The Journal of Mental Health Policy and Economics
J. Mental Health Policy Econ. **2**, 111–121 (1999)

Measuring Costs of Guideline-Driven Mental Health Care: The Texas Medication Algorithm Project

T. Michael Kashner[1,2]*, A. John Rush[1] and Kenneth Z. Altshuler[1]

[1]*Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA*
[2]*Dallas VA Center for Health Services Research, Dallas, TX, USA*

Abstract

Background: Algorithms describe clinical choices to treat a specific disorder. To many, algorithms serve as important tools helping practitioners make informed choices about how best to treat patients, achieving better outcomes more quickly and at a lower cost. Appearing as flow charts and decision trees, algorithms are developed during consensus conferences by leading experts who explore the latest scientific evidence to describe optimal treatment for each disorder. Despite a focus on 'optimal' care, there has been little discussion in the literature concerning how costs should be defined and measured in the context of algorithm-based practices.
Aims of the study: This paper describes the strategy to measure costs for the Texas Medication Algorithm project, or TMAP. Launched by the Texas Department of Mental Health and Mental Retardation and the University of Texas Southwestern Medical Center at Dallas, this multi-site study investigates outcomes and costs of medication algorithms for bipolar disorder, schizophrenia and depression.
Methods: To balance costs with outcomes, we turned to cost-effectiveness analyses as a framework to define and measure costs. Alternative strategies (cost–benefit, cost–utility, cost-of-illness) were inappropriate since algorithms are not intended to guide resource allocation across different diseases or between health- and non-health-related commodities. 'Costs' are operationalized consistent with the framework presented by the United States Public Health Service Panel on Cost Effectiveness in Medicine.

Patient specific costs are calculated by multiplying patient units of use by a unit cost, and summing over all service categories. Outpatient services are counted by procedures. Inpatient services are counted by days classified into diagnosis groups. Utilization information is derived from patient self-reports, medical charts and administrative file sources. Unit costs are computed by payer source. Finally, hierarchical modeling is used to describe how costs and effectiveness differ between algorithm-based and treatment-as-usual practices.
Discussion: Cost estimates of algorithm-based practices should (i) measure opportunity costs, (ii) employ structured data collection methods, (iii) profile patient use of both mental health and general medical providers and (iv) reflect costs by payer status in different economic environments.
Implication for health care provision and use: Algorithms may help guide clinicians, their patients and third party payers to rely on the latest scientific evidence to make treatment choices that balance costs with outcomes.
Implication for health policies: Planners should consider consumer wants and economic costs when developing and testing new clinical algorithms.
Implications for further research: Future studies may wish to consider similar methods to estimate costs in evaluating algorithm-based practices.

Received 6 April 1999; accepted 16 August 1999.

Mental health clinical practice guidelines, protocols, preferred practices, clinical pathways, and algorithms generally embrace a concept of optimal care that balances patient outcomes with costs.[1] Thus, it is not surprising to see clinicians applying guidelines to put the latest medical discoveries into practice;[2–5] at the same time managed care organizations use guidelines to contain health care costs.[6,7] Despite this emphasis on costs, studies developing and evaluating guideline-driven practices have focused little attention on how these costs should be measured.[8–10] The issue is exacerbated by a health economic literature that supports different approaches to measuring costs in applications that usually focus on general medical patients. Such a literature ignores problems of homelessness, criminal

*Correspondence to: T. Michael Kashner, Ph.D., J.D., Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235-9086, USA.
Tel: (214) 648-4608.
Fax: (214) 648-4612.
Source of Funding
Contract grant sponsor: Robert Wood Johnson Foundation
Contract grant number: 031023
Contract grant sponsor: Meadow Foundation
Contract grant number: 97040055
Contract grant sponsor: National Institute of Mental Health
Contract grant number: MH53799
Contract grant sponsor: Bristol-Myers Squibb Company
Contract grant sponsor: Eli Lilly and Company
Contract grant sponsor: Glaxo Wellcome, Inc.
Contract grant sponsor: Janssen Pharmaceutica
Contract grant sponsor: Novartis Pharmaceutical Corporation
Contract grant sponsor: Pfizer, Inc.
Contract grant sponsor: Wyeth-Ayerst Laboratories
Contract grant sponsor: Mental Health Connections
Contract grant sponsor: Betty Jo Hay Distinguished Chair in Mental Health
Contract grant sponsor: Rosewood Corporation Chair in Biomedical Science
Contract grant sponsor: Department of Veterans Affairs Health Services Research and Development Research Career Scientist Award
Contract grant number: RCS 92-403

CCC 1091–4358/99/030111–11$17.50

incarceration, frequent use of non-specialty providers and complex and inconsistent health care financing confronting many patients with severe and persistent mental illnesses.

The goal of this discussion paper is to explain to administrators, policy makers, researchers and the health professional how costs may be calculated under guideline-driven care for patients with severe mental illness. Specifically, we describe how investigators with the Texas Medication Algorithm Project (TMAP) are applying the Utilization and Cost (UAC)[11] methodology to measure the cost of guideline-driven practices to treat schizophrenia and major depressive and bipolar disorders. TMAP is a public–academic collaborative effort to develop, implement and evaluate medication algorithms to treat patients at 20 state-supported mental health clinics across Texas.[12–14,5,15] These facilities are affiliated with the Texas Department of Mental Health and Mental Retardation. TMAP medication algorithms are intended to aid clinical decision-making by organizing strategic (what treatments) and tactical (how to treat) decisions into sequential stages.[16] Presented as flowcharts, these decision trees help clinicians consider patient status and responses to prior treatment to prioritize among medication options. Investigators plan to evaluate TMAP by comparing patients whose care follows a medication algorithm with those following usual care during a 12–24 month follow-up period.

We begin by describing *what* costs should be measured when developing and evaluating mental health clinical guidelines. This is done by describing cost–outcome studies as an appropriate framework to evaluate practice guidelines as optimal care. We also give a rationale for limiting calculations to direct health care costs, while expanding the service scope to include services from both mental health specialty and general medical care providers. Next, we describe *how* costs should be measured specifying data requirements. Finally, we describe analytic methods to link changes in costs to algorithm adherence.

What Costs Should Be Measured

Clinical practice guidelines are the result of experts applying the latest medical knowledge to chart the course of treatment that either optimizes patient health outcomes (outcome driven) or achieves the biggest 'bang for the buck' (value driven). Since the latter agenda requires balancing health outcomes against costs, we use the cost–outcome study as the appropriate framework to measure costs and evaluate guideline-driven clinical programs.

Frameworks for cost–outcome studies vary in the literature. Between 1991 and 1996, the medical literature contained some 3500 cost–outcome publications,[17] with methods cited in economics[18] and human resources;[19,20] environmental regulation;[21] pharmacy;[22] clinical decision making[23] and management;[24] public health[25] including health services,[26] medical care quality[27] and program evaluation[28] and clinical trials.[29] Cost–outcome methods have also been summarized by medical specialty, including medicine,[30,31] oncology,[32] cardiology[33] and mental health.[34–37] These studies vary with respect to (i) how outcomes are defined, (ii) what costs are included and (iii) how costs and outcomes are compared.

Outcome Measures

Choice of outcome measures for cost–outcome studies include *effects* (symptoms, health functioning, satisfaction), *benefits*[18] (economic consequences measured in dollars), *utility*[38] (health states weighted by consumer preferences), and quality adjusted life years, or QALYS[39] (life years adjusted for morbidity due to illness). TMAP uses effectiveness measures that are administered during face-to-face interviews at baseline and 3 month follow-ups. Instruments assess (i) psychiatric symptoms including the Scale for Assessing Negative Symptoms,[40] the Brief Psychiatric Rating Scale,[41,42] the Calgary Depression Scale,[43] the Clinician Administered Rating Scale for Mania,[44] the Internal State Scale for manic and depressive symptoms[45] and Clinical and Self-Reported versions of the Inventory of Depressive Symptomatology;[46,47] (ii) side-effects including Systematic Assessment for Treatment Emergent Events[48] and the Barnes Rating Scale for Drug-Induced Akathisia;[49] (iii) functioning and quality of life including the 12-Item Short-Form Health Survey, or SF-12,[50] the Lehman Work and Productive Activity Subscale[51,52] and the Quality of Life Scale[53] and (iv) medication compliance including the Medication Compliance Scale.[54]

Defining Costs

Cost–outcome studies define costs as the societal value of alternative uses of expended resources, or 'opportunity costs'. Studies do vary, however, with respect to what elements are costed. All investigators include some measure of *direct* cost for diagnosing, treating and assessing patients, and the cost of accessing these services. Other costs are *indirect* (work loss, absenteeism, reduced productivity) and *intangible* (pain, suffering) associated with morbidity and premature mortality.

TMAP investigators measure only direct costs that encompass all of the patient's care, with sums calculated for each payer source. This is consistent with the US Public Health Service sponsored Panel on Cost-Effectiveness in Health and Medicine.[55–58] Considered as program outcomes, changes in indirect costs are treated on the 'benefits' side of the cost–benefit scale, and thus are excluded from the cost calculus to avoid double counting.[59] Global health care includes all payer perspectives (providers, third party payers and patient out of pocket), service perspectives (care for schizophrenia, major depressive and bipolar disorders, other psychiatric, addiction and general medical conditions), and provider perspectives (Texas state mental health clinics, other mental health and addiction treatment providers and general medical providers). Our global emphasis on all of the patient's health care providers is designed to capture costs incurred by patients who, in an effort to avoid algorithm restrictions, seek care from non-algorithm and non-specialty providers.

Our focus on the cost of both mental and general medical services, as recommended by the panel, is particularly compelling in light of empirical studies evidencing an association between use of general medical and mental health care. First, patients frequently seek general medical providers to care for psychiatric problems.[60–62] Second, the presence of psychiatric symptoms tend to be associated with more medical care costs.[63–69] Third, patients in clinical environments who use specialty mental health care tend to use fewer general medical care services, including outpatient visits,[70] inpatient days[71] and peak use of primary care services,[72] and at lower general medical costs.[73,74] Fourth, clinical observers report reductions in use of general medical care following use of mental health care,[75–77] treatment for undiagnosed panic disorder[78] and psychotherapy,[79] while others have reported no change,[63] or an increase[80] in general medical care following use of mental health services. Finally, the effectiveness of general medical care may be influenced by psychiatric status.[81] For example, psychiatric status has been observed to be related to mortality rates among the elderly with hip fractures,[82,83] and myocardial infarction.[84]

Finally, economists generally prefer to report costs in terms of a total burden on society. While important for social policy decisions, such information would do little to motivate individual providers to adhere to, patients to comply with or third parties to finance care based on, clinical guidelines. Thus, we calculate payer-specific costs to describe how the cost burden is shared among patients, third party payers and the health care providers.

Comparing Costs and Outcomes

There is disagreement in the literature concerning how costs and outcomes are to be compared. In the simple case, the choice as to whether to accept or reject the TMAP algorithm as a preferred practice becomes clear whenever guideline adherence is shown to lead to lower (higher) costs and better (worse) outcomes. The choice is unclear whenever either protocol or treatment-as-usual achieves a better outcome, but at higher direct treatment costs. Finally, the two choices are equivalent if both result in the same outcome and incur similar costs. The decision rule is summarized in **Table 1**.

Whenever the choice is unclear, two cost–outcome statistics are useful to policy makers to judge the value of guideline-driven practices. The cost–outcome 'ratio of differences' statistic is based on the formulation of Jerrell and Hu[85] and will measure how much additional outcome is produced for each additional health care dollar patients in an algorithm based practice incur over patients in usual care. For randomly assigned patients, the statistic is computed by

$$\text{Cost–outcome ratio of differences} = \frac{\bar{o}_A - \bar{o}_U}{\bar{c}_A - \bar{c}_U} \qquad (1a)$$

where $\bar{o}_A$ and $\bar{o}_U$ are average outcomes and $\bar{c}_A$ and $\bar{c}_U$ are average costs for algorithm-based and treatment-as-usual patients, respectively. This statistic is intended to help policy makers balance the algorithm's better health outcomes with anticipated increases in treatment costs.

Most scientific-based algorithms are expected to put more, rather than fewer, clinical resources into caring for the patient than what is currently available in traditional care. Thus, better outcomes from algorithm-based practices may be the result of more care, rather than from more efficacious treatment. The cost–outcome 'difference in ratios' statistic measures the extent, if any, to which algorithms will yield a greater outcome, dollar for dollar, than treatment-as-usual. Calculations are based on the Siegel[29] revitalization of Grossman's health production model[86] in which health care costs are treated as inputs (independent variable) to produce health outcomes as outputs (dependent variable). Let $o_A(c)$ and $o_U(c)$ be mathematical functions describing the relationship between health care costs 'c' for patients assigned to algorithm and treatment-as-usual, respectively, on health outcomes 'o'. The term 'marginal productivity' refers to the change in outcome associated with an addiction dollar invested in health care under a given treatment protocol. At a given initial investment in health care of c_0, marginal productivity may be computed as: $[o(c_0 + \$1) - o(c_0)]/\$1$. The 'difference in ratios' cost–outcome statistic equals the difference in marginal productivity between algorithm and treatment-as-usual patients when both groups have received an equivalent dollar investment in health care, or

$$\text{Cost–outcome difference in ratios} = \frac{[o_A(c_0 + \$1) - o_A(c_0)]}{\$1} - \frac{[o_U(c_0 + \$1) - o_U(c_0)]}{\$1} \qquad (1b)$$

Table 1. Policy choices when comparing guideline-driven clinical practice (ALGO) versus treatment-as-usual (TAU)

Outcome	Costs		
	higher	no difference	lower
better	*unclear*	ALGO	ALGO
no change	TAU	*equivalent*	ALGO
worse	TAU	TAU	*unclear*

Measuring Direct Health Care Costs

We compute direct health care costs with respect to a given payer source. Computationally, we multiply a patient-specific use rate for each health service by a payer-specific unit cost, and sum over all services. Mathematically, if TC_{ip} is the total cost to care for patient i with respect to payer p, then

$$TC_{ip} = u_{1i}c_{1p} + u_{2i}c_{2p} + \ldots + u_{ji}c_{jp} + \ldots u_{ni}c_{np} \qquad (2)$$

with u_{ji} as the quantity of service j that patient i consumed at a cost c_{jp} per unit to payer p. For this study, payer groups include study patients, their local providers and third party payers and an all-payer class to represent global costs. Overlapping estimates prevent calculating global costs by merely summing payer-specific totals.

The strategy relies on (i) a classification of health services into categories of care, $j = 1, 2, \ldots, n$, that are mutually exclusive, exhaustive and homogenous with respect to costs, (ii) payer-specific unit cost schedules $\{c_{1p}, c_{2p}, \ldots, c_{jp}, \ldots, c_{np}\}$, that reflect cost of services to appropriate payer groups and (iii) patient-specific utilization rates, $\{u_{1i}, u_{2i}, \ldots, u_{ji}, \ldots, u_{ni}\}$, determined from hybrid use of care data that integrates available sources of information to create a comprehensive profile encompassing all of the patient's health care providers.

Classification of Health Services

We measure utilization by counting the number of outpatient procedures by physicians' *Current Procedural Terminology* codes[87] (CPT), inpatient days by Diagnosis Related Groupings[88] (DRG) and prescription fills by National Drug Code. Other services are measured as days by facility type (e.g., long term psychiatric care, nursing home, addiction rehabilitation, domiciliary care, transitional residence), day-treatment days and home care visits.

These systems were selected because they (i) were designed to classify clinical activities into homogenous categories with respect to the cost of production and (ii) are standards for medical billing.

Outpatient procedures under a guideline-driven protocol are expected to cost differently from those under treatment-as-usual. We thus distinguish between procedures produced under an algorithm and treatment-as-usual by a two-digit extension to the five-digit CPT code. Thus, 'psychiatric interviews' become *CPT#90820-xy* and 'medication management' *CPT#90862-xy*, with '*x*' representing primary psychiatric diagnosis ('S' schizophrenia, 'B' bipolar, 'D' depression) and '*y*' representing protocol adherence (0, complete adherence; 1, partial adherence; 2, no adherence).

Assessing Unit Costs

Unit cost schedules are determined for each payer source and contain estimates for each CPT outpatient procedure, DRG inpatient day and medication. To provide the information needs of local sponsors, we price TMAP care based on the cost environments of local payer groups. However, the methodology allows us to model how estimates would change if the study were performed in other regions of the country and clinical settings, and with a different mix of players.

Payer groups include practitioners serving study patients who incur costs to provide care, including tax-supported county and state hospitals and clinics and the medical centers with the US Department of Veterans Affairs, as well as community providers and facilities. Third party payers include welfare (Texas Medicaid) and US social security (Medicare) programs helping patients pay for health care, as well as private health plans serving study patients. Consumer payments include patient out-of-pocket expenses incurred to access Texas state mental health facilities, other mental health and general medical providers and patient cost shares calculated for Medicare and other health plans servicing TMAP study patients.

Unit costs are calculated by multiplying a service-specific weight, or relativity, by a payer-specific conversion factor. The relativity, wj, depends only on the service category j and not on the payer source. Thus, a service with a relativity of 2 would cost twice as much as a service with a relativity of 1. A payer-specific conversation factor, F_p, depends only on the identity of the payer p and is the payer's cost for services with a relativity of one. The unit cost, c_{jp}, to payer p for service j may be calculated by

$$c_{jp} = w_j F_p \,. \tag{3}$$

For outpatient care, relativities for CPT procedures are taken from the total work, practice and malpractice components of the Resource Based Relative Value Units (RBRVU) of Health Care Financing Administration's Medicare Physician Fee Schedule.[89,90] These relativities were scientifically calculated to reflect time, skill and experience required to perform each procedure. For inpatient care, *per diem* relativities by DRGs are computed by dividing the DRG weight for the entire inpatient stay by a national average length of stay for the respective DRG. Data come from sampled participants of the US Medicare program, as published by the Health Care Financing Administration.[91] Finally, patient out-of-pocket expenses by outpatient visit and inpatient admission are computed from TMAP patient survey responses.

Conversion factors for inpatient and for outpatient services are determined separately for each payer. For government-supported institutions, conversion factors are computed by adding their annual services budget plus imputed rents for space and depreciation for equipment, and dividing by total relativities of all reported services. For community providers, conversion factors are computed by multiplying a cost to charge ratio to the 50th or 80th percentile of billed charges converted to a 'charge per relativity'. A billed charge per relativity is calculated by dividing the total charge by the total relativity for all services itemized in each bill. Charges are translated into costs by a 'cost to charge' factor. The US Health Care Financing Administration uses a 72% cost to charge ratio when calculating hospital payments from charge data under its Medicare program for elderly citizens.[91,92] Ancillary services have higher ratios: 80% for radiology, 85% for laboratory, and 120% for pharmacy.[92] Conversion factors for third-party payers, including government programs (welfare or Medicaid, social security or Medicare), private insurance companies and health plans, can be determined directly from the third party's reimbursement rates.

Relativities should differ between psychiatric procedures in algorithm-based clinics and those in treatment-as-usual facilities. To account for these differences, we estimate a relativity for each CPT procedure in algorithm-based practices, with Health Care Financing Administration relativities applicable for treatment-as-usual practices. Relativity estimates are based on encounter data obtained from computerized records maintained by the Texas state mental

health facilities. For each non-algorithm-based encounter, medical procedures listed in the data are translated into CPT codes, assigned an RBRVU and the corresponding relativities summed to compute a total relativity for the encounter. Total physician time is also computed for the encounter. The association between the encounter's physician time (independent variable) and total relativity (dependent variable) is computed using ordinary least squares regression for a total relativity production model:

$$R_e = \beta_0 + \beta_1 \ (\text{MDtime})_e + \beta_2 \ (\text{MDtime})_e^2 + v_e \qquad (4)$$

where for each encounter *e*, R_e is the total RBRVU relativity and MDtime_e is total physician time. β are coefficients. v is a normally independent and identically distributed random variate with zero mean and constant variance. The model's fit gauges the accuracy of this strategy. The estimated model may then be used to compute a relativity for algorithm-based and CPT-coded procedures by entering average physician time determined to produce algorithm-based procedures into the estimated regression model. That is, compared to usual care, algorithm-based relativities are higher or lower depending on physician time inputs.

The methodology has been applied to the cost accounting and clinical databases for the Department of Veterans Affairs health care system.[93] Total costs for clinical care, exclusive of research and education programs, were computed for 1 October 1996 through 30 March 1997. Estimates were based on actual salaries paid for professional, support and administrative staff, purchased supplies, paid building maintenance and utilities and estimates of depreciation of building and equipment calculated as a straight-line depreciation of the actual purchase price over the expected life of the item. Data came from computerized cost accounting and clinical files housed at the Department of Veterans Affairs centralized computer in Austin, TX.[94] For the six-month study period, data were obtained from outpatient departments of 147 medical centers that produced nearly 35.9 million CPT procedures with an average RBRVU per procedure of 1.4. Dividing the cost to generate these procedures into the total RBRVU (35.9 million procedures × 1.4 RBRVU/procedure) yielded a conversion factor of $36.88 per RBRVU for Veterans medical facility costs. This compares with $36.6873 national conversion factor for physician payments under the US Medicare program for 1998.[89] Thus, the unit cost incurred by the Veterans medical center to produce a psychiatric interview (CPT #90801) may be computed by $36.88/RBRVU × 3.27 RBRVU for CPT90801, or $120.60 per interview.

The stability of physician time inputs as predictors of outpatient productivity, measured in terms of total relativities, has been tested across outpatient clinics at 146 US Veterans medical facilities.[93] In this example, the unit of analysis is the clinic level, rather than encounter-level relativities recommended for the TMAP project. Our clinic-level analyses were dictated by data that reported quarterly physician time by facility. MD time is measured as an FTE, or full time equivalent, equaling an average assigned work time of 2080 hours per year, minus sick time and vacation leave. Results are reported in **Table 2**. The estimated model suggests that one additional physician hour will produce on average 6.2 RBRVU procedures in psychiatry. This is sufficient to generate 1.7 psychiatric interviews, 1.8 diagnostic interviews, three 20-minute psychotherapy sessions or 4.5 medication management sessions. The high percent of explained variance ($R^2 = .640$) reveals how important physician time is to production of medical procedures, even when using aggregated clinic data that include the services of psychiatric nurses, physician assistants, social workers and psychologists.

This strategy has several advantages. First, differences between algorithm-based and treatment-as-usual practice costs are reflected by changes in the mix of medication prescribed and health services consumed as well as from differences in physician time inputs. However, this method will not reflect differences in the intensity of non-physician inputs. Second, estimating unit costs by payer source improves external validity that permits policy makers to better understand the cost implications of guideline-driven practices across consumers, providers and third party payers. By applying other unit cost schedules, the methodology permits investigators to estimate how TMAP may cost in other economic environments. Third, we rely on microcosting studies, cost accounting databases and market rates to profile a best estimate of health care costs. By separating previously determined service-specific relativities from payer-specific conversion factors, investigators can calculate complete unit cost schedules for each payer even though not all procedures or patient types are well represented in the data.

Measuring Use of Care

Data to compute patient use of *all* health care come from different sources. *Medical charts* are considered by many to be a gold standard, though their completeness and accuracy have not gone unquestioned.[95–97] Computerized *administrative files* from providers, payers and managers are popular in the literature because they are inexpensive to access[98–103] but otherwise present problems with[104,105] (i) incomplete data, (ii) non-reporting of patient 'out-of-plan' use, (iii) non-standardized definitions and collection procedures making inter-file merges difficult,[106] (iv) licensing laws that complicate medical file transfers[107] and (v)

Table 2. Estimated coefficients for a total relativity production model based on resource based relative value scale for outpatient psychiatry and addictions treatment clinics at VA medical centers from 1 October 1996 through 30 March 1997[93]

	Beta coefficient	Standard error	*t*-statistic	$p<$
Constant	7 954.572	6859.609	1.160	.248
(MD time)	12 785.941	2239.535	5.709	.000
(MD time) squared	–62.744	135.570	–0.463	.644
$R^2 = .640$				
$n = 146$				

uncertain data quality.[108] On the positive side, recent studies suggest that the billing function and performance evaluation purposes of administrative files may lead some systems to include checks and balances to ensure data accuracy and completeness.[109,110] Finally, *consumer surveys* are well known among epidemiologists who measure national use rates (National Institute of Mental Health Epidemiologic Catchment Area study,[111] the National Health Interview Survey,[112] the Medical Outcome Study[113] and the National Medical Care Expenditure Surveys[114,115]). Surveys have been the only method available to collect data from many hard-to-reach groups, such as the homeless.[116–118] The accuracy of self-reports, however, will vary with the (i) saliency of the medical event (painful, interferes with lifestyle, life threatening, recent, more frequent),[119–123] (ii) intensity of care including longer hospital stays[124] or surgery,[125] (iii) characteristics of the patient, with underreporting likely for older[126] patients and for those with low annual incomes,[119] (iv) characteristics of the interviewer,[127,128] (v) respondent's expectations and interviewer attitudes and styles[129,130] and (vi) presence of collateral family members or friends.[131] On the other hand, recall error[132] and motivational and cognitive states[133] may create special problems when interviewing patients with mental illness. Estimates of data validity vary, with studies suggesting that surveys underreport,[134–137] overreport[11,138] and have modest agreement[133,139] with provider accounts of patient use of mental health care.

To account for patients' total use of care, UAC constructs use-of-care profiles based on multiple, or hybrid, sources of information. The concept is not new. Utilization informatioon has been constructed by replacing patient self-reports whenever unavailable or unreliable with information taken from collateral interviews of family members and friends,[131] combining patient self-reports and provider records to construct a 'best profile',[140] 'hybrid'[141] or 'best estimate',[138] combining multiple claims databases[142] and merging automated medical record and chart-review data.[143]

Figure 1 shows that UAC extends these approaches by constructing utilization databases from abstracts of written medical and billing records obtained from the patient's health care providers. If reliable records for a specific provider are unavailable, we turn to administrative files assembled from the patient's third party payers (Texas Medicaid, Federal Medicare programs, private health plans). In the absence of administrative files, the information is taken directly from face-to-face interviews with the patient (UAC Questionnaire).

The process begins when each TMAP patient signs informed consent permitting investigators to access patient medical records. Next, a search is conducted to pull patient event files from computerized administrative databases maintained by Texas state mental health clinics, federally supported clinics (e.g. Department of Veterans Affairs facilities) and government supported welfare (Medicaid) and social security (Medicare) programs. These data are restructured to fit a common architecture that include unique patient and provider identifiers, activity dates, diagnosis codes and use-of-care information coded into a standard

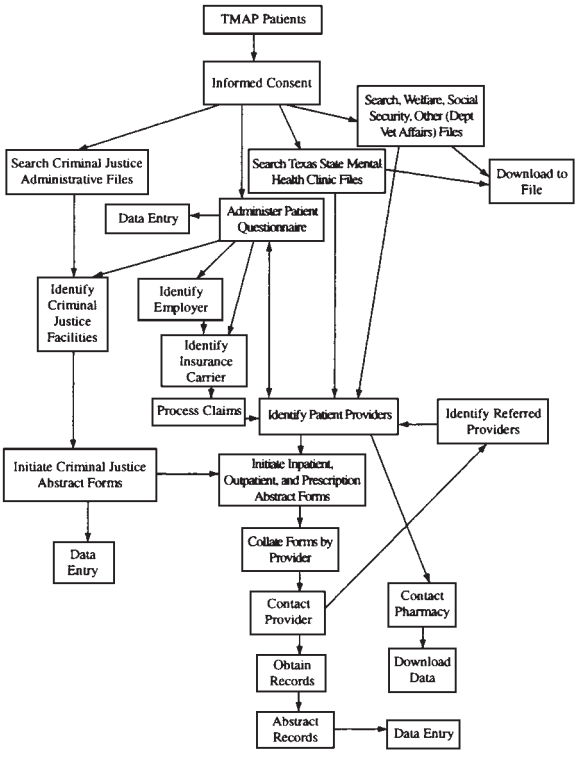

Figure 1. Flow of data collection activities

utilization format described above. For example, medical procedures coded locally using unique activity codes are mapped into a 'best' CPT code with the aid of a panel of local clinicians and information management personnel from the reporting facility who are familiar with both CPT and the local activity codes.

Subjects are administered the Utilization and Cost Questionnaire (UAC-Q) at intake and at three-month follow-ups throughout the 12–24 month study period. As described elsewhere,[11] subjects are asked to describe for the past three months (i) where they worked, (ii) their private health insurance coverage, (iii) their participation in the Texas Medicaid program, (iv) the amount of time and method they used to travel to their assigned Texas state mental health clinic, (v) the name and address of each provider where they had obtained services, including the volume of care (inpatient days, day hospital days, clinic visits, home visits) and the reason for encounters (psychiatric, medical), by type of facility (clinic, emergency rooms, hospitals or other institutions), (vi) where they get prescriptions filled and (vii) the location of correctional institutions where they served time.

A list of all the patient's health care providers is constructed from UAC-Q responses, a review of administrative files, insurance company claims data and records from correctional institutions. The patient's insurance company is identified from his/her employer named in the UAC-Q. The employer is asked only general information about their employee health coverage. Correctional facilities are identified in the UAC-Q and from searches through the Texas criminal justice system. Each institution is contacted and the location

of their health care providers determined. Finally, the location of each pharmacy the patient uses is identified from patient responses to UAC-Q.

Different professionals are listed as one provider if they share a common medical record and billing system. Thus, an entire community mental health center may be treated as a single provider if the patient's medical record for all attending clinicians is centrally maintained. Information about individual providers (name, address, telephone number, affiliations) is identified over the Internet and in published medical directories, and confirmed by contacting each provider. Providers are contacted, and appropriate patient records obtained, abstracted and double entered into the computer.

Next, medical record abstracts, computerized administrative files and patient self-reports are converted into standard response codes. For example, UAC-Q reports only the number of outpatient visits and inpatient days. A total RBRVU weight for an average patient reported outpatient visit, and a total DRG weight for an average patient reported inpatient day, is estimated by patient diagnosis, age, gender and provider type from a regression analysis estimated using data from study patients who completed UAC-Qs and for whom provider records were also available.

A final database is constructed by assembling patient use-of-care information from each listed health care provider. Data come from the provider's medical records or, if unavailable, from administrative files obtained from providers and third-party payers. In the absence of any records for a given provider, we turn to patient self-reports in the UAC-Q to describe the study patient's use of care for the given provider. Recognizing limitations in written medical records, we also explore including events reported in third-party administrative files but not confirmed in patient medical records. This strategy is based on the theory that administrative files may be a richer source for cost content when used for performance evaluation and revenue generating purposes.[109,110]

This method has several advantages. First, we profile patient use of care from all available data sources, relying on patient self-reports whenever provider records are unavailable. This is possible because, unlike other health survey instruments, the UAC-Q asks subjects to describe their use of care separately for each health care provider. Second, requests to a given provider for records are limited to only those subjects who were, in fact, patients of the provider. We thus avoid the cumbersome task of asking providers about subjects they have not treated as patients.

Analytic Methods

Guideline-driven practices are difficult to study because providers caring for patients within the same facility are likely to share information concerning treatments. To avoid treatment blends, the schizophrenia and bipolar and major depressive disorder algorithms were matched across 20 clinics so that no clinic was supporting algorithms for more than one disorder. In addition, control clinics were selected where no algorithm was implemented. Patients attending each clinic enter the study at the time of a change in psychiatric medication. Health outcomes are assessed at intake, and every three months thereafter. The UAC longitudinal estimates of health care costs span one year prior to intake, and continuously thereafter.

Our cost data are time dependent events which are nested within patients who, in turn, are nested within clinics. Clinics, rather than patients, are assigned to algorithm-based or treatment-as-usual practices. While many methods are available, we summarize costs over three-month periods that end at intake and at each three-month follow-up when the UAC-Q and health outcome questionnaires are administered.

To handle these data, we employ a three-level growth curve[144,145] analysis. These models are also known as random-effects,[146,147] random regression,[148,149] empirical Bayes,[150] general mixed linear[151] and hierarchical linear[152] models. Estimates are calculated using HLM/3L software.[153]

Mathematically, the three-month cost (c_{sit}) for patient i assigned to facility s at time t is assumed to be a linear function of an initial cost level (β_0^c), a time growth rate, β_1^c and random patient–time effects, v. In level two, the cost growth rate is determined from a facility specific average rate, fixed characteristics associated with the patient (x_{si}) and random patient effects (e^c). In level three, the facility specific cost growth rate is determined by a constant term (φ_0^c), a facility-specific treatment indicator variable (I_S) and a random facility effect (u). The treatment indicator variable assumes the value of one if the facility where the patient receives care has been assigned to an algorithm-based practice, and zero if assigned to treatment-as-usual. The three-level model is represented by

$$c_{sit} = \beta_{0si}^c + \beta_{1si}^c t + v_{sit}^c \qquad \text{level I} \qquad (4a)$$

$$\beta_{1si}^c = \gamma_{0s}^c + \gamma_1^c x_{si} + e_{si}^c \qquad \text{level II} \qquad (4b)$$

$$\gamma_{0s}^c = \varphi_0^c + \varphi_1^c I_s + u_s^c \qquad \text{level III} \qquad (4c)$$

where v, e and u are independently normal random variates with zero mean and constant variances. These analyses may be expanded in several ways, with the appropriate model selected to have the best overall fit of the data. These include expanding level 1 to include higher orders of time (t^2, t^3 etc) to account for non-linear changes in the growth rate with time. The random effects may be expanded to include heteroscedastic and autocorrelated level-1 co-variance structures,[154–156] and censored samples that include many non-users of services. The model will be expanded by the approach first explored by Duan *et al.*[157] extended by Pohlmeier and Ulrich[158] and described econometrically by Maddala[159] in which use versus non-use (logistic regression) is modeled separately from the volume of use among health care users (linear regression). Volume variables are normalized with a log, or other appropriate, transformation.[160–162]

The difference in cost between algorithm based and treatment-as-usual is estimated from the parameter φ_1^c. A similar model may be formed to examine the impact of a practice protocol on health outcomes. If these parameters

are denoted by the superscript '*o*', then the cost-outcome ratio of differences would be computed by $\varphi_1^o / \varphi_1^c$.

Alternatively, we construct cost models to estimate the cost–outcome difference in ratios. These models take the form

$$o_{sit} = \beta_{0si}^p + \beta_{1si}^p c_{sit} + v_{sit}^p \quad \text{level I} \quad (5a)$$

$$\beta_{1si}^p = \gamma_{0s}^p + \gamma_1^p x_{si} + e_{si}^p \quad \text{level II} \quad (5b)$$

$$\gamma_{0s}^p = \varphi_0^p + \varphi_1^p I_s + u_s^p \quad \text{level III} \quad (5c)$$

where the superscript '*p*' refers to health care production. The cost–outcome difference in ratios for algorithm-based practices is computed from an estimate of φ_1^p. While computations for the confidence intervals and significance tests for the cost–outcome statistic are simple, the 'outcome as a function of cost' equations ignore 'costs' as an outcome variable that may be subject to the same influences as patient health outcomes. This may be addressed statistically using predicted costs as an instrumental variable.[19,163,164]

Our analytic approach takes into account treatment assignment by facility, does not require fixed intervals between actual follow-up observations, can measure how practice effects vary with time and allows for more flexible covariance structures for a better model fit.

Conclusion

Clinical guidelines, protocols, preferred practices, clinical pathways and algorithms all embrace a concept of optimal care in which the latest medical knowledge is applied to finding treatments that balance health outcomes with the costs of care. For TMAP, we therefore measure cost in the context of a cost–effectiveness study in which 'effectiveness' is measured in terms of symptoms, functioning and satisfaction. To define costs, we rely on the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine to focus on the direct cost of both mental health and general medical care. Services include those provided by the program, other mental health specialty providers and general medical care, including services patients receive while homeless or incarcerated in a correctional institution. The UAC methodology allows us to compute costs with respect to different payers and in varying economic environments. Payers include the practitioner who produces the care, public and private third-party payers and the consumer. Costs are calculated by multiplying unit costs by patient utilization rates. Unit costs are calculated by multiplying a payer-specific conversion factor by a relative value weight assigned to each outpatient procedure or inpatient day. Patient use of care is determined by integrating medical records, administrative files and patient self-reports to obtain a comprehensive profile of the patient's use of care. Differences in costs between guideline-driven and treatment-as-usual care will be based on differences in the quantity and mix of services utilized and the time physicians spend diagnosing, treating and assessing the patient. Hierarchical models are used to compute differences between practitioners who follow the medication algorithm and treatment-as-usual in costs per patient, in costs per unit change in health outcomes and in outcome achieved per health care dollar spent.

In conclusion, our cost estimates are based on a structured data collection methodology that profiles patient use of all providers and health services, that analytically handles missing data and that describes how the burden of costs falls on different payer groups. We thus believe our methodology represents the most complete system to assess patient care costs that is available to date.

Acknowledgements

This project was supported, in part, by the Texas Medication Algorithm Project (TMAP) through grants from the Robert Wood Johnson Foundation (No. 031023), Meadows Foundation (No. 97040055), the National Institute of Mental Health (MH53799), Bristol-Myers Squibb Company, Eli Lilly and Company, Glaxo Wellcome, Inc., Janssen Pharmaceutica, Novartis Pharmaceuticals Comporation, Pfizer, Inc., Wyeth-Ayerst Laboratories, the Mental Health Connections, Betty Jo Hay Distinguished Chair in Mental Health, Rosewood Corporation Chair in Biomedical Science and a Department of Veterans Affairs Health Services Research and Development Research Career Scientist Award (RCS No. 92-403). The opinions expressed herein, however, are those of the authors and do not necessarily reflect the views of supporting agencies and sponsors.

References

1. Field MJ, Lohr KN, eds. *Clinical Practice Guidelines: Directions for a New Program*. US Institute of Medicine Committee to Advise the Public Health Service on Clinical Practice Guidelines. US Dept Health Human Services, Washington, DC: National Academy Press, 1990.
2. *Clinical Practice Guideline Number 5: Depression in Primary Care, Vol 1. Detection and Diagnosis*, AHCPR Publication 93-0550. Rockville, MD: US Department of Health and Human Services, Agency for Health Care Policy and Research, 1993.
3. American Psychiatric Association. APA Practice Guidelines. Washington, DC: American Psychiatric Association, 1996.
4. Jobson KO, Potter WZ. International Psychopharmacology Algorithm Project Report: introduction. *Psychopharmacol. Bull.* 1995; **31**: 457–459.
5. Crismon ML, Madhukar T, Pigott TA, Rush AJ, Hirschfeld MA, Kahn DA, DeBattista C, Nelson JC, Nierenberg AA, Sackeim HA, Thase ME, and the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder. The Texas Medication Algorithm Project: Report of the Texas Consensus Conference Panel on medication treatment of major depressive disorder. *J. Clin. Psychiatry* 1999; **60**: 142–156.
6. Managed Care Services. Behavioral Health: Principles and Procedures for Quality Care Management. Houston, TX: Behavioral Health Services, 1993.
7. GreenSpring Health Services. Care Management Model: Clinical Policies and Procedures. Columbia, MD: GreenSpring Health Services, 1992.
8. Lubarsky DA, Glass PSA, Ginsberg B, Dear G de L, Dentz ME, Gan TJ, Sanderson IC, Mythen MG, Dufore S, Pressley C, Gilbert WC, White WD, Alexander ML, Coleman RL, Rogers M, Reves JG. The successful implementation of pharmaceutical practice guidelines: analysis of associated outcomes and cost savings. *Anesthesiology* 1997; **86**: 1145–1160.
9. La Ruche G, Lorougnon F, Digbeu N. Therapeutic algorithms for the management of sexually transmitted diseases at the peripheral

level in Côte d'Ivoire: assessment of efficacy and cost. *Bull. World Health Organization* 1995; **73**: 305–313.
10. McFadden ER Jr, Elsanadi N, Dixon L, Takacs M, Deal EC, Boyd KK, Idemoto BK, Broseman LA, Panuska J, Hammons T, Smith B, Caruso F, McFadden CB, Shoemaker L, Warren EL, Hejal R, Strauss L, Gilbert IA. Protocol therapy for acute asthma: therapeutic benefits and cost savings. *Am. J. Med.* 1995; **99**: 651–661.
11. Kashner TM, Suppes T, Rush AJ, Altshuler KZ. Measuring use of outpatient mental health services: a comparison of self reports and provider records. *Evaluation Program Planning* 1999; **22**: 31–39.
12. Gilbert DA, Altshuler KZ, Rago WV, Shon SP, Crismon ML, Toprac MG, and Rush AJ. Texas Medication Algorithm Project: Definitions, rationale and methods to develop medication algorithms. *J. Clin. Psychiatry* 1998; **59**: 345–351.
13. Rush AJ, Rago WV, Crismon ML, Toprac MG, Shon SP, Suppes T, Miller AL, Trivedi MH, Swann AIC, Biggs MM, Shores-Wilson K, Kashner TM, Pigott T, Chiles JA, Gilbert DA and Altshuler KZ. Medication treatment for the severely and persistently mentally ill: The Texas Medication Algorithm Project. *J. Clin. Psychiatry* 1999; **60**: 284–291.
14. Rush AJ, Crismon ML, Toprac M, Trivedi MH, Rago WV, Shon S, Altshuler KZ. Consensus guidelines in the treatment of major depressive disorder. *J. Clin. Psychiatry* 1998; **59** (suppl. 20): 73–84.
15. Chiles JA, Miller AL, Crismon ML, Rush AJ, Krasnoff AS, Shon SS. The Texas Medication Algorithm project: Development and implementation of the schizophrenia algorithm. *Psychiatr. Services* 1999; **50**: 69–74.
16. Rush AJ, Prien RF. From scientific knowledge to the clinical practice of psychopharmacology: can the gap be bridged? *Psychopharmacol. Bull.* 1995; **31**: 7–20.
17. Elixhauser A, Halpern M, Schmier J, Luce BR. Health care CBA and CEA from 1991 to 1996: An Updated Bibliography. *Med. Care* 1998; **36**: MS1–MS9.
18. Mishan EJ. Cost–Benefit Analysis. New York: Praeger, 1976.
19. Heckman J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *J. Human Resources* 1997; **32**: 441–462.
20. Manski CF. Learning about treatment effects from experiment with random assignment of treatments. *J. Human Resources* 1996; **31**: 707–733.
21. Bentkover JD, Covello VT, Mumpower J, Spicker SF. Benefits Assessment: the State of the Art. Boston, MA: Reidel, 1986.
22. Morris CB and Rittenhouse BE (eds). Standards for Economic Evaluation of Drugs: Issues and Answers. *Med. Care* 1996; **34**: DS1–DS239.
23. Eddy MH. Clinical decision making: from theory to practice. Applying cost-effectiveness analysis. *J. Am. Med. Assoc.* 1992; **268**: 2575–2582.
24. Yates B. Cost-effectiveness analysis, cost–benefit analysis, and beyond. Evolving models for the scientist–manager–practitioner. *Clin. Psychol. Sci. Practice* 1995; **2**: 385–398.
25. Cooper BS and Rice DP. The economic cost of illness revised. *Social Security Bull.* 1976; **39**: 21–36.
26. Warner KE, Luce BR. Cost–Benefit and Cost Effectiveness in Health Care. Ann Arbor: Health Administration Press, 1982.
27. Donabedian A. Benefits in Medical Care Program. Cambridge, MA: Harvard University Press, 1976.
28. Drummond MF, O'Brien BJ, Stoddard GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press, 1997.
29. Siegel JE, Laska E, Meisner M. Statistical methods for cost-effectiveness analyses. *Controlled Clin. Trials* 1996; **17**: 387–404.
30. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *New Engl. J. Med.* 1977; **296**: 716–721.
31. Heithoff KA and Lohr KN, eds. US Institute of Medicine Division of Health Care Services. (corporate author) *Effectiveness and Outcomes in Health Care*. Washington, DC: National Academy Press, 1990.
32. Schulman KA, Yabroff KR. Measuring the cost-effectiveness of cancer care. *Oncology* 1995; **9**: 523–538.
33. Krumholz HM, Pasternak RC, Weinstein MC, Friesinger GC, Ridker PM, Tosteson ANA, Goldman L. Cost effectiveness and thrombolytic therapy with streptokinase in elderly patients with suspected acute myocardial infarction. *New Engl. J. Med.* 1992; **327**: 7–13.
34. Wells KB, Sturm R, Sherbourne CD, Meredith LS. Caring for Depression. Cambridge, MA: Harvard University Press, 1996.
35. Miller NE, Magruder KM (eds). The Cost Effectiveness of Psychotherapy: A Guide for Practitioners Researchers and Policymakers. New York: Oxford University Press, 1999.
36. Hargreaves WA, Shumway M, Hu TW, Cuffel B. Cost–outcome Methods for Mental Health. San Diego, CA: Academic, 1998.
37. Kashner TM, Rush AJ, Altshuler KZ. Managed care and the focus on outcomes research. *J. Practical Psychiatry Behav. Health* 1997; **3**: 135–145.
38. Patrick DL, Erickson P. Health Status and Health Policy: Allocating Resources to Health Care. New York: Oxford University Press, 1993.
39. Torrance GW, Feeny D. Utilities and Quality-Adjusted Life Years. *Int. J. Technol. Assessment Health Care* 1989; **5**: 559–575.
40. Andreasen NC. The scale for the assessment of negative symptoms (SANS): Conceptual and theoretical foundations. *Br. J. Psychiatry* 1989; **155** (suppl. 7): 49–52.
41. Overall JE, Gorham DR. Introduction – the Brief Psychiatric Rating Scale (BPRS): Recent developments in ascertainment and scaling. *Psychopharmacol. Bull.* 1988; **24**: 97–99.
42. Ventura J, Green MF, Shaner A, Liberman RP. Training and quality assurance with the Brief Psychiatric Rating Scale: 'The Drift Busters'. *Int. J. Methods Psychiatric Res.* 1993; **3**: 221–244.
43. Addington D, Addington J, Maticka-Tyndale E. Assessing depression in schizophrenia: the Calgary Depression Scale. *Br. J. Psychiatry* 1993; **163** (suppl. 22): 39–44.
44. Altman E, Hedeker D, Janicak P, Peterson J, Davis J. The clinician-administered rating scale for mania (CARS-M): Development, reliability, and validity. *Biol. Psychiatry* 1994; **36**: 124–134.
45. Bauer MS, Crits-Christoph P, Ball WA, Dewees E, McAllister T, Alahi P, Cacciola J, Whybrow PC. Independent assessment of manic and depressive symptoms by self-rating: Scale characteristics and implications for the study of mania. *Arch. Gen. Psychiatry* 1991; **48**: 807–812.
46. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger JE, Burns CT. The inventory of depressive symptomatology (IDS): Preliminary findings. *Psychopharmacol. Bull.* 1986; **22**: 985–990.
47. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The inventory of depressive symptomatology (IDS): Psychometric properties. *Psychol. Med.* 1996; **26**: 477–486.
48. Levine J, Schooler N. SAFTEE: A technique for the systematic assessment of side effects in clinical trials. *Psychopharmacol. Bull.* 1986; **22**: 343–381.
49. Barnes TRE. A rating scale for drug-induced akathisia. *Br. J. Psychiatry* 1989; **154**: 672–676.
50. Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey (SF-12): Construction of scales and preliminary tests of reliability and validity. *Med. Care* 1996; **34**: 220–233.
51. Lehman AF. Quality of life interview (TI30S) Self-administered questionnaire. Unpublished instrument.
52. Lehman AF. A quality of life interview for the chronically mentally ill. *Evaluation Program Planning* 1988; **11**: 51–62.
53. Heinrichs DW, Hanion TE, Carpenter WT Jr. The quality of life scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bull.* 1984; **10**: 388–398.
54. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med. Care* 1986; **24**: 67–74.
55. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. *J. Am. Med. Assoc.* 1996; **276**: 1253–1258.
56. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. *J. Am. Med. Assoc.* 1996; **276**: 1172–1177.
57. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. *J. Am. Med. Assoc.* 1996; **276**: 1339–1341.
58. Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
59. Kashner TM. Cost effectiveness of lead screening, comment. *New Engl. J. Med.* 1982; **307**: 1260.
60. Taube CA, Mechanic D, Hohmann A, eds. The Future of Mental Health Services Research. DHHS Publication No. (ADM)89-1600. Washington, DC: US Government Printing Office, 1989.
61. Eisenberg L. Treating depression and anxiety in primary care. *New Engl. J. Med.* 1992; **326**: 1080–1083.
62. Danton WG, Altrocchi J, Antonuccio D, Basta R. Nondrug treatment of anxiety. *Am. Family Physician* 1994; **10**: 161–166.

63. Simon GE, VonKorff M, Barlow W. Health care costs of primary care patients with recognized depression. *Arch. Gen. Psychiatry* 1995; **52**: 850–856.
64. Manning W, Wells KB. The effects of psychological distress and psychological well-being on use of medical services. *Med. Care* 1992; **30**: 541–553.
65. Johnson J, Weissman M, Klerman G. Service utilization and social morbidity associated with depressive symptoms in the community. *J. Am. Med. Assoc.* 1992; **267**: 1478–1483.
66. Levenson J, Hamer RM, Rossiter LF. Relation of psychopathology in general medical inpatients to use and cost of services. *Am. J. Psychiatry* 1990; **147**: 1498–1503.
67. Kessler LG, Burns BJ, Shapiro S, Tischler GI, George LK, Hough RL, Bodison D, Miller RH. Psychiatric diagnoses of medical service users: evidence from the Epidemiologic Catchment Area program. *Am. J. Public Health* 1987; **77**: 18–24.
68. Mechanic D, Cleary PD, Greenley JR. Distress syndromes, illness behavior, access to care and medical utilization in a defined population. *Med. Care* 1982; **20**: 361–372.
69. Borus JF, Olendzki MC. The offset effect of mental health treatment on ambulatory medical care utilization and charges. *Arch. Gen. Psychiatry* 1985; **42**: 573–580.
70. Patterson D, Bise B. Report to NIMH, contract 282-77-0219 MS. Rockville, MD, 1978.
71. Duehrssen A. Katamnestische Ergebnisse bei 1004 Patienten nach analytischer Psychotherapie. *Z. Psychosomat. Med*. 1962; 8.
72. Hankin JR, Kessler LG, Goldberg ID, Steinwachs DM, Starfield BH. A longitudinal study of offset in the use of nonpsychiatric services following specialized mental health care. *Med. Care* 1983; **21**: 1099–1110.
73. Jameson J, Shuman LJ, Young WW. The effects of outpatient psychiatric utilization on the costs of providing third party coverage. *Med. Care* 1978; **16**: 383–399.
74. Jones KR, Vischi TR. Impact of alcohol, drug abuse and mental health treatment on medical care utilization. *Med. Care* 1979; 12 (suppl.).
75. Mumford E, Schlesinger HJ, Glass GV, Patrick C, Cuerdon T. A new look at evidence about reduced cost of medical utilization following mental health treatment. *Am. J. Psychiatry* 1984; **141**: 1145–1158.
76. Rupp A. The economic consequences of not treating depression. *Br. J. Psychiatry* 1995; **166** (suppl. 27): 29–33.
77. Holder HD, Blose JO. Changes in health care costs and utilization associated with mental health treatment. *Hosp. Community Psychiatry* 1987; **38**: 1070–1075.
78. Salvador-Carulla L, Segui J, Fernandez-Cano P, Canet J. Costs and offset effect in panic disorder. *Br. J. Psychiatry* 1995; **166** (suppl. 27): 23–28.
79. Cummings NA. Prolonged (ideal) versus short-term (realistic) psychotherapy. *Prof. Psychol.* 1977; **8**: 491–501.
80. McHugh JP, Kahn MW, Heiman E. Relationships between mental health treatment and medical care utilization among low-income Mexican–American patients: some preliminary findings. *Med. Care* 1977; **15**: 439–444.
81. Friedman R, Myers P, Sobel D, Caudill M, Benson H. Behavioral medicine, clinical health psychology, and cost offset. *Health Psychol*. 1995; **14**: 509–518.
82. Magaziner J, Simonsick EM, Kashner TM, Hebel JR. Predictors of functional recovery one year following hospital discharge for hip fracture: a prospective study. *J. Gerontol. Med. Sci*. 1990; **45**: M101–107.
83. Mossey JM, Mutran E, Knott K, Craik R. Determinants of recovery 12 months after hip fracture: the importance of psychosocial factors. *Am. J. Public Health* 1989; **79**: 279–286.
84. Frasure-Smith N, Lespérance F, Talajic M. Depression following myocardial infarction. Impact on six-months survival. *J. Am. Med. Assoc*. 1993; **270**: 1819–1825.
85. Jerrell JM, Hu T-W. Cost-effectiveness of intensive clinical and case management compared with an existing system of care. *Inquiry* 1989; **26**: 224–234.
86. Grossman M. On the concept of health capital and the demand for health. *J. Polit. Econ.* 1972; **80**: 223–255.
87. American Medical Association. Current Procedural Terminology, 4th edn. Reston, VA: St. Anthony, 1998.
88. 1998 St. Anthony's DRG Guidebook. Reston, VA: St. Anthony, 1998.
89. Health Care Financing Administration. Federal Register, Vol. 62, No. 211 (31 October 1997), 59 048–59 269.
90. Hsiao WC, Braun P, Dunn DL, Becker ER, Yntema D, Verrilli DK, Stamenovic E, Shiao-Ping C. Overview of the development and refinement of the resource-based relative value scale. *Med. Care* 1992; **30** (suppl): NS1–NS12.
91. Health Care Financing Administration. Medicare Program: Changes to the Inpatient Hospital Prospective Payment System and Fiscal Year 1998 Rates: Final Rule. Federal Register, Vol. 62, No 168; 29 August, 1997: 45 966–46 140.
92. Health Care Financing Administration. Medicare Program: Prospective Payments for Medicare Inpatient Hospital Services. Federal Register, Vol. 48, No. 171; 1 September 1983: 39 752–39 890.
93. Kashner TM. Final Report to the Rate Alternative Technical Evaluation Project: Phase IV: Estimating VA Cost-Based Itemized Physician Charges for Inpatient and Outpatient Care Services. Medical Care Cost Recovery Office, Department of Veterans Affairs, Washington, DC, 1997.
94. Menke TJ, Homan RH and Kashner TM. Issues in determining costs in the Department of Veterans Affairs. *Med. Care* 1999; **37** (suppl.): AS18–AS26.
95. Romm FJ, Putnam SM. The validity of the medical record. *Med. Care* 1981; **19**: 310–315.
96. Feigl P, Glaefke G, Ford L, Diehr P, Chu J. Study patterns of cancer care: how useful is the medical record. *Am. J. Public Health* 1988; **78**: 526–533.
97. Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA. Coding of acute myocardial infarction: clinical and policy implications. *Ann. Intern. Med.* 1988; **109**: 745–751.
98. Iezzoni LI, Daley J, Heeren T, Foley SM, Fisher ES, Duncan C, Hughes JS, Coffman GA. Identify complications of care using administrative data. *Med. Care* 1994; **32**: 700–715.
99. Agency for Health Care Policy and Research (AHCPR). Report to Congress: The feasibility of linking research-related data bases to Federal and non-Federal medical administrative databases. AHCPR publication No. 91-003. Rockville, MD: AHCPR, 1991.
100. Tierney WM, McDonald CJ. Practice databases and their uses in clinical research. *Stat. Medicine* 1991; **19**: 541–557.
101. Roper WL, Winkenwerder W, Hackbarth GM, Krakauer H. Effectiveness in health care: an initiative to evaluate and improve medical practice. *New Engl. J. Med.* 1988; **319**: 1197–1202.
102. Ellwood PM. Shattuck lecture – outcomes management. A technology of patient experience. *New Engl. J. Med.* 1988; **318**: 1549–1556.
103. Dannenberg AL, Shapiro AR, Fries JF. Enhancement of clinical predictive ability by computer consultation. *Methods Information Med.* 1979; **18**: 10–14.
104. Grady ML, Schwartz HA. Automated Data Sources for Ambulatory Care Effectiveness Research (AHCPR Publication No. 93-002). Rockville, MD: AHCPR, 1993.
105. Hannan EL, Racz MJ, Jollis JG, Peterson ED. Using Medicare claims data to assess provider quality for CABG surgery: does it work well enough? *Health Services Res*. 1997; **31**: 659–678.
106. Dick RS, Steen EB (eds). US Institute of Medicine Committee on Improving the Patient Record. (corporate author) The computer based patient record: An essential technology for health care. Washington: Institute of Medicine, 1995.
107. Gardner E. Red tape, antiquated laws keep computerized records on shaky legal ground. *Mod. Healthcare* 1991; **21**: 27, 30–31.
108. Eddy DM. Variation in physician practice: The role of uncertainty. *Health Aff.* 1984; **3**: 74–89.
109. Fowles JB, Lawthers AG, Weiner JP, Garnick DW, Petrie DS, Palmer RH. Agreement between physicians' office records and Medicare Part B Claims data. *Health Care Financing Rev.* 1995; **16**: 189–199.
110. Kashner TM. Agreement between administrative files and written medical records: A case of the Department of Veterans Affairs. *Med. Care* 1998; **36**: 1324–1336.
111. Narrow WE, Regier DA, Rae DS, Manderscheid RW, Locke BZ. Use of services by persons with mental and addictive disorders: Findings from the National Institute of Mental Health Epidemiologic Catchment Area Program. *Arch. Gen. Psychiatry* 1993; **50**: 95–107.
112. Kovar MG, Poe GS. *The National Health Interview Survey design: 1973–84, and procedures, 1975–83. Vital and health statistics*, Series 1, No. 18, Department of Health and Human Services Publication No. (PHS) 85-1320. Hyattsville, MD: National Center for Health Statistics, 1985.
113. Tarlov AR, Ware JE, Greenfield EC, Perrin NE, Zubkoff M. The Medical Outcomes Study: An application of methods for monitoring the results of medical care. *J. Am. Med. Assoc*. 1989; **262**: 925–930.

114. Bonham GS. *Procedures and questionnaires of the National Medical Care and Expenditure Survey*, Series A, Methodological Report No. 1. Publication No. (PHS) 82-20001. Hyattsville, MD: National Center for Health Statistics, 1983.
115. Horgan CM. Specialty and general ambulatory mental health services. *Arch. Gen. Psychiatry* 1985; **42**: 565–572.
116. Gelberg L, Linn LS, Leake BD. Mental health, alcohol and drug use, and criminal history among homeless adults. *Am. J. Psychiatry* 1988; **145**: 191–1988.
117. Koegel P, Burnam A, Farr RK. The prevalence of specific psychiatric disorders among homeless individuals in the inner city of Los Angeles. *Arch. Gen. Psychiatry* 1988; **45**: 1085–1092.
118. Padgett D, Struening EL, Andrews JH. Factors affecting the use of medical, mental health, alcohol and drug treatment services by homeless adults. *Med. Care* 1990; **8**: 806–821.
119. Donabedian A. Aspects of Medical Care Administration. Cambridge: Harvard University Press, 1973.
120. Trussell RE, Elinson J, Levin ML. Comparison of various methods of estimating the prevalence of chronic disease in a community — the Hunterdon County Study. *Am. J. Public Health* 1956; **46**: 173–182.
121. National Center for Health Statistics. *Reporting Hospitalization in the Health Interview Survey*, PHS Publication No. 1000, series 2, No. 6. Washington, DC: US Government Printing Office, 1965.
122. National Center for Health Statistics. *Interview data on chronic condition compared with information derived from medical record*, PHS Publication No. 1000, series 2, No. 23. Washington, DC: US Government Printing Office, 1967.
123. National Center for Health Statistics. *Net differences in interview data on chronic conditions and information derived from medical records*, PHS Publication No. HSM 73-1331, series 2, No. 57. Washington, DC: US Government Printing Office, 1973.
124. Simmons WR, Bryant EE. An evaluation of hospitalization data from the health interview survey. *Am. J. Public Health* 1962; **52**: 1638–1647.
125. Mechanic D, Newton M. Some problems in the analysis of morbidity data. *J. Chronic Dis.* 1966; **36**: 223.
126. Brown JB, Adams ME. Patients as reliable reporters of medical care process: Recall of ambulatory encounter events. *Med. Care* 1992; **30** (5): 400–411.
127. Colombotos J, Elinson J, Loewenstein R. Effect of interviewer's sex on interview responses. *Public Health Rep*. 1968; **83**: 685–690.
128. U.S. Department of Health, Education and Welfare, Public Health Service. *Health stastistics from the US National Health Survey: health interview responses compared with medical records*, PHS Publication No. 584 – D5. Washington, DC: US Government Printing Office, 1961.
129. Kahn RL, Cannell CF. The Dynamics of Interviewing. New York: Wiley, 1957.
130. Cannell CF, Fowler FJ, Marquis KH. *The influence of interview and respondent psychological and behavioral variables on the reporting in household interviews*, PHS Publication No. 10000, series 2, No. 26. Washington, DC: US Government Printing Office, 1968.
131. Magaziner J, Simonsick EM, Kashner TM and Hebel JR. Patient–proxy response comparability on measures of patient health and functional status. *J. Clin. Epidemiol.* 1988; **41**: 1065–1074.
132. Simon GE, VonKorff M. Recall of psychiatric history in cross-sectional surveys: Implications for epidemiologic research. *Epidemiol. Rev.* 1995; **17**: 221–1227.
133. Golding JM, Gongla P, Brownell A. Feasibility of validating survey self-reports of mental health service use. *Am. J. Community Psychol.* 1988; **16**: 39–51.
134. Manderscheid RW, Rae DS, Narrow WE, Locke BZ, Regier DA. Congruence of service utilization estimates from the epidemiologic catchment area project and other sources. *Arch. Gen. Psychiatry* 1993; **50**: 108–114.
135. Manderscheid RW, Witkin MJ, Rosenstein MJ, Bass RD. The national reporting program for mental health statistics: history and findings. *Public Health Rep.* 1986; **101**: 532–539.
136. Cannell CF, Fisher G, Bakker T. *Reporting of hospitalization in the Health Interview Survey*, US Department of Health, Education and Welfare, National Center of Health Statistics, series 2, No. 6. Washington, DC: US Government Printing Office, 1965.
137. Spector PE, Bedell JR. Measuring program effectiveness: Self-report versus objective indicators of recidivism. *Am. J. Community Psychol.* 1982; **10**: 613–616.
138. Taube CA, Schlenger WE, Rupp A, Whitmore RW. Validity of Medicaid household respondent reporting of ambulatory visits for mental disorders. *J. Econ. Social Meas.* 1986; **14**: 243–256.
139. Calsyn RJ, Morse GA, Klinkenberg WD, Trusty ML. Reliability and validity of self-report data of homeless mentally ill individuals. Evaluation Program Planning 1997; **20**: 47–54.
140. Kashner TM, Rost K, Smith GR, Lewis S. An analysis of panel data: The impact of a psychiatric consultation letter on the expenditures and outcomes of care for patients with somatization disorder. *Med. Care* 1992; **30**: 811–821.
141. Clark RE, Ricketts SK, McHugo GJ. Measuring hospital use without claims: a comparison of patient and provider reports. *Health Services Res.* 1996; **31**: 153–169.
142. Quam L, Ellis LBM, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiological research: the concordance of claim-based criteria with the medical record and patient survey for identifying a hypertensive population. *Med. Care* 1993; **31**: 498–507.
143. Dresser MVB, Feingold L, Rosencranz SL, Coltin KL. Clinical quality measurement: Comparing chart review and automated methodologies. *Med. Care* 1997; **35**: 539–552.
144. Stanek EJ III, Diehl SR. Growth curve models of repeated binary response. *Biometrics* 1988; **44**: 973–983.
145. Willett JB, Ayoub C, Robinson D. 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. *J. Consulting Clin. Psychol.* 1991; **59**: 38–47.
146. Laird NM, Ware H. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–974.
147. Ware JH. Linear models for the analysis of longitudinal studies. *Am. Stat.* 1985; **39**: 95–101.
148. Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, Shea T, Imber SD, Sotsky SM, Watkins JT. 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.
149. Jennrich RI, Schluchter MD. Unbalanced repeated measures models with structured covariance matrices. *Biometrics* 1986; **42**: 805–820.
150. Maritz TS. Empirical Bayes Methods. London: Methuen, 1970.
151. Goldstein HI. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 1986; **73**: 43–56.
152. Bryk AS, Raudenbush SW. Hierarchical Linear Models. Newbury Park, CA: Sage, 1992.
153. Bryk AS, Raudenbush SW, Congdon RT Jr. *Hierarchical Linear and Nonlinear Modeling, with HLM/2I and HLM/3L Programs* version 4. Chicago, IL: Scientific Software, 1996.
154. Louis TA, Spiro A III. *Fitting first order auto-regressive models with covariates*, Technical Report. Cambridge, MA: Harvard University School of Public Health, Department of Biostatistics, 1984.
155. Hedeker D. *Random Regression Models with Autocorrelated Errors*, Chicago, IL: University of Chicago, 1989 (doctoral dissertation).
156. Chi EM, Reinsel GC. Models of longitudinal data with random effects and AR(1) errors. *J. Am. Stat. Assoc.* 1989; **84**: 452–459.
157. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. *J. Bus. Econ. Stat.* 1983; **1**: 115–126.
158. Pohlmeier W. and Ulrich V. An econometric model of the two-part process in the demand for health. *J. Human Resources* 1995; **30**: 339–361.
159. Maddala GS. Limited-Dependent and Qualitative Variables in econometrics. New York: Cambridge University Press, 1988.
160. Theil H. Principles of Econometrics. New York: Wiley, 1971.
161. Kmenta J. Elements of Econometrics. New York: Macmillan, 1973.
162. Amemiya T. Advanced Econometrics. Cambridge: Harvard University Press, 1985.
163. Heckman J and Hotz VJ. Choosing among alternative non-experimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc*. 1989; **84**: 862–874.
164. Heckman J. Varieties of selection bias. *Am. Econ. Rev.* 1990; **80**: 313–318.

