibited the greatest power under a decreasing-effects pattern (4–5). When outcomes exhibited a constant-effects pattern (3), effect-statistics from a Declining-Effects Model had better power than from a Second-Order Growth Model to detect if DMP had any favorable impact on health outcomes (77 percent versus 68 percent). While the single parameter growth-effect estimate from the First-Order Growth Model did display better power (87 percent) than the initial-effect estimate from the Declining-Effects Model (77 percent), the former model is misspecified, and leads to an incorrect interpretation that effect sizes grow throughout follow-up, when in fact the effect size remains constant after a one-time lump-sum increase accruing during an initial period. Thus, for both constant and declining-effects conditions, a Declining-Effects Model is the best choice to derive an effects-statistic. Note that under constant and declining-effect conditions, Kendal's τ was virtually powerless to detect any program differences.

The Declining-Effects Model was also superior to Second-Order Growth Models in detecting declining-effect patterns, a power gap that tended to widen as effect sizes declined at a faster rate. That is, the Declining-Effects Model provides the more appropriate test to determine if TAU caught up with their DMP counterparts after DMP gained an advantage during the initial period.

The results from these simulations meant that in the absence of theory-driven hypotheses, exploratory analyses should begin with a test for declining-effects patterns using Declining-Effects Models. In the absence of declining- or constant-effects patterns, investigators should proceed with Growth Models. To rely on Growth Models when declining or constant effects exist, investigators risk either having insufficient power to detect any program advantages, or risk incorrectly interpreting the data to suggest that group differences continue to grow throughout follow-up.

Detecting Declining-Effect Conditions: TMAP

Declining-Effects Models are relevant only if there is empirical support for declining-effects patterns in nature. To test for such patterns, we use TMAP data for 267 adult patients with a Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnosis of bipolar disorder (American Psychiatric Association 1994). All subjects had signed informed consent,

provided baseline and at least one follow-up IDSC assessment, and had been assigned to one of 4 ALGO clinics ($n = 141$) or 7 TAU clinics ($n = 126$). The sample was diverse ethnically (27 percent Hispanic, 12 percent African American, 60 percent Caucasian, and 1 percent other), with a mean age of 39.0 years (SD = 10.3), and 68 percent female, 24 percent married, 26 percent living alone with household sizes averaging 1.7 plus patient (SD = 1.8), were high school graduates (78 percent), but mostly unemployed (73.6 percent) with mean monthly household disposable income (after housing costs) of $450/month (SD = $599) (1999 US$), with income assistance or food-stamps (49 percent) and on Medicaid during six months prior to baseline (53 percent). Clinically, patients at baseline reported a mean length of illness of 14.7 years (SD = 12.5), IDSC score of 30.9 (SD = 14.5), 24-item Brief Psychiatric Rating Scale ($BPRS_{24}$) (Overall and Gorham 1988; Ventura et al. 1993) of 52.8 (SD = 13.5), and Medical Outcome Study 12-item Short Form (SF-12) (Ware, Kosinski, and Keller 1996) mental functioning of 35.0 (SD = 11.3) and physical functioning of 42.6 (SD = 11.6), with 20 percent reporting substance abuse during the prior six months. A 10-item Patient Perception of Benefits scale was constructed for TMAP to measure patient attitudes concerning whether health care can improve patient functioning. Patients were asked "If I can get the help I need from a doctor, I believe that I will be much better able to …manage problems at home, earn a living or go to school, enjoy things that interest me, feel good about myself, handle emergencies and crises, get along with friends, get along with my family, control my life, do things on my own, and make important decisions that affect my life and my family." Patients ranked each item as strongly agree (= 1) to strongly disagree (= 5). For this sample, baseline scores ranged from 10 to 50, with mean 19.1 (SD = 7.6), with higher scores indicating greater pessimism about care.

Declining-Effects Models (equations 8b and 9b in Appendix) were estimated using HLM/3L software (Bryk, Raudenbush, and Congdon 1996) with subjects grouped by baseline symptoms into very severe ($IDSC \geq 46$, $n = 48$), severe (IDSC 17–45, $n = 170$), and mild ($IDSC \leq 16$, $n = 46$). Estimates were adjusted to reflect patient differences in baseline $BPRS_{24}$ score, age (years), family size, disposable income, years of education, patient perception of benefits, gender, African American status, and Hispanic status.

Results are reported in Table 2. Examining only those subjects who began the study with the most severe depressive symptoms at baseline, patients in both TAU and DMP tracks experienced a reduction in symptoms during the initial three months of the program by –6 IDSC and –16.5 IDSC, respectively. The greater reduction in symptoms during the initial period

Table 2: IDSC Adjusted Estimates of Initial-Change, Growth-Rate, and Differences between DMP and TAU Patients with Bipolar Disorder, by Baseline IDSC Scores

	TAU (n = 123)				*DMP (n = 141)*				*DIFFERENCE*			
	γ	*df*	*t*	*P*	γ	*df*	*t*	*P*	γ	*df*	*t*	*P*
Initial-Change												
Very severe	−6.00±2.79	33	2.15	.039	−16.54±2.46	33	6.73	<.001	−10.54±4.05	165	2.60	.010
Severe	−2.59±1.42	137	1.82	.069	−3.72±1.32	137	2.82	.005	−1.13±1.99	603	0.57	.57
Mild	4.73±1.63	32	2.91	.007	5.81±1.86	32	3.13	.004	1.08±2.74	170	0.39	.69
Growth-Rate												
Very severe	−3.18±0.92	165	3.45	.001	0.58±0.59	165	0.98	.33	3.75±1.10	165	3.43	.001
Severe	0.45±0.46	603	0.97	.33	−0.19±0.34	603	0.56	.58	−0.64±0.57	603	1.12	.27
Mild	−0.34±0.42	170	0.81	.42	−0.95±0.33	170	2.88	.004	−0.61±0.53	170	1.15	.25

Coefficients represented as mean value ± standard error.

Adjusted for mean values by group (all, high, medium, and low clinical depressive symptoms) with respect to baseline $BPRS_{24}$ score, age (years), family size, disposable income, years of education, patient perception of benefits from treatment, gender, African American status, Hispanic status.

Depression symptom subsamples created by mean ± 1 standard deviation IDSC scores. Sample sizes are for *very severe* (IDSC ≥ 46) DMP $n = 24$, TAU $n = 24$; *severe* (IDSC 17-45) DMP $n = 89$, TAU $n = 81$; and *mild* (IDSC ≤ 16) DMP $n = 28$, TAU $n = 18$.

among DMP patients (DMP initial-effect) was significant (-10.5 IDSC), indicating a more favorable DMP outcome than TAU at the end of the initial period. During the postperiod, however, TAU patients continued to improve at a growth-rate of –3.2 IDSC/qtr, while symptoms for DMP patients were unchanged. The difference in growth-rates was significant (DMP growth-effect) suggesting that TAU patients caught up with their DMP counterparts during the postperiod as IDSC outcome differences between tracks diminished by 3.7 IDSC/qtr.

In contrast, patients who entered the study with severe baseline depressive symptoms showed only modest reductions in symptoms after the initial period (–3.7 IDSC for DMP and –2.6 IDSC for TAU), though DMP and TAU differences were not significant. On the other hand, patients with mild symptoms actually showed a worsening of symptoms at the end of the initial period as both DMP and TAU patients regressed into depressive symptom episodes. An apparent inverse relationship between initial reductions and baseline symptoms suggest a regression to the mean confound. The impact on DMP initial- and growth-effects is small, as baseline IDSC values were 30.3 ± 14.5 for DMP and 31.5 ± 14.5 for TAU, or a mean difference of -1.2 ± 1.8 ($t = 0.67$, $df = 261$, $p = .51$).

In summary, DMP versus TAU exhibited a declining-effects pattern for only the very severely depressed patients.

CONCLUSION

Investigators often borrow from the clinical trials literature applying test statistics (Kendall's τ, growth models) to determine program effectiveness by detecting if the effect sizes grow with time. This is appropriate when treatments under study "cure" debilitating illnesses that, if left untreated, would grow worse.

However, a growth hypothesis may not be reasonable in intervention trials focusing on provider behaviors, chronic conditions, and comparisons with TAU controls. Examples include the TMAP algorithm-driven DMP for patients with mental disorders designed to help clinicians find the right mix of available treatments and assist compliance through patient education.

Pattern of outcomes may deviate from a simple growth. Disease management program practitioners may revert to former methods as TAU practitioners accept new ones, causing practice patterns to blend, and

outcomes to become similar. TAU practitioners may eventually prescribe an appropriate mix of services that may eventually work, helping TAU outcomes to catch up to their DMP counterparts, though more time is required to reach desired levels. On the other hand, treatments for intractable chronic conditions may eventually wear off, patients may stop care due to intolerable side-effects, or poor prognoses may mean that a DMP merely delays an inevitable decline in health.

In such cases, an intervention could exhibit significant improvement during an initial period, with the advantage declining in subsequent periods.

Our analyses of TMAP data suggested that declining-effects patterns do exist when algorithm-driven DMPs are evaluated against TAU controls. Declining-effects also appear in work therapy programs for homeless, substance dependent veterans (Kashner, Rosenheck et al. 2002), and in acute randomized controlled trials whenever placebos impact outcomes (Hrobjartsson and Gotzsche 2001; Andrews 2001; Kirsch 2000; Lavori 2000).

Our power calculations suggested that under declining-effects patterns, growth models are inappropriate. They either lack power to detect program effects, or misspecify how program advantages unfolded over time. We recommend investigators begin with a test for declining-effects patterns before proceeding with growth models when evaluating DMP for patients with chronic conditions.

In summary, in addition to the psychometric properties of outcome measures, the research design, and the power and type II error of the statistic itself, services researchers should carefully consider how outcomes are expected to unfold over time before selecting an appropriate program effect statistic.

APPENDIX: INCREASING-EFFECTS MODEL

A First-Order Growth Model is:

$$y_{it} = \alpha_0 + \alpha_{01} I_i + \alpha_{10} t + \alpha_{11} I_i t + u_i + v_{it}, \tag{1a}$$

where y_{it} is health status assessed at time t from baseline ($t = 0$) when patient i was assigned to either treated ($I_i = 1$) or controls ($I_i = 0$). Patient and time level random effects are u_i and v_{it}. The growth-effect is α_{11}, or difference in growth-rates in outcome between treated ($a_{10} + a_{11}$) and control (α_{10}) patients. For symptom outcomes (fewer symptoms reflect better health), a favorable

growth-effect suggests $\alpha_{11}<0$. α_{01} is the average baseline difference between treated $(\alpha_0+\alpha_{01})$ and controls (α_0).

A Second-Order Growth Model allowing growth-rates to change with time and adjusting for baseline covariates is:

$$y_{it} = \alpha_0 + \alpha_{01} I_i + \alpha_{10} t + \alpha_{11} I_i t + \alpha_{20} t^2 + \alpha_{21} I_i t^2 + \alpha_3 (X_{it} - \bar{X}) + u_i + v_{it}, \qquad (1b)$$

where α_{21} is the rate of change in growth-rates over time between treated $(\alpha_{20}+\alpha_{21})$ and controls (α_{20}). How health varies as patient characteristics (X_{it}) differ from a reference, or average, patient $(\bar{X})$ is described by α_3.

Derivation of Declining-Effects Model

Declining-Effects Models compare health outcomes of patients assigned to a disease management program (DMP) with those assigned to treatment-as-usual (TAU). Let y_{it} be a ratio scale measuring health status for patient i assessed at time t following baseline $(t\geq 0)$. Health status is assumed to be a linear function of: (1) patient characteristics $(\beta_0,)$, (2) a growth-rate $(\beta_1,)$, (3) a one-time, lump-sum change in outcome (β_2) occurring at t^0 following baseline, and (4) patient (u_i) and time (v_{it}) level random variates, or:

$$y_{it} = \beta_0 + \beta_1 t + \beta_2 B_t + u_i + v_{it}, \qquad (2)$$

where B_t is a step function that equals one whenever $t\geq t^0$, and zero otherwise, with random variates assumed to be identically distributed with zero mean and constant variance: $\mathrm{COV}(u_i, u_j) = 0$ for $i\neq j$; and $\mathrm{COV}(v_{is}, v_{it}) = 0$ for $s\neq t$.

Describing the patient effect, let:

$$\beta_0 = \beta_{00} + \beta_{01}(X_{it} - \bar{X}), \qquad (2a)$$

where β_{00} equals baseline health for the average patient with characteristics, $\bar{X}$; and the vector β_{01} is the rate health varies for patients at time t whose characteristics, X_{it}, differ from the average patient.

Describing the growth-effect, let:

$$\beta_1 = \beta_{10} + \beta_{11}(X_{it} - \bar{X}), \qquad (2b)$$

where β_{10} is the change in health per unit time for the average patient and the parameter vector β_{11} describes how the growth-rate varies for patients at time t whose characteristics differ from the average patient. Equation 2b can be expanded to include nonlinear time trends.

Describing the lump-sum effect, let:

$$\beta_2 = \beta_{20} + \beta_{21}(X_{it} - \bar{X}), \qquad (2c)$$

where β_{20} is the lump-sum change for the average patient and the parameter vector β_{21} describes how the lump-sum varies for patients whose characteristics differ from the average patient.

Finally, all parameters may be a function of which service system protocol patients were assigned:

$$\beta_{jk} = \beta_{jk0} + \beta_{jk1} I_i \quad for\, j = 0, 1, 2; \quad and \quad k = 0, 1; \tag{2d}$$

where I_i is a dichotomous variable that assumes the value of 1 if the patient were assigned to DMP services protocol, and = 0 for TAU, with β_{jk1} as the difference between DMP and TAU. Alternatively, β_{jk1} can be described as a rate when I_i is measured as a continuous variable describing provider adherence and/or patient compliance with the service protocol.

Balancing model parsimony with misspecification error, we: (1) assumed growth-rates were unchanged over relatively short observation periods; (2) focused on how lump-sum and growth-rates vary by treatment group; and (3) focused on lump-sum changes that were allowed to vary with patient characteristics. Thus:

$$\begin{aligned} y_{it} =& \beta_{000} + \beta_{001} I_i + \beta_{01}\left(X_{it} - \bar{X}\right) + \beta_{100} t + \beta_{101} I_i t \\ &+ \beta_{200} B_t + \beta_{201} I_i B_t + \beta_{21} B_t \left(X_{it} - \bar{X}\right) + u_i + v_{it}. \end{aligned} \tag{3}$$

where, β_{000} is the average baseline value for TAU patients; β_{001} is the average adjusted difference in baseline values between DMP and TAU patients; β_{100} is the growth-rate calculated for TAU patients; β_{101} is the adjusted difference in growth-rates between DMP and TAU; β_{200} is the one-time lump-sum change in outcomes for the average TAU patient; and β_{201} is the adjusted difference in lump-sum changes in outcomes between DMP and TAU.

Equation 3 was simplified by substituting change scores for raw health values. Common in clinical trials literature, change scores are calculated by subtracting baseline values from each follow-up assessment. Setting $t = 0$ and $B_{t=0} = 0$, equation 3 becomes:

$$y_{it=0} = \beta_{000} + \beta_{001} I_i + \beta_{01}\left(X_{it} - \bar{X}\right) + u_i + v_{it=0}. \tag{4}$$

Subtracting equation 4 from 3, assuming the first assessment was administered after the lump-sum change ($B_{t \geq t^0} = 1$), and recalibrating time so the first assessment occurs at $t = 1$, then:

$$\Delta y_{it} = \beta_{100} t + \beta_{101} I_i t + \beta_{200} + \beta_{201} I_i + \beta_{21}\left(X_{it} - \bar{X}\right) - v_{it=0} + v_{it0}, \tag{5}$$

with change scores $\Delta y_{it} = y_{it} - y_{it=0}$. Rearranging terms and substituting $u^*_i = -v_{it=0}$, equation 5 is reconfigured as:

$$\Delta y_{it} = \beta_{200} + \beta_{201} I_i + \beta_{100} t + \beta_{101} I_i t + \beta_{21}\left(X_{it} - \bar{X}\right) + u_i^* + v_{it}, \quad \textit{for } t > 0. \tag{6}$$

Equation 6 is a more familiar random regression equation, but with change scores replacing raw values for the dependent variable. Setting $t = 1$ in equation 6, the average initial-change in outcomes between baseline $(t = 0)$ and the first assessment $(t = 1)$ can be computed by:

$$\Delta y_{it=1} = \beta_{200} + \beta_{201} I_i + \beta_{100} + \beta_{101} I_i. \tag{7}$$

for DMP $(I_i = 1)$ and for TAU $(I_i = 0)$ patients. The initial-change between baseline and the first assessment equals a lump-sum change $(\beta_{200} + \beta_{201} I_i)$ plus an accrued growth-rate during the first period $(\beta_{100} + \beta_{101} I_i)$. Adding and subtracting equation 7 from equation 6 yields:

$$\begin{aligned} \Delta y_{it} = &(\beta_{200} + \beta_{100}) + (\beta_{201} + \beta_{101}) I_i + \beta_{100}[t-1] + \beta_{101} I_i [t-1] \\ &+ \beta_{21}\left(X_{it} - \bar{X}\right) + u_i^* + v_{it}. \end{aligned} \tag{8a}$$

or

$$\Delta y_{it} = \gamma_0 + \gamma_1 I_i + \gamma_2 [t-1] + \gamma_3 I_i [t-1] + \gamma_4\left(X_{it} - \bar{X}\right) + u_i^* + v_{it}, \tag{8b}$$

with TAU initial-change (γ_0), TAU growth-rate (γ_2), initial-effect (γ_1), growth-effect (γ_3), the impact on change scores when patient characteristics differ from an average patient (γ_4), where $\gamma_0 = \beta_{200} + \beta_{100}$; $\gamma_2 = \beta_{100}$; $\gamma_1 = \beta_{201} + \beta_{101}$; $\gamma_3 = \beta_{101}$; and $\gamma_4 = \beta_{21}$, and u_i^* and v_{it} as patient and time-level random variates. Note, the model can be modified to account for bivariate and ordinal value outcome measures. The DMP initial-change and growth-rate can be calculated directly by substituting:

$$I_i = 1 - I_i^*. \tag{9a}$$

into Eq. 8b to obtain:

$$\Delta y_{it} = \gamma_0^* + \gamma_1^* I_i^* + \gamma_2^* [t-1] + \gamma_3^* I_i^* [t-1] + \gamma_4^*\left(X_{it} - \bar{X}\right) + u_i^* + v_{it}, \tag{9b}$$

with DMP initial-change (γ_0^*) and growth-rate (γ_2^*), and identical but opposite effect size estimates $(\gamma_1 = -\gamma_1^*)$, $(\gamma_3 = -\gamma_3^*)$.

Effects are said to be *declining* if following an initial-effect $(\gamma_1 < 0)$, TAU has a more favorable growth-rate, $\gamma_2 < \gamma_2^*$; or $\gamma_2^* - \gamma_2 = \gamma_3 > 0$. The TAU is *catching up* if the declining-effects were the result of TAU patients improving $(\gamma_2 < 0)$ while DMP patients got no worse during the postperiod $(\gamma_2 < \gamma_2^* \leq 0)$. On the other hand, both DMP and TAU patients are said to face an *inevitable decline* if symptoms return to patients in both groups during the postperiod $(0 < \gamma_2 < \gamma_2^*)$. Effects are *constant* if following an initial-effect $(\gamma_1 < 0)$, DMP and

TAU growth-rates are equal ($\gamma_2^* = \gamma_2$ or $\gamma_3 = 0$). Effects are *increasing* if DMP patients continue to improve at a faster rate than TAU ($\gamma_3 < 0$).

DETAILS OF SIMULATION MODELS

Let $B_{it} = 1$ for all patient i with $t < 1$ and $B_{it} = 0$ for $t \geq 1$. Data were generated to represent different test patterns including: [#0] No-effects: $y_{it} = 50 + u_i + v_{it}$; [#1] Increasing-effects with Lump-sum: $y_{it} = 50 - t - 2I_i t - B_{it} - 2I_i B_{it} + u_i + v_{it}$; [#2] Increasing-effects that diminish: $y_{it} = 50 - 2.25t - 4.5I_i t + 0.25\, t^2 + 0.5\, I_i t^2 + u_i + v_{it}$; [#3] Constant-effects: $y_{it} = 50 - t - B_{it} - 4I_i B_{it} + u_i + v_{it}$; [#4] Small declining-effects: $y_{it} = 50 - t + 1_t I_i - B_{it} - 5I_i B_{it} + u_i + v_{it}$; and [#5] Declining-effects: $y_{it} = 50 - t + 1.33tI_i - B_{it} - 5.33I_i B_{it} + u_i + v_{it}$.

Models testing for initial-, declining- and growth-effects included: (A) First-Order Growth: $\Delta y_{it} = at + btI_i$ (growth-effect: $b < 0$); (B) Second-Order Growth: $\Delta y_{it} = at + btI_i + ct^2 + dt^2 I_i$ (growth-effect: $b < 0$; diminishing growth-effect: $d > 0$; declining-effects: $b < 0$ and $b + 8d > 0$); and (C) Declining-Effect: $\Delta y_{it} = e + fI_i + g\,[t - 1] + hI_i[t - 1]$ (initial-effect: $f < 0$; growth-effect: $h > 0$; and declining-effects: $f < 0$ and $h > 0$).

REFERENCES

American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV). Washington, DC: American Psychiatric Association.

———. 1995. *American Psychiatric Association Practice Guideline for Psychiatric Evaluation of Adults*. Washington, DC: American Psychiatric Association.

Andrews, G. 2001. "Placebo Response in Depression: Bane of Research, Boon to Therapy." *British Journal of Psychiatry* 170 (3): 192–194.

Arndt, S., C. Turvey, W. H. Coryell, J. D. Dawson, A. C. Leon, and H. S. Akiskal. 2000. "Charting Patients' Course: a Comparison of Statistics Used to Summarize Patient Course in Longitudinal and Repeated Measures Studies." *Journal of Psychiatric Research* 34 (2): 105–13.

Berndt, E. R., S. N. Finkelstein, P. E. Greenberg, R. H. Howland, A. Keith, A. J. Rush, J. Russell, and M. B. Keller. 1998. "Workplace Performance Effects from Chronic Depression and its Treatment." *Journal of Health Economics* 17 (5): 511–35.

Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical Linear Models*. Newbury Park, CA: Sage.

Bryk, A. S., S. W. Raudenbush, and R. T. CongdonJr. 1996. *Hierarchical Linear and Nonlinear Modeling, with HLM/2l and HLM/3L Programs*, (release 4). Chicago: Scientific Software International.

Chi, E. M., and G. C. Reinsel. 1989. "Models of Longitudinal Data with Random Effects and AR(1) Errors." *Journal of the American Statistical Association* 84 (406): 452–9.

Chiles, J. A., A. L. Miller, M. L. Crismon, A. J. Rush, A. S. Krasnoff, and S. S. Shon. 1999. "The Texas Medication Algorithm Project: Development and Implementation of the Schizophrenia Algorithm." *Psychiatric Services* 50 (1): 69–74.

Crismon, M. L., M. Trivedi, T. A. Pigott, A. J. Rush, M. A. Hirschfeld, D. A. Kahn, C. DeBattista, J. C. Nelson, A. A. Nierenberg, H. A. Sackeim, M. E., and Thasethe Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder. 1999. "The Texas Medication Algorithm Project: Report of the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder." *Journal of Clinical Psychiatry* 60 (3): 142–56.

Donabedian, A. 1976. *Benefits in Medical Care Programs.* Cambridge, MA: Harvard University Press.

Field, M. J., and K. N. Lohr. 1990. *Clinical Practice Guidelines: Directions for a New Program. U.S. Institute of Medicine Committee to Advise the Public Health Service on Clinical Practice Guidelines.* Washington, DC: U.S. Department of Health and Human Services, National Academy Press.

Gibbons, R. D., D. Hedeker, I. Elkin, C. Waternaux, H. C. Kraemer, J. B. Greenhouse, T. Shea, S. D. Imber, S. M. Sotsky, and J. T. Watkins. 1993. "Some Conceptual and Statistical Issues in Analysis of Longitudinal Psychiatric Data: Application to the NIMH Treatment of Depression Collaborative Research Program Dataset." *Archives General Psychiatry* 50 (9): 739–50.

Gilbert, D. A., K. Z. Altshuler, W. V. Rago, S. P. Shon, M. L. Crismon, M. G. Toprac, and A. J. Rush. 1998. "Texas Medication Algorithm Project: Definitions, Rationale and Methods to Develop Medication Algorithms." *Journal of Clinical Psychiatry* 59 (7): 345–51.

Goldstein, H. I. 1986. "Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares." *Biometrika* 73 (1): 43–56.

Grossman, M. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55.

Hedeker, D. 1989. "Random Regression Models with Autocorrelated Errors." Doctoral dissertation, University of Chicago.

Hedeker, D., and R. D. Gibbons. 1994. "A Random-Effects Ordinal Regression Model for Multilevel Analysis." *Biometrics* 50 (4): 933–44.

Hrobjartsson, A., and P. C. Gotzsche. 2001. "Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo with No Treatment." *New England Journal of Medicine* 344 (21): 1594–1601.

Jennrich, R. I., and M. D. Schluchter. 1986. "Unbalanced Repeated Measures Models with Structured Covariance Matrices." *Biometrics* 42 (4): 805–20.

Jobson, K. O., and W. Z. Potter. 1995. "International Psychopharmacology Algorithm Project Report: Introduction." *Psychopharmacology Bulletin* 31 (3): 457–9.

Kashner, T. M., R. Rosenheck, A. B. Campinell, A. Suris, R. Crandall, N. J. Garfield, P. Lapuc, K. Pyrcz, T. Soyka, and A. Wicker 2002. "Impact of Work Therapy on Health Status among Homeless, Substance Dependent Veterans: A Randomized Controlled Trial." *Archives of General Psychiatry* 59 (10): 938–44.

Kashner, T. M., A. J. Rush, and K. Z. Altshuler. 1999. "Measuring Costs of Guideline-Driven Mental Health Care: The Texas Medication Algorithm Project." *Journal of Mental Health Policy and Economics* 2 (3): 111–21.

Kirsch, I. 2000. "Are Drug and Placebo Effects in Depression Additive?" *Biological Psychiatry* 47 (8): 733–5.

Laird, N. M., and H. Ware. 1982. "Random-Effects Models for Longitudinal Data." *Biometrics* 38 (4): 963–74.

La Ruche, G., F. Lorougnon, and N. Digbeu. 1995. "Therapeutic Algorithms for the Management of Sexually Transmitted Diseases at the Peripheral Level in Côte d'Ivoire: Assessment of Efficacy and Cost." *Bulletin of the World Health Organization* 73 (3): 305–13.

Lavori, P. W. 1990. "ANOVA, MANOVA, and My Black Hen." *Archives of General Psychiatry* 47 (8): 775–8.

———. 2000. "Placebo Control Groups in Randomized Treatment Trials: A Statistician's Perspective." *Biological Psychiatry* 47 (8): 717–23.

Louis, T. A., and A. Spiro III. 1984. "Fitting First Order Auto-regressive Models with Covariates (Technical Report)." Cambridge, MA: Harvard University School of Public Health, Department of Biostatistics.

Lubarsky, D. A., P. S. A. Glass, B. Ginsberg, G. de L. Dear, M. E. Dentz, T. J. Gan, I. C. Sanderson, M. G. Mythen, S. Dufore, C. Pressley, W. C. Gilbert, W. D. White, M. L. Alexander, R. L. Coleman, M. Rogers, and J. G. Reves. 1997. "The Successful Implementation of Pharmaceutical Practice Guidelines: Analysis of Associated Outcomes and Cost Savings." *Anesthesiology* 86 (5): 1145–60.

McDowell, I., and C. Newell. 1996. *Measuring Health: A Guide to Rating Scales and Questionnaires.* New York: Oxford University Press.

McFadden Jr, E. R., N. Elsanadi, L. Dixon, M. Takacs, E. C. Deal, K. K. Boyd, B. K. Idemoto, L. A. Broseman, J. Panuska, T. Hammons, B. Smith, F. Caruso, C. B. McFadden, L. Shoemaker, E. L. Warren, R. Hejal, L. Strauss, and I. A. Gilbert. 1995. "Protocol Therapy for Acute Asthma: Therapeutic Benefits and Cost Savings." *American Journal of Medicine* 99 (6): 651–61.

Maritz, T. S. 1970. *Empirical Bayes Methods.* London: Methuen.

Overall, J. E., and D. R. Gorham. 1988. "Introduction—The Brief Psychiatric Rating Scale (BPRS): Recent Developments in Ascertainment and Scaling." *Psychopharmacological Bulletin* 24 (1): 97–9.

Rush, A. J., M. L. Crismon, M. Toprac, M. H. Trivedi, W. V. Rago, S. Shon, and K. Z. Altshuler. 1998. "Consensus Guidelines in the Treatment of Major Depressive Disorder." *Journal of Clinical Psychiatry* 59 (20, supplement): 73–84.

Rush, A. J., D. E. Giles, M. A. Schlesser, C. L. Fulton, J. E. Weissenburger, and C. T. Burns. 1986. "The Inventory of Depressive Symptomatology (IDS): Preliminary Findings." *Psychiatry Research* 18 (1): 65–87.

Rush, A. J., C. M. Gullion, M. R. Basco, R. B. Jarrett, and M. H. Trivedi. 1996. "The Inventory of Depressive Symptomatology (IDS): Psychometric Properties." *Psychological Medicine* 26 (3): 477–86.

Rush, A. J., and R. F. Prien. 1995. "From Scientific Knowledge to the Clinical Practice of Psychopharmacology: Can the Gap Be Bridged?" *Psychopharmacology Bulletin* 31 (1): 7–20.

Rush, A. J., W. V. Rago, M. L. Crismon, M. G. Toprac, S. P. Shon, T. Suppes, A. L. Miller, M. H. Trivedi, A. I. C. Swann, M. M. Biggs, K. Shores-Wilson, T. M. Kashner, T. Pigott, J. A. Chiles, D. A. Gilbert, and K. Z. Altshuler. 1999. "Medication Treatment for the Severely and Persistently Mentally Ill: The Texas Medication Algorithm Project." *Journal of Clinical Psychiatry* 60 (5): 284–91.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

Stanek, E. J. III, and S. R. Diehl. 1988. "Growth Curve Models of Repeated Binary Response." *Biometrics* 44 (4): 973–83.

Suppes, T., A. Swann, E. B. Dennehy, E. Habermacher, M. Mason, L. Crismon, M. Toprac, A. J. Rush, S. Shon, and K. Z. Altshuler. 2001. "Texas Medication Algorithm Project: Development and Feasibility Testing of a Treatment Algorithm for Patients with Bipolar Disorder." *Journal of Clinical Psychiatry* 62 (6): 439–47.

Ventura, J., M. F. Green, A. Shaner, and R. P. Liberman. 1993. "Training and Quality Assurance with the Brief Psychiatric Rating Scale: 'The Drift Busters'." *International Journal of Methods for Psychiatric Research* 3 (4): 221–44.

VHA Directive 96-053. 1996. Washington, DC: Department of Veterans Affairs, Veterans Health Administration.

Ware, J. E., M. Kosinski, and S. D. Keller. 1996. "A 12-item Short-form Health Survey (SF-12): Construction of Scales and Preliminary Tests of Reliability and Validity." *Medical Care* 34 (3): 220–33.

Ware, J. H. 1985. "Linear Models for the Analysis of Longitudinal Studies." *American Statistician* 39 (2): 95–101.

Willett, J. B., C. C. Ayoub, and D. Robinson. 1991. "Using Growth Modeling to Examine Systematic Differences in Growth: An Example of Change in the Functioning of Families at Risk of Maladaptive Parenting, Child Abuse, or Neglect." *Journal of Consulting Clinical Psychology* 59 (1): 38–47.