Volume Title: Medical Care Output and Productivity

Volume Author/Editor: David M. Cutler and Ernst R. Berndt, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-13226-9

Volume URL: http://www.nber.org/books/cutl01-1

Publication Date: January 2001

Chapter Title: Treatment Price Indexes for Acute Phase Major Depression

Chapter Author: Ernst R. Berndt, Susan Busch, Richard Frank

Chapter URL: http://www.nber.org/chapters/c7638

Chapter pages in book: (p. 463 - 508)

12

Treatment Price Indexes for Acute Phase Major Depression

Ernst R. Berndt, Susan H. Busch,
and Richard G. Frank

It is not well to sneer at political economy in its relation to the insane poor. Whether we think it right or not, the question of cost has determined and will continue to determine their fate for weal or woe.
—Asylum Superintendent George Cook, 1866

12.1 Introduction

Much has been written in the last decade on broad trends in medical care spending in the United States. Although the most recent evidence is somewhat ambiguous, the apparent slowdown in the rate of increase in aggregate health expenditures over the last five years has been welcomed by many governments, employers, patients, and insurers. Relatively little attention, however, has focused on components of the change in expenditures. Is the trend change in medical care expenditures due to changes in price, quantity, quality, or a combination of all three? What has been the role of organizational change, such as the growth of managed care, relative to technological/pharmaceutical innovation, such as new medications, on expenditure trends?

A standard procedure for economists analyzing changing expenditures over time is to employ official government price indexes, divide nominal expenditures by such a price index, and thereby decompose expenditures

Ernst R. Berndt is professor of applied economics at the Sloan School of Management of the Massachusetts Institute of Technology and director of the NBER's Program on Technological Progress and Productivity Measurement. Susan H. Busch is assistant professor in the Division of Health Policy and Administration of Yale Medical School. Richard G. Frank is professor of health economics at Harvard Medical School and a research associate of the National Bureau of Economic Research.

The authors gratefully acknowledge research support from Eli Lilly and Company, the U.S. Bureau of Economic Analysis, National Science Foundation grant SBR-9511550, and National Institute of Mental Health grant MH43703. The authors are also grateful to Douglas Cocks, Thomas Croghan, David Cutler, Dennis Fixler, Mark McClellan, Will Manning, Thomas McGuire, Joseph Newhouse, Charles Phelps, Jack Triplett, and seminar participants at the Carnegie Mellon–University of Pittsburgh Applied Micro Workshop for helpful comments on earlier drafts.

into price and quantity components. Further analysis might then focus on factors affecting real quantity growth, such as productivity gains.

In fact, the Bureau of Labor Statistics (BLS) medical consumer price index (CPI) is often used by analysts interested in undertaking such expenditure decompositions. It is well known that such a practice is frequently inappropriate and misleading, for the medical CPI deals only with consumers' direct out-of-pocket payments and does not include payments from employers to insurers.

A deeper issue, however, involves the conceptual foundations underlying the medical CPI and producer price indexes (PPIs). Although some revisions are gradually being implemented, the BLS price indexes are based largely on the repricing over time of a fixed bundle of inputs. For decades, health economists have argued that a more appropriate price index is one based on the entire episodic treatment of selected illnesses and conditions, incorporating technological and institutional innovations that change the mix of inputs used to treat the condition, and including any effects on changed medical outcomes.

Developing a price index in health care can be viewed in terms of the unique characteristics of the health services in question. Consider creating a price index for a good that displays the following features:

- Consumers pay only a portion of the gross price due to insurance. The insurance is associated with moral hazard resulting in "too much" use of the good.
- There have been dramatic technical change and great advances in the benefits offered by the good, yet the full range of benefits is difficult to measure.
- There is great variety in the forms that the good takes and consumers value those forms differently.
- The supply side of the market for the good has experienced fundamental changes in its organization.
- The good is viewed as so important in some cases that the government compels people to use it.

The treatment of major depression, one of the most prevalent and disabling mental disorders, exhibits all of the features listed above. In this paper we report on the first three years of a research program aimed at measuring prices and output for the treatment of this important disease. The approach taken in this program of research builds on several recent efforts to construct price indexes for medical care. That recent research includes work by Cutler et al. (1998) that contrasts input price indexes for the treatment of heart attacks that rose by 6.7 percent per year over 1983–94, with an outcomes-adjusted index that takes into account a conservative valuation for the extension of life expectancy attributable to new heart attack treatments and increases by only 2.3 percent per year, implying a

net bias of 4.4 percent annually. Focusing on a price index for cataract surgery, Shapiro, Shapiro, and Wilcox (chap. 10 in this volume) find that a CPI-like fixed-weight input-based index increases by a factor of about nine between 1969 and 1993, whereas their preferred alternative index incorporating realized reductions in hospital lengths of stay, but ignoring any improvements in the quality of medical outcomes, increases by only a factor of three, implying an annual differential of 4.6 percent.

Because depression is a chronic recurring illness and mortality is not an endpoint, we must rely on indirect methods of incorporating information on clinical effectiveness and outcomes. In our research we have made use of results from the published clinical literature, and from official treatment guideline standards, to identify therapeutically similar treatment bundles that can then be linked and weighted to construct price indexes for treatment of specific forms of major depression. Preliminary results from this research have been published in Frank, Berndt, and Busch (1999) and Frank, Busch, and Berndt (1998).

In this paper we consolidate our findings and extend our analysis considerably. Specifically, we refine our definition of eligible treatment episodes, develop new imputations for missing data, expand the set of treatment bundles from five to seven, and estimate hedonic price regressions.

We begin the paper with an overview of current BLS procedures for constructing medical care price indexes, and then provide a background on the nature of and alternative treatments for acute phase major depression. We outline features of the retrospective medical claims database from MEDSTAT, and then discuss our implementation of treatment episode definition and identification. We then report on quantities and prices of the treatment bundles from 1991 to 1995, we construct and interpret CPI-like and PPI-like aggregate price indexes, and we comment on an initial analysis involving hedonic price procedures. We end the paper with a discussion section, followed by concluding remarks.

12.2 Current Procedures for Measuring Medical Care Price Indexes

In the U.S. context, it is useful to distinguish "supply" and "demand" prices for medical care. By supply prices we mean the total payments received by health care providers for a particular medical treatment, consisting of the sum of payments from health insurance plans (both private and public) plus that from patients' direct out-of-pocket payments (OOPPs). At this level of generality, our supply price concept relates to components of the Producer Price Index (PPI) published by the BLS.

The demand price notion we employ here is the consumers' direct OOPPs, consisting of copayments and deductible cash payments made by the patient/consumer to providers for a particular medical treatment. Thus, in our price index changes in supply price are indicative of changes

in the cost of treatment, while changes in the demand price can be traced both to changes in the cost of treatment and changes in benefit design. Our demand price concept relates to the medical care component of the BLS Consumer Price Index (CPI). As we discuss below, the presence of insurance clouds the interpretation of the demand price index for depression as a CPI. We also discuss important differences concerning the treatment of indirect consumer payments via, for example employees' contributions to employment-related health insurance.

To associate the notion of a price index for treatment episodes of care with official price indexes published by the BLS, we now describe index number procedures currently employed by the BLS, beginning with its PPI.[1]

12.2.1 BLS Procedures for Medical Care Related PPIs

The PPI "measures average changes in selling prices received by domestic producers for their output," separately by industry.[2] The PPI takes as its definition of an industry that based on the Standard Industrial Classification (SIC) code, using four-digit SIC industries and their aggregates.[3] Since its inception in 1902, the PPI has focused rather heavily on the goods-producing sectors of the U.S. economy. In 1986, in recognition of the growing importance of services, the BLS gradually began to broaden the PPI's scope of coverage in the services sectors, including medical care.

Within each industry, the BLS calculates aggregate PPIs using a modified fixed-weight Laspeyres price index formula over individual price quotes, where fixed weights are based on value of shipments data. The primary price quote at the "cell index" level of disaggregation is "the net revenue accruing to a specified producing establishment from a specified kind of buyer for a specified product shipped under specified transactions terms on a specified day of the month."[4] Although the BLS seeks transaction rather than list prices for its price quotes, the agency is well aware that compliance by firms is easier with list than with actual transaction prices. Participation by firms in the PPI is voluntary, with a "productive" compliance rate being about 63 percent in 1992 (Catron and Murphy 1996, 131, table A-2).

The BLS currently draws a sample of items for each industry on average about every seven years, and then reprices this fixed set of items monthly until an entirely new sample is drawn. In 1995 the BLS announced that for certain industries, particularly technologically dynamic ones such as pharmaceuticals and electronics in which seven-year time lags could yield a sample of products and services much older and quite unrepresentative

1. A more detailed discussion of the medical CPI and PPI data construction procedures is given in Berndt et al. (chap. 4 in this volume).

2. U.S. Department of Labor, Bureau of Labor Statistics (1992), 140.

3. See Triplett (1990) for economic issues involved in defining and aggregating industries within the SIC system.

4. U.S. Department of Labor, Bureau of Labor Statistics (1992), 141.

of market transactions, samples would be supplemented at one- or two-year intervals.

For quite some time, the BLS has published PPIs for certain medical-related manufacturing industries, such as pharmaceuticals and diagnostic equipment. It is only recently, however, that the BLS has begun publishing PPIs for medical service industries such as hospital services and physician services. The BLS initiated a PPI for an aggregate of health services in December 1994, for offices and clinics of doctors of medicine in December 1993, and for hospitals in aggregate and by type in December 1992.

A central measurement issue in the construction of medical price indexes involves the specification and implementation of a concept of industry output. For obvious reasons, the SIC structure is not well equipped to provide guidance to the BLS on what is the appropriate real output concept in medical care industries, nor on how this output quantity and output price can best be measured.

An alternative source for guidance on implementation of a medical care output concept is provided by the Health Care Financing Administration (HCFA). In 1983 HCFA implemented a prospective payment schedule for inpatient hospital care whereby hospitals received a fixed payment for each Medicare patient admission, regardless of the amount or duration of services actually provided the patient. These Medicare prospective payment schedules distinguish treatments by twenty-four major diagnostic categories, which are broken down further into 495 groupings of medical diagnoses and surgical procedures, known as diagnostic related groups (DRGs).[5]

DRGs provide one possible output concept, but currently DRGs only measure output for inpatient hospital care; many outpatient commodities (e.g., prescription pharmaceuticals) and services are not included in the DRG system. Classification schemes used by health care payers include version 4 of Current Procedural Terminology (CPT4) codes, a list containing thousands of procedures for which physicians and hospitals can bill; these CPT4 codes can be envisaged as inputs into the treatment of an illness or condition.[6] A systematic structure of diagnostic codes for illnesses and conditions is version 9 of the International Classification of Diseases (ICD-9).[7] Relationships among ICD-9, CPT4, and DRG codes are multifaceted; a single DRG encompasses treatment of somewhat arbitrary aggregations of distinct ICD-9 diagnoses, alternative combinations of CPT4 codes can be used in the treatment of a particular ICD-9 diagnosis, and a given CPT4 procedure can be used in the treatment of various ICD-9 diagnoses. The DRG system makes use of both ICD-9 and CPT4 coding

5. For a recent list of DRGs, see Prospective Payment Assessment Commission (1995), appendix E. Note also that some private insurers and Medicaid programs make use of DRGs for purposes of hospital payment.

6. For a discussion of CPT4, see American Medical Association (1990).

7. ICD-9 codes are discussed and listed in U.S. Department of Health and Human Services (1980). The ICD-9 has recently been updated to version 10.

systems. Other diagnostic-related systems used in setting risk-adjusted capitation rates include the Ambulatory Care Group algorithm (Weiner et al. 1996) and the Hierarchical Coexisting Conditions model (Ellis et al. 1996).

The BLS PPI program has initiated procedures for construction of medical service PPIs at two rather aggregate levels, physician services and hospital services. In turn these encompass a variety of more detailed industries, such as physician services from psychiatrists, general/family practitioners, internists, and hospital services from general medical and surgical hospitals, psychiatric hospitals, and specialty nonpsychiatric hospitals.

With respect to physician services, based on a sampling universe of all physician practices in the United States, the BLS constructs a sample frame of physician practice units. From this sample unit, the BLS randomly chooses a bill that measures the net price paid to the practice for the entire set of services and procedures provided during an office visit, distinguished by payer type. The physician's output from this visit is represented by the content of the patient's bill, including CPT codes associated with that visit, and is related to a particular ICD-9 diagnosis. Given this sample bill, the BLS contacts the physician practice unit each month and asks it to reprice what the current net transaction prices would be for that particular fixed bundle/payer of services. Since the organization of physician practices has undergone considerable upheaval and consolidation in the last few years, however, sample attrition in the BLS physician services PPI has been considerable, and maintaining a sufficiently large response rate for repricing has proven to be difficult.[8]

The hospital services PPI measures net prices paid to hospitals for the entire bundle of services received during a hospital stay associated with a particular ICD-9 diagnosis, given the payer type. Hospital output is represented by the content of a patient's bill, including all charges for room, medical supplies, drugs, and ancillary services provided the patient during a single hospital stay.

The sampling universe for the hospital services PPI is taken from the American Hospital Association, stratified by hospital size, public versus private ownership, and type of medical specialty. Once a hospital is identified as a sample unit, the BLS chooses a fixed subset of DRGs, and each hospital is then asked on a monthly basis to report on net transaction prices of a single representative patient bill (typically the last patient bill on file for that DRG). Because the identical treatment bundle is not always observed in subsequent months, BLS reporters construct subsequent fictitious DRG bundle prices by repricing the identical inputs.[9]

8. For further discussion, see Fixler and Ginsburg (1997).

9. For further details, see Catron and Murphy (1996), Fixler and Ginsburg (1997), and U.S. Department of Labor, Bureau of Labor Statistics (n.d.).

It is worth noting that with both the physician services and hospital services PPIs, BLS use of fixed itemized components for obtaining price quotes does not capture major input substitution of treatment for a condition, such as the changing mix of psychotherapy and psychotherapeutic drugs used for the treatment of acute phase depression. Such a zero-substitutability definition of physician and hospital output leads to the existence of a substitution bias, but that bias is of course not confined to medical PPIs, for it pervades the entire fixed-weight Laspeyres price index computational procedures.

Finally, for our purposes it is also useful to note that the BLS publishes a PPI for prescription pharmaceuticals. Although pricing the output of prescription pharmaceuticals presents some particularly interesting issues involving treatment of generic drugs and of quality improvements in new products, those issues are beyond the scope of this paper.[10]

12.2.2 BLS Procedures for Medical Care Related CPIs

According to the BLS, the CPI is "a measure of the average change in the prices paid by urban consumers for a fixed market basket of goods and services."[11] Based on data from its Consumer Expenditure Surveys (CEX), the BLS identifies and defines a fixed "market basket" of goods, employing a classification system known as the item structure, which is updated approximately every ten years. In January 1998 the BLS introduced its most recent major item structure and fixed weight revisions, based on the 1993–95 CEX; from January 1987 until January 1998, fixed weights had been based on the 1982–84 CEX.

In contrast to the BLS's recently initiated medical PPI program, the medical CPI has been published regularly since 1935.[12] Currently the BLS publishes a monthly aggregate medical care CPI (MCPI), subindexes for medical care commodities and medical care services, as well as for prescription drugs and medical supplies, nonprescription drugs and medical supplies, physician services, hospital and related services, and health insurance. Each of these medical related CPIs is based on consumers' OOPPs, and thereby excludes all medical payments by governments, as well as employers' contributions to employee health insurance; only medical OOPPs plus that portion of third-party insurance paid for out-of-pocket by employees is included within the scope of the MCPI. Thus, although national health spending in 1996 constituted 13.6 percent of GDP, the total weight given medical care items in the CPI in 1997 is only 5.4 percent.[13]

10. For further discussion, see Berndt, Cockburn, and Griliches (1996), Griliches and Cockburn (1994), Kanoza (1996), and Kelly (1997).

11. U.S. Department of Labor, Bureau of Labor Statistics (1992), 176.

12. For historical discussions, see Langford (1957) and Getzen (1992).

13. Levit et al. (1998); Ford and Ginsburg (1997).

Recently the medical CPI has introduced a number of changes, some of which are similar to those implemented several years earlier in the PPI. Until at least 1990, in most cases the MCPI priced specific input items at list prices (e.g., "chargemaster" fees for x-rays, laboratory tests, and physicians' office visits) rather than at the average actual transaction price for treatment of, say, a child's forearm fracture to a managed care organization obtaining a hospital discount.[14] Since 1993 the BLS has attempted to obtain hospital transaction rather than list prices, but through 1996, efforts "yielded slow progress to date" (Cardenas 1996b, 40). Beginning with the January 1998 major revisions, more aggressive efforts have been made to obtain actual hospital prices by payer, but information is not yet available on the composition and nature of hospital quotes now being obtained. While DRG hospital quotes have been considered for use in the MCPI, the BLS is instead contemplating pricing "package" treatments, consisting of "highly standardized and tightly defined components and risk factors" for conditions such as appendectomies, tonsillectomies, and cataract surgery (Cardenas 1996b, 40). Details on how such treatment packages would be defined have not been released. Finally, while revisions involving hospital service components of the MCPI have been announced, the BLS has not yet published comparable information on revisions to the physician service components of the MCPI.[15]

12.2.3 BLS Procedures for Pricing Medical Treatments: Comments

For some time now, health economists and government statisticians have pointed to directions toward which the pricing of medical care services should move. Here we briefly summarize several of the directions suggested by this literature. First, as early as 1962 Anne Scitovsky proposed "an index which would show changes, not in the costs of such items of medical care such as drugs, physicians' visits, and hospital rooms, but in the average costs of the complete treatment of individual illnesses such as, for example, pneumonia, appendictis, or measles."[16] In Scitovsky (1967), this treatment episode approach was implemented on an illustrative basis for five medical conditions. Discussions of shortcomings and biases inherent in the BLS MCPI approach, and of the preference for the treatment episode approach to medical price measurement, have appeared steadily since 1967; see, for example, the chapter "Measuring Changes in the Price of Medical Care" in various editions of the well-known health economics textbook by Paul Feldstein (1979, 1983, 1988) and the Baxter Foundation Prize address by Joseph Newhouse (1989).

Second, experts agree, the measurement of changes in the price of medi-

14. See, for example, Armknecht and Ginsburg (1992) and Cardenas (1996a, 1996b).

15. For a discussion of the pricing of prescription pharmaceuticals in the MCPI, and treatment of generic drugs, see Armknecht, Moulton, and Stewart (1994), and U.S. Department of Labor, Bureau of Labor Statistics (1995).

16. Scitovsky (1964); also see Scitovsky (1967).

cal services should incorporate major quality changes, such as those involving adjustments for improvements in health care outcomes. In 1996, for example, Paul Armknecht noted that "A new dimension needs to be included in the pricing of medical services that includes outcomes, so that if cancer treatment results in improved survival rates, this is reflected in the index" (33). The Boskin Commission final report went further, asserting that "we strongly endorse a move in the CPI away from the pricing of health care inputs to an attempt to price medical care outcomes" (U.S. Senate Finance Committee 1996, 60). Although incorporation of quality and outcomes changes into medical price indexes presents some significant conceptual challenges[17] and implementation difficulties, it is clear that failing to address these aspects is likely to result in price measures that are unreliable and inaccurate.

Third, use of fixed weights over extended periods such as ten to twelve years in constructing sub- and aggregate price indexes is particularly inappropriate in the medical care sector, where both institutional and technological changes are rapidly occurring. Note that the frequency with which fixed weights are updated is distinct from the issue of which index number formula (e.g., Laspeyres, Paasche, or Törnqvist approximation to the Divisia) is preferable. For the rapidly changing medical care sector, decennial updates of fixed weights with old weights being fifteen years out of date before the new revision occurs (i.e., use of 1982–84 CEX weights through December 1997) results in price indexes whose accuracy and reliability can legitimately be called into question.

In the research reported here, we extend previous research aimed at constructing CPI- and PPI-like medical price indexes that deal with prices of treatment episodes rather than prices of discrete inputs, that are based on transaction rather than list prices, that take quality changes and expected outcomes into account, and that employ more current expenditure weights in the aggregation computations. Before describing the components of this research, however, we digress to discuss the illness whose treatment price we measure, namely, acute phase major depression.

12.3 The Nature and Prevalence of Major Depression

Depression is commonly characterized by melancholy, diminished interest or pleasure in all or most activities, sleep disorders, and feelings of worthlessness. In order for a patient's condition to be considered an episode of major depression, a clinician must determine that a very specific set of clinical criteria have been met. According to the fourth edition of the *Diagnostic and Statistical Manual* (DSM-IV) of the American Psychiatric Association (APA), major depression is diagnosed when the following is observed:

17. See, for example, the discussion between Gilbert (1961, 1962) and Griliches (1962).

The presence of one of the first two symptoms, as well as at least five of nine total symptoms. The symptoms must be present most of the day almost every day, for at least two weeks. The symptoms include:

1) depressed mood most of the day nearly every day;
2) markedly diminished interest or pleasure in almost all activities most of the day;
3) significant weight loss/gain;
4) insomnia/hypersomnia;
5) psychomotor agitation/retardation;
6) feelings of worthlessness (guilt);
7) fatigue;
8) impaired concentration (indecisiveness); and
9) recurrent thoughts of death or suicide. (APA 1994, 161)

Two dimensions of depression involve its persistence. A single episode of the illness is self-explanatory given the above diagnostic criteria. A recurrent depression is defined by two or more major depressive episodes each separated by at least eight weeks of return to usual functioning (APA 1993). Episodes of illness come and go, last from several weeks to several months, and are followed by periods of relatively normal mood and behavior. Untreated, the average depressive episode lasts from four to six months. Although the vast majority of individuals who experience an episode of major depression return to their original level of functioning, between 20 and 35 percent experience persistent symptoms; when these persistent symptoms last for twenty-four months or longer, these cases are referred to as chronic depression.[18] Approximately 50 percent of all people having a depressive episode can be expected to have a recurrence, usually within two or three years. Once an individual has a second episode, additional recurrence is 70 percent likely (APA 1993). The lifetime average for depressive episodes is five to seven, but as many as forty episodes have been reported (Papolos and Papolos 1992, 7).

A number of studies have shown that depression has similar or greater functional impairments than those attributed to other episodic and chronic medical illnesses.[19] Episodes of depression can be classified according to severity: mild, moderate, or severe. Mild depression typically involves the minimum number of symptoms required to meet clinical criteria and minor functional impairment. Moderately severe episodes are characterized by an excess in symptoms above the minimum to meet clinical criteria and by greater degrees of functional impairment. Severe major depression involves a number of excess symptoms above the minimum and significant degrees of functional impairment including the ability to work or conduct usual activities.

Epidemiological research indicates that in the early 1990s, 10.3 percent

18. See Keller et al. (1984), and Keller, Lavori, Rice, et al. (1986).
19. See, for example, Broadbent et al. (1990); Wells et al. (1989); Hays et al. (1995).

of the U.S. population met the criteria for major depression at some time during a twelve-month period (Kessler et al. 1994). Depression is often accompanied by other forms of ill health, such as anxiety, eating disorders, substance abuse, or other medical conditions (Kendler et al. 1995). Although the reasons are still not well understood, women—particularly women under age twenty-five—are much more likely to suffer from depression than are men; the relative lifetime female/male prevalence rate is about 1.7. Rates of recurrence and chronicity for major depression appear to be no different for women and men (Kessler et al. 1993).

12.4 Alternative Treatments for Acute Phase Major Depression

The acute phase of major depression is typically defined as the stabilization of most acute symptoms. In practice, standards of care typically identify the acute phase as occurring over a twelve-week period. In this research we allow protocol levels of acute phase treatment to occur over a six-month period to recognize that actual practice departs from the controlled environment of efficacy research. In clinical practice, continuation therapy frequently follows on acute phase treatment in hopes of preventing relapse. If a recurrent episode of major depression occurs, maintenance treatments are often initiated to prevent further recurrences (Kupfer 1991).

Given the information available in claims data, distinguishing between acute and continuation phase treatment is difficult. We employ acute phase standards of care to establish expected outcomes, but in our empirical implementation we undoubtedly will mix acute and other phases of care. Clinical research on the continuation phase of treatment is less developed, and definitive protocols have not been as widely adopted in many clinical settings.

12.4.1 Recent Developments in Psychotherapy and Antidepressant Medication

Treatments for acute phase major depression have advanced considerably during the last two decades. In the area of psychotherapy, a variety of new techniques has expanded treatment options well beyond traditional psychodynamic or psychoanalytic approaches. Interpersonal therapy (IPT), behavior therapy (BT), family therapy, and cognitive behavior therapy (CBT) are all relatively new. Evidence from controlled clinical trials suggests that when applied as the single mode of treatment for less severe forms of acute phase major depression, each of these therapies reduces depressive symptoms. Moreover, relative to antidepressant medication as the sole treatment, each has generally been shown to perform at comparable levels of efficacy and to have similar outcomes.[20]

20. See Beck et al. (1979), Elkin et al. (1989), Frank et al. (1990), Kupfer et al. (1992), and Beach, Sandeen, and O'Leary (1990).

Extraordinary advances have been achieved in the last two decades in the area of antidepressant medication. Although very recent developments expand the therapy set even further, over the 1991–95 period examined here three general classes of antidepressant medications were employed. These are (1) cyclic antidepressants that include the widely used tricyclic antidepressants (TCAs) and a number of lesser-known drugs such as trazodone; (2) selective serotonin-reuptake inhibitors (SSRIs), which include brand name drugs such as Prozac, Zoloft, Paxil, and Luvox;[21] and (3) monoamine oxidase (MAO) inhibitors which, due to considerable side effects and dangerous interactions, are generally used only for patients resistant to other forms of treatment.

Newer SSRIs offer some distinct advantages over older TCAs, although in randomized controlled trials clinical efficacy rates are similar. SSRIs are associated with lower risk of overdose, and reduced levels and numbers of side effects. Important side effects frequently associated with TCAs include drowsiness, dry mouth, impaired ability to concentrate, seizures, and weight gain.[22] The side effects most prominently associated with use of SSRIs relate to sexual dysfunction (particularly for males) and anxiety. The advantages of SSRIs come at a significantly higher pecuniary cost than most TCAs.

Psychotherapeutic interventions have frequently been combined with antidepressant medication as a combination strategy for treating major depression. The specific interventions that have been most intensively studied are the use of TCAs in combination with either IPT, BT, or a general unspecified form of short-term psychotherapy. To date, no clinical studies have been reported in the literature which systematically assess the combination of the newer SSRIs and psychotherapy.[23] It is generally presumed, however, that such combinations will be at least as efficacious as the combination of TCAs and psychotherapy.

Finally, electroconvulsive therapy (ECT) is an effective treatment, but ECT is typically limited to rather special circumstances when the patient's depression is severe and is complicated by a number of other psychiatric symptoms including psychosis, catatonic stupor, or high risk of suicide.

In the analysis reported below, we focus on outpatient treatments for major depression; these constitute the vast majority of treatment episodes (75 to 80 percent). We do this for several reasons. First, inpatient claims data typically do not contain information on the drugs prescribed for treat-

21. Recent variations slightly distinct from the SSRIs include the brand name drugs Effexor, Serzone, and Remeron.

22. Considerable variation exists in side effects among the TCAs; see Berndt, Cockburn, and Griliches (1996, 142–43, table 1) for details.

23. Using MEDSTAT retrospective claims data, Croghan et al. (1999) provide evidence suggesting that the combination of SSRIs and psychotherapy is more effective than psychotherapy alone in patients receiving continuous treatment. For a discussion of related SSRI-psychotherapy research, see Wilde and Benfield (1998).

ment; thus characterizations of inpatient care are inherently incomplete.[24] Second, because of other incomplete information regarding illness severity and comorbid conditions, it is difficult to use administrative claims data to characterize fully an illness diagnosis, and therefore to make judgments about the appropriate use of hospital services for treating major depression. Third, few clinical trials specifically address inpatient treatment for major depression, making it difficult to assign outcomes to treatments. Finally, because there was considerable evidence of overuse of hospital services in the aggregate during the late 1980s and early 1990s, the inclusion of hospital services in our 1991 base year could make interpretation of price changes troublesome.[25] This strategy of limiting severity of cases by excluding individuals hospitalized for treatment of depression is likely to be only partly successful. During the 1990s there were a substantial reduction in inpatient psychiatric admissions and in the length of stay for hospitalized cases. The implication of this is that the population of people treated only on an outpatient basis may be getting sicker over time.

In our analysis, we therefore focus on the use of various antidepressant medications alone, several forms of psychotherapy alone, and several drug-psychotherapy combination treatments. Because of very small sample sizes, we do not incorporate treatment involving ECT or the MAO inhibitors.

12.4.2 Results from Comparative Efficacy Studies

We now provide a brief summary of research results comparing the efficacy of alternative treatments for depression of varying severity. We divide depression severity into two classes: severe and less severe (hereafter, mild). We have reviewed approximately thirty major clinical trials and meta-analyses from the clinical literature on comparative efficacy of acute phase treatments (Busch, Frank, and Berndt 1996). This literature points to several key conclusions.

First, psychotherapies of all kinds have been shown to result in superior outcomes compared to no treatment. When compared amongst themselves, the different forms of psychotherapy appear to have no significant differences in outcomes.[26]

Second, for less severe forms of depression, psychotherapies alone, TCAs with medical management, and SSRIs with medical management appear to produce comparable outcomes. Each of these therapies produced significantly better outcomes than placebo treatments. Versions of these results have been reported in numerous large treatment trials and by

24. A significant portion of inpatient episodes have unspecified outpatient follow-up, thereby limiting that avenue for identifying treatments.

25. For a discussion, see Mechanic (1989) and McGuire (1989).

26. The AHCPR Depression Guidelines Panel (U.S. Department of Health and Human Services 1993) provides a summary and interpretation of the evidence on this point.

meta-analyses of smaller clinical trials. Combination treatments with these as components also generate equivalent levels of efficacy for less severe forms of depression.

Third, for more severe forms of depression, the bulk of the evidence suggests that TCAs alone, SSRIs alone, and combinations of drugs and psychotherapy have comparable levels of efficacy, and each results in superior outcomes compared to psychotherapy alone. Recently some evidence has emerged showing some extra improvement from the combination treatments relative to medication alone (APA 1993). We believe it is premature to conclude that combination treatments offer significantly higher levels of efficacy than do antidepressant medications alone (or with medical management, as is typically the case).

Based on these observations from the literature, we view all the major treatment technologies as offering comparable expected outcomes for the average care of less severe acute phase depression. For severe depression, we view TCAs and SSRIs alone as comparable to each other and to combinations of TCAs and SSRIs with psychotherapy.

12.5 Identifying Comparable Treatment Bundles from Claims Data

The results from our review of the clinical trials literature enable us to develop a set of treatment "bundles" that group together therapeutically similar treatments of a specific form of major depression. This identification of treatment bundles that result in similar expected health outcomes is a crucial step in the construction of medical treatment price indexes that take expected outcomes into account. The implicit assumption we adopt here is that obtaining therapeutically similar outcomes from alternative treatments provides a useful approximation to achieving similar expected utility levels.[27]

12.5.1 The MEDSTAT Data

To identify empirically comparable treatment bundles for acute phase major depression, we employ a data set consisting of retrospective insurance claims from four large self-insured employers that offered twenty-five health plans to 428,168 employees and their dependents for the years 1991 through 1995. The data were obtained from MEDSTAT, Inc., and contain information on prescription drug claims, inpatient hospital treatment, outpatient visits, ICD-9 diagnoses, CPT4 procedures, and the demographic characteristics of all covered individuals. The health benefits offered to

27. We recognize this is only an approximation. This is particularly the case for depression, where the constellation of side effects across treatment can vary significantly and can lead to differential patient compliance and patient preferences between the SSRIs and TCAs. See Crown et al. (1996) and Wilde and Benfield (1998) for evidence on differential TCA-SSRI compliance among patients.

enrollees in this database are quite generous relative to the general market for private health insurance in the United States.

During the five years we observed, important changes occurred in the terms of insurance coverage for mental health care. While the vast majority of plans represent so-called managed indemnity plans (90 percent–94 percent), the management of mental health care benefits changed for a substantial number of enrollees between 1991 and 1995. Beginning on 1 January 1994, about 8 percent of enrollees had their mental health coverage "carved out" to a specialty mental health managed care company.[28] In January 1995 an additional 35 percent of enrollees had their mental health benefits carved out.

It is reasonable to expect that these carve-out arrangements affected both the input prices and the quantities of specific delivered services, such as visits.[29] Changes in the extent of carve-out arrangements might also have affected the general clinical strategies used in treating major depression. Recent analyses by Wells et al. (1996) and by Berndt, Frank, and McGuire (1997) show clear differences in treatment patterns between carve-out managed care plans and treatment of mental health care financed by general health insurers, with carve-out arrangements being associated with a higher likelihood of using prescription drug treatments.

Each outpatient and prescription drug claim can accommodate two ICD-9 diagnostic codes. In identifying cases of major depression, we used ICD-9 codes 296.2 (major depressive disorder, single episode) and 296.3 (major depressive disorder, recurrent episode) to define patients diagnosed with major depression. We do this for three reasons. First, the clinical trials literature has for the most part employed these definitions in their entry criteria. Second, chart reviews have indicated that the specificity of these two diagnoses is high (i.e., the proportion of true positives and false negatives is high, while the proportion of false positives is very low). Third, clinicians could employ a more ambiguous diagnosis such as "depression not elsewhere classified" or "neurotic depression." That clinicians designated their diagnosis as either 296.2 or 296.3 indicates a conscious act of volition.[30]

Using the diagnostic information and dates contained in the claims, we construct episodes of treatment. Because we do not directly observe symptoms in retrospective claims data, we cannot make our claims-based definition of an episode of treatment correspond directly to an episode of

28. Specialty carve-out management occurs when a portion of the health risk is managed separately from the rest of health care. See Frank, McGuire, and Newhouse (1995) for further discussion.

29. For discussion, see Goldman, McCulloch, and Sturm (1998), and Ma and McGuire (1998).

30. Thus we exclude other 296 ICD-9 diagnoses, depression not elsewhere classified, as well as some other broad depression-related conditions such as neurotic depression.

the illness.[31] When claims data indicate that psychotherapeutic drugs were prescribed, we consider the number of days of treatment provided by the prescription as the time period over which an individual received care. We follow American Psychiatric Association (1993) guidelines in defining an episode of depression as new if a diagnosis is preceded by a period of at least eight weeks of not meeting clinical criteria for depression.[32] Thus we use an eight-week period without treatment to define new treatment episodes. In preliminary analyses, we experimented with alternative definitions, such as those involving six- or twelve-week intervals; results were essentially unaffected. For treatments lasting longer than six months we count quantities of visits and drugs meeting guideline standards for acute care occurring with the first six months of care (e.g., fifteen psychotherapy visits even if year totals were twenty-five visits). To ensure that we consider the full set of claims associated with the acute phase of treatment, we exclude episodes beginning in the last six months of 1995, the last year in our sample.

When these criteria were applied to the MEDSTAT data set, we defined 18,920 episodes of acute phase outpatient care over the 1991–95 period. Because we cannot fully observe the treatment received for censored cases, we confine our attention to the 15,750 uncensored episodes in which we observe at least eight weeks without treatment before the beginning of an episode; as noted above, 1,548 episodes beginning in the last six months of 1995 were also excluded. In order to limit the sample to less severe forms of depression, we eliminated individuals with episodes involving inpatient hospital treatment for a mental illness any time during the five years. This reduced the number of episodes to 10,067.

Using information on procedures (e.g., type of visit, whether drug prescribed) as given by the CPT4 codes, we can describe the composition of treatment that occurred within a treatment episode. Drug treatment is based on the national drug codes (NDC) reported on the claim. The NDC classification of antidepressant medications revealed use of seven TCAs, three SSRIs, two other serotonin-related drugs,[33] three MAO inhibitors, four anxiolytics, and four heterocyclics for treatment of depression. In terms of composition, 54 percent of the drug claims involved SSRIs, 19 percent TCAs, 19 percent anxiolytics, and 7 percent heterocyclics.[34]

In previous research, we have reported that in the MEDSTAT claims

31. For discussion of defining episodes of care, see Keeler et al. (1986) and Wingert et al. (1995).

32. We count days without treatment only after the number of days of supply in a drug prescription has been exhausted, thereby assuming full compliance with the daily recommended dosage.

33. These were brand name drugs Effexor and Serzone.

34. These compositional figures are consistent with IMS aggregate national sales data for antidepressant medication over this time period, as reported by Berndt, Cockburn, and Griliches (1996).

database, the share of treatment accounted for by treatment involving psychotherapy claims appeared to increase significantly in 1995, after having fallen steadily from 1991 to 1994.[35] Subsequent research has indicated to us that while procedure codes are missing for many outpatient claims—a common problem with claims data, the extent of missing claims in the MEDSTAT data set is particularly large in 1991–94. Thus, the apparent sudden increase in psychotherapy claims in 1995 could simply reflect considerable missing claims from 1991 through 1994. Hereafter we call the data underlying this previous research our "old" data set.

To identify missing psychotherapy claims not explicitly delineated by CPT4 codes, for our "new" data set we have developed an algorithm in which a psychotherapy procedure code is assigned to claims with missing CPT4 codes if two visits are within fourteen days of each other, if they have the same charge, and if a previously psychotherapy-identified claim in the episode has the same charge.[36] Use of this algorithm contributed to an increase in the number of identified episodes.

Our primary analysis of the claims data considers only "pure" or "strict guideline" treatments. That is, initially we will only consider episodes of care that adhere strictly with treatment definitions as tested in the clinical trials literature. In this way we can directly link the "price" of an episode of a well-defined treatment to the price of other therapeutically similar treatments. However, we will also report results based on findings derived from episode treatment definitions that relax this stringent guideline restriction, in which we include treatments slightly below guideline criteria and those involving some continuation phase treatment.

In our strict guideline analyses we identify seven major classes of treatments that have been proven effective in the treatment of depression: (1) psychotherapy alone, six to fifteen visits "PT alone"; (2) short-term TCA treatment alone or with medical management—"TCA alone"; (3) short-term SSRI treatment alone or with medical management, 31–180 days—"SSRI alone"; (4) short-term TCA treatment (31–180 days) with some psychotherapy—"TCA+PT"; (5) short-term SSRI treatment (31–180 days) with some psychotherapy—"SSRI+PT"; (6) short-term combined TCA/SSRI treatment (31–180 days) with some psychotherapy—"TCA/SSRI+PT." With the exception of psychotherapy alone, episodes with anxiolytics also being prescribed are combined into the appropriate class defined above. Use of seven major classes of treatments in this new data set represents an expansion from the five used in our previously published research. Another significant change in the new data was to include epi-

35. See Frank, Berndt, and Busch (1999) and Frank, Busch, and Berndt (1998).

36. Additional checks were done to ensure that these newly identified psychotherapy visits were not in fact instead a sequence of medical management visits. Robustness checks revealed that only a small number of claims were reclassified when cost thresholds of about $50 were used to mark medical management from psychotherapy.

sodes that involved long-term treatments or those extending beyond twenty visits or six months. For those episodes we included the components of care consistent with the acute phase bundles specified above and excluded spending on the remainder of the "episode." Using this expanded definition of an episode, we are more likely to include continuation treatment in the analysis. This may be especially true in the later years as new recommendations about continuation treatment diffused into practice. Thus, the new data set may be "pricing" acute plus early continuation phase care in the later years. This would tend to bias an index upward. Evidence for this is reflected by an upward trend in the average duration of SSRI treatment during the five years (from 87 days in 1991 to 126 in 1995). This change contributed to a substantial increase in the number of cases used in the analyses.

The new data set also improved the definition of censoring. In the "old" data set, to limit our analysis to episodes for which we had information on the entire episode, all cases which included treatments in the last eight weeks of 1995 were eliminated. This means that more episodes were eliminated from 1995 than from any other year. More severe episodes in need of longer term treatment may therefore have been disproportionately dropped from previous analyses. This may have biased the price index downward. We have now corrected that "late '95" censoring problem by limiting our analysis to the first six months of treatment. As mentioned previously, to ensure we consider all treatments in the first six months of the episode, we do not consider episodes begun in the last half of 1995.

Finally, the new data set allows drug switching if it is consistent with standards of care in clinical trials. Again, the change would tend to increase the severity and complexity of the treated population. Drug switching has been increasing over time as new pharmaceutical agents have become available (SSRIs and serotonin-norepinephrine reuptake inhibitors [SNRIs]).

It is informative to compare episodes defined in our new data set with those in the old. As seen in table 12.1, the most obvious change is that the number and proportion of episodes receiving guideline-compatible treatments increase very substantially, from 2,348 to 5,039, or from 23 percent to 50 percent. There are 2,980 guideline-compatible episodes in the new data set that were not in the old, and 289 in the old that do not appear in the new.

There is also some evidence suggesting that in the new data set, a greater number of episodes are associated with other mental health comorbidities. While new-old differences are relatively small in terms of substance abuse comorbidities in guideline-compatible episodes (2.5 versus 2.8 percent), 17.9 percent of acute phase depression episodes in the new data set simultaneously involve treatment of schizophrenia or bipolar depression, while in the old data set this proportion was a smaller 15.8 percent. For the various panic and anxiety comorbidities, the new-old proportions are 3.3

Table 12.1 Comparison of Data Sets Based on New and Old Episode Definitions

	New Data	Old Data	In New but Not in Old Data	In Old but Not in New Data
Total episodes	10,067	10,368		
Episodes guideline compatible	5,039	2,348	2,980	289
Mental health comorbidities (%)[a]				
Substance abuse	2.5	2.8	2.6	4.2
Schizophrenia/bipolar	17.9	15.8	20.1	23.2
Anxiety/panic	3.3	2.9	3.9	6.2
Sum of other medical comorbidities (%)[b]				
0	48.2	48.3	47.5	41.5
1	37.9	36.2	39.3	38.8
≥2	13.9	15.6	13.2	19.7

[a]Panic/anxiety diagnoses are ICD-9 300, 300.1–300.4; schizophrenia/bipolar are all 295s, 296.0–296.1, 296.4–296.8; other includes obsessive/compulsive disorders (300.3), bulimia (783.6), depression not otherwise specified (311), and dysthymia (300.4).

[b]Other medical comorbidities are sense organs, circulatory system, cerebrovascular, digestive system, kidneys, prostate, pregnancy, and nervous system.

versus 2.9 percent. The literature, as well as our own discussions with clinicians, suggests that (1) major depression and bipolar illness are often mistaken for one another, and (2) treatment of depression with comorbid psychoses and substance abuse is both common and difficult. We interpret differences between the new and old data sets as therefore moving us away from the more pristine world of clinical trials to a more realistic and more complicated treatment environment in which mental illness comorbidities are more prevalent.

It is worth noting that differences between the new and old data sets in terms of other, nonmental health comorbidities are rather minor. As seen in the bottom panel of table 12.1, for example, the proportion having no such comorbidities is 48.2 percent in the new data set, and 48.3 percent in the old; relative to the old data set, the proportion having one other medical comorbidity is slightly higher (37.9 percent versus 36.2 percent), but is slightly lower for two or more comorbidities (13.9 versus 15.6 percent).

12.5.2 On Guideline Compatibility

Earlier we noted that of the 10,067 episodes identified in the new data set from 1991 through 1995, 5,039 (50 percent) were guideline compatible. It is interesting to note that in our new data set the proportion of episodes receiving guideline-compatible treatment rose from 35 percent in 1991 to 42 percent in 1992, then increased sharply in 1993 to 56 percent, and remained at 56 percent in 1994 and 1995.

The interpretation of guideline-incompatible treatments is difficult with

claims data. For example, 1,308 episodes (33 percent of the 3,965) treated with psychotherapy alone consisted of a single visit. In addition, 1,672 or 16 percent of all episodes received neither psychotherapy nor an antidepressant drug. Such single visits might have taken place for the purpose of "ruling out" major depression as the relevant condition to be treated in favor of a somatic condition or another mental disorder. In such cases, the visit should not be viewed as "inappropriate treatment" but rather as an appropriate assessment, consistent with the Depression Guideline Panel statement (1993, 36): "Effective treatment rests on accurate diagnosis. The practitioner must first determine whether the patient has a clinical depression or is simply suffering normal sadness or distress. . . . For patients who have very mild cases of major depression or whose diagnosis is unclear and who are not in immediate danger or are not suffering significant functional impairment, the practitioner may want to schedule one or two additional weekly evaluation visits to determine whether symptoms will abate without formal treatment. . . ." The implication of this is that with retrospective claims data, distinguishing treatment and assessment is quite uncertain. Hence we are somewhat unclear as to whether only 50 to 60 percent of care lies on the production frontier, or whether another 10 to 20 percent of treatments were dealt with properly but did not require treatment of the type studied here. It is important to note here that our use of guideline standards of care imposes a rather unrealistic shape on the production function for treatment of depression. It takes on a step function form. For example, if one were to receive six psychotherapy visits for treatment of depression, our analysis would treat the case as "effective," whereas four or five visits would be viewed as "ineffective." This is unlikely to be an accurate representation of clinical reality. Nevertheless, there is little systematic analysis upon which to make alternative assumptions. Thus, we use the step function production model as a point of departure.

With these difficulties in appropriate interpretation as a caveat, we now examine in greater detail the extent of guideline compatibility in the various treatment bundles of the new data set.

Of the 10,067 episodes, 3,765 were treated with psychotherapy alone, and 376 with psychotherapy and anxiolytics. Within the claims data, no distinction was made between different types of psychotherapy. Considering the 3,044 treatments involving fifty-minute psychotherapy sessions, we find that 1,658 received four or fewer visits, and 800 only a single visit. Of the 921 episodes involving twenty-minute psychotherapy alone, 508 involved just one brief visit, while 178 episodes were treated with between two and five brief sessions. An additional twenty episodes involved some form of group therapy treatment, with only 25 percent of those having more than three visits.

While clinical trials data indicate that individuals show partial response to psychotherapy with six weeks of treatment (with weekly sessions) and

remission in twelve weeks, published guidelines for the treatment of acute phase depression do not indicate any demonstrated effectiveness for fewer than six visits. The benefits of short psychotherapy visits, in the absence of antidepressant medication, have not been studied and therefore cannot be considered either effective or ineffective treatment. Although clinical trials and published treatment guidelines indicate psychotherapeutic treatment alone is an effective treatment, in our data, of the 3,044 episodes given this treatment, only 1,386 episodes (46 percent) can be considered to have completed a psychotherapy regimen that is consistent with guideline treatments.[37]

Turning attention now to treatments involving antidepressant medications, we begin by noting that claims data do not include information on how many days medication was actually taken; as a proxy, we use the number of days of treatment for which a prescription was filled.

Of the 10,067 episodes considered, 231 were treated with TCA either alone (176) or in combination with anxiolytic medication (55). Of these 231 episodes, 55 (24 percent) were treated with fewer than thirty days of medication and 29 were treated with ten or fewer days. Generally, the clinical literature (and the APA guidelines) indicates that while patients may show some improvement from antidepressant medication by the end of the first week, full response to acute phase depression may take four to six weeks.

An additional 741 episodes were treated with SSRIs either alone (604) or in combination with anxiolytic medication (137); 77 (10 percent) of these episodes were treated with fewer than thirty days of medication. Because individuals have differing reactions to drugs, some individuals are appropriately treated with one class of antidepressants, and then switched to another class. In our sample, 95 episodes were treated with both TCAs and SSRIs, but 20 of these episodes had less than thirty days of both drugs.

Several episodes were treated with medications other than SSRIs and TCAs. In our sample 10 episodes were treated with MAO inhibitors, while 164 were treated with heterocyclics. An additional 124 episodes were treated with antianxiety medication alone, a protocol which has not been approved by the FDA for the treatment of major depression. The use of antianxiety medication in the treatment of depression remains controversial. The use of alprazolam may be appropriate, if other medication is contraindicated. There is no clinical trial evidence for the efficacious use of other anxiolytic medications.[38]

Finally, in terms of combination treatments, the share of episodes treated with psychotherapy and antidepressant medication grew over time. Over the entire five-year period, 552 treatment episodes involved both

37. For a related discussion, see Katon et al. (1992).
38. See Wells et al. (1994, 1996) for further discussion.

some TCA and some psychotherapy, while 2,169 included both SSRI and some psychotherapy. A large share (36 percent) of the episodes treated with combination treatments had three or fewer psychotherapy visits, and an additional 382 episodes were treated with some TCA, some SSRI, and some psychotherapy.

12.5.3 Comparison with Results from Other Studies

The patterns of care observed in this data set raise issues related to the likely effectiveness of treatment, given the substantial proportion of episodes involving guideline-incompatible treatments. One potential criticism of the patterns of treatment bundle data presented above is that they are based on retrospective claims data. Claims data are useful in that the retrospective medical treatment of many individuals can be analyzed efficiently and at minimal expense. In addition, such observational data reveal the "real world" practice of medicine, not the pristine clinical trial environment. Claims data are also used for quality assessment by organizations constructing "report cards" on health care organizations. Yet claims data have been fairly criticized for several reasons. The accuracy of diagnoses and recorded data are sometimes questioned, and omissions in records are common. For example, depression has been shown to be underdiagnosed by primary care physicians and overdiagnosed by psychiatric clinicians (Schulberg et al. 1985).

Other studies have found treatment patterns for depression that are generally consistent with the patterns we observe here. The Medical Outcomes Study (MOS) consisted of 635 individuals diagnosed with depression or with current depressive symptoms for whom data were collected by self-administered questionnaires, patient diaries, phone interviews, and health examinations. The MOS found that only 23 percent of depressed patients had used an antidepressant medication in the prior month or used it daily for a month or more in the prior six months. Of those patients using an antidepressant medication, 39 percent used an inappropriately low dose (Wells et al. 1994).

The MOS did not report number of psychotherapy visits. Instead it reported "counseling," defining it as three or more minutes of counseling during the screening interview. This makes comparison with published standards of care difficult. Although 90 percent of patients of mental health specialists were counseled, among general medical practitioners where most study participants were treated, only 20 percent of managed care and 40 percent of fee-for-service patients were counseled.

Another study of eighty-eight outpatients enrolled in the Clinical Research Collaborative Program on the Psychobiology of Depression of the National Institute of Mental Health (NIMH) found that prior to entry into the study, only 19 percent of patients received an adequate dose and duration of antidepressant medication, while 24 percent received some

antianxiety medication. Regarding psychotherapy visits, 44 percent were seen for at least one hour weekly (Keller, Lavori, Klerman et al. 1986).

Thus, the substantial proportion of patients in the MOS and NIMH studies apparently not obtaining efficacious treatments, while perhaps surprising, is consistent with the treatment patterns found in the claims data we observe here.

For our purposes of constructing price indexes for the treatment of acute phase major depression, we must decide whether to utilize the data suggestive of treatment not consistent with Food and Drug Administration (FDA) approvals and Agency for Health Care Policy and Research (AHCPR) guidelines. Since the interpretation of such treatments is problematic, in our "strict guideline" analysis we confine our attention to episodes of treatment defined as being consistent with AHCPR guidelines; however, in a separate analysis we also include treatment bundles that are "close" to guideline standards. Additional research on guideline-incompatible care and the shape of the production function for treatment of depression is currently under way.

12.6 Quantities and Prices of Alternative Treatment Bundles

We begin our empirical analysis by characterizing quantities and prices of the seven treatment bundles. In table 12.2 we report guideline-compatible treatment quantities by year. Censoring at the beginning and end of our time span implies that in 1991 and in 1995, about 1,500 episodes are identified, whereas in 1992–94 the number is considerably larger at about 2,400 per year. The two most common treatment bundles are SSRI+PT and PT alone. Because of varying sample size by year, quantity trends are more easily discerned by examining quantity proportions rather than absolute levels; guideline-compatible treatment bundle quantity proportions are reported in table 12.3.

A number of notable trends appear in table 12.3. First, there has been

Table 12.2 **Guideline-Compatible Treatment Bundle Quantities, by Year**

Treatment	1991	1992	1993	1994	1995	Total
PT alone	214	311	321	311	229	1,386
TCA alone	32	45	43	37	19	176
SSRI alone	52	113	188	205	110	668
TCA+PT	63	111	129	73	33	409
SSRI+PT	125	292	528	644	400	1,989
TCA/SSRI	9	14	23	23	6	75
TCA/SSRI+PT	27	53	121	98	37	336
Sum	522	939	1,353	1,391	834	5,039
Total episodes	1,479	2,211	2,426	2,468	1,483	10,067
Percent guideline-compatible	35.3	42.5	55.8	56.4	56.4	50.1

Table 12.3 Guideline-Compatible Treatment Bundle Quantity Proportions, by Year

Treatment	1991	1992	1993	1994	1995	Total
PT alone	0.409	0.331	0.237	0.224	0.275	0.275
TCA alone	0.061	0.048	0.032	0.027	0.023	0.035
SSRI alone	0.100	0.120	0.139	0.147	0.132	0.133
TCA+PT	0.121	0.118	0.095	0.052	0.040	0.081
SSRI+PT	0.239	0.311	0.390	0.463	0.480	0.395
TCA/SSRI	0.017	0.015	0.017	0.017	0.001	0.015
TCA/SSRI+PT	0.052	0.056	0.089	0.070	0.044	0.067

a very substantial decline in PT alone treatments, from 41 percent in 1991 to 22 percent in 1994, and then up slightly to nearly 28 percent in 1995. Second, for the medication only treatments, SSRI alone has grown from 10 to 13 percent, even as TCA alone declined from 6 to 2 percent; the sum of the two medication only treatments has remained relatively constant at 15 to 16 percent. Third, most of the compositional change among treatment bundles has involved the medication-psychotherapy combination treatments. While the TCA+PT combination fell from 12 to 4 percent between 1991 and 1995, the SSRI+PT treatment share doubled, from 24 to 48 percent of all treatments. By 1995, the SSRI+PT combination had become the modal treatment bundle.

In table 12.4 we report the average supply price for each treatment bundle, by year; recall that this supply price captures the sum of payments from insurers and patients/consumers to providers. As is seen there, annual price movements represent a mix of increases and decreases. Between 1993 and 1994, for example, the price of one bundle increases and that of all other six bundles decreases, while for 1994–95. price movements are reversed—six up, and one down. Of particular interest is the extent to which compositional changes in quantities of treatment bundles appear to be negatively related to changes in relative prices. When entries in tables 12.3 and 12.4 are considered together, it becomes clear that compositional quantity changes do not follow simple conceptions of downward-sloping demand curves where quantity and "price" changes are clearly negatively related. From 1991 to 1992, for example, even as the supply price of SSRI alone more than doubled from $52 to $113, the SSRI alone share increased from 4 to 6 percent; similarly, the cost of the SSRI+PT combination bundle increased by about 150 percent from $125 to $292 between 1991 and 1992, yet the quantity share grew from 24 percent to 31 percent. It is important to note that the quality of treatment bundles may vary (e.g., side effects). Also, insurance coverage drives a wedge between the supply and demand prices. Both of these factors may help explain observed patterns and they are discussed below.

This positive relationship between changes in supply price and quantity

share is observed in other years as well. From 1992 to 1993, for example, the supply price of TCA+PT fell by about 18 percent from $882 to $745, but its quantity share also dropped from 12 percent to 10 percent. Between 1993 and 1994, while the cost of PT alone fell about 5 percent from $866 to $818, the PT quantity share fell from 26 percent to 23 percent. Similarly, for both TCA alone and TCA+PT, between 1993 and 1994 both prices and quantity shares fell.

What these quantity and supply price data suggest, therefore, is that factors other than supply price are likely to have induced compositional changes among treatment bundles for acute phase major depression. Prominent among these, we conjecture, is the increased knowledge and experience gained by physicians on the efficacy and effectiveness of the SSRI medications, particularly in combination with a limited amount of psychotherapy and changing insurance arrangements.

Although compositional changes among treatment bundles may reflect a host of price and nonprice influences, it is clear that the dollar expenditure shares among the seven treatment bundles have been greatly affected between 1991 and 1995. These dramatic changes are evident in table 12.5. While the expenditure share going to PT alone fell from 45 percent to

Table 12.4 **Average Supply Price of Treatment Bundle, in Dollars, by Year**

Treatment	1991	1992	1993	1994	1995	Average[a]
PT alone	876	853	866	818	819	846
TCA alone	236	310	318	204	226[b]	267
SSRI alone	311	387	309	308	344	328
TCA+PT	932	882	745	706	941	820
SSRI+PT	941	983	949	958	972	961
TCA/SSRI	649[b]	361[b]	449	423	328[b]	439
TCA/SSRI+PT	1,002	1,021	1,062	953	977	1,010
Average[a]	805	817	802	788	825	

[a]Denotes a weighted average.

[b]Denotes a cell with less than twenty observations.

Table 12.5 **Total Expenditure Shares by Treatment Bundle, by Year**

Treatment	1991	1992	1993	1994	1995	Average
PT alone	0.446	0.346	0.256	0.232	0.273	0.310
TCA alone	0.018	0.018	0.013	0.007	0.006[a]	0.012
SSRI alone	0.038	0.057	0.054	0.058	0.055	0.052
TCA+PT	0.140	0.128	0.089	0.047	0.045	0.090
SSRI+PT	0.280	0.374	0.461	0.563	0.565	0.449
TCA/SSRI	0.014[a]	0.007[a]	0.010	0.009	0.003[a]	0.008
TCA/SSRI+PT	0.064	0.071	0.118	0.085	0.053	0.078

[a]Denotes a cell with less than twenty observations.

27 percent between 1991 and 1995, by 1995 the SSRI+PT bundle was responsible for more than half (56 percent) of all dollar expenditures, up from 28 percent in 1991. In 1991 the three single bundles of PT alone, TCA alone, and SSRI alone accounted for 50 percent of all treatment dollars in 1991, but by 1995 they only captured 33 percent of expenditures.

Together, the patterns observed in tables 12.2 through 12.5 suggest to us that rather than PT and the SSRIs being viewed as simple single substitutes in the production of treatments for acute phase depression, they are more accurately envisaged as being complementary in the sense that the PT+SSRI combination has now become the modal treatment.

To this point, our discussion of treatment bundle price has focused on the supply price—the sum of payments received from insurers and patients/consumers. The demand price notion reflects the fact that patients receiving treatment for acute phase depression provide direct payments in the form of copayments and deductibles. In table 12.6 we report average demand prices for each of the seven treatment bundles, annually. Trends in the demand price differ quite markedly from those of the supply price.

Recall from table 12.4 that the average supply price across all bundles fell slightly from $805 in 1991 to $788 in 1994, and then increased to $825 in 1995. In contrast, as seen in table 12.6, from 1991 to 1994 the average demand price increased from $100 to $105, and then it increased very sharply to $128 in 1995. This sharp increase of 22 percent between 1994 and 1995 in the average demand price across all treatment bundles reflects corresponding jumps in the PT alone (26 percent), SSRI alone (19 percent), SSRI+PT (17 percent), and TCA/SSRI+PT (21 percent) bundles.

Two main factors may account for changes in demand prices that differ so starkly from supply price changes. First are the changes in cost-sharing provisions. These are documented in the MEDSTAT data set. The MEDSTAT data indicated to us that in 1995, for a number of insurance plans, benefit changes were introduced that substantially increased the copayments and deductibles required of plan enrollees.[39] Second is the tendency to use out-of-network providers in the context of managed care. Coverage of care provided by nonnetwork providers typically carries higher copayments than does care received from a network clinician. During the time period observed, networks were becoming more restrictive and behavioral health carve-outs were introduced, which may have increased the share of care provided by nonnetwork providers. The impact of these benefit design changes can be seen by computing the portion of the total treatment cost

39. Data from Marion Merrell Dow's *Managed Care Digest* (1993–1994, 1996) indicate that over all HMOs surveyed, average copayment per prescription rose from $6.78 in 1993 to $7.43 in 1995 for brand name drugs, and from $4.75 to $5.22 for generics. While the MEDSTAT increases appear to be larger, they also include deductibles for psychotherapy visits, not just copayments for drugs.

Table 12.6 **Average Demand Price of Treatment Bundle, in Dollars, by Year**

Treatment	1991	1992	1993	1994	1995	Average[a]
PT alone	121	122	122	123	155	128
TCA alone	39	46	46	48	51	46
SSRI alone	23	30	24	27	32	27
TCA+PT	128	124	116	113	117	120
SSRI+PT	102	105	115	121	142	120
TCA/SSRI	43[b]	47[b]	46	51	21[b]	45
TCA/SSRI+PT	102	121	131	135	163	132
Average[a]	100	101	102	105	128	

[a]Denotes a weighted average.

[b]Denotes a cell with less than twenty observations.

borne directly by the patient, that is, computing the ratio of the demand price in table 12.6 to the corresponding supply price in table 12.4; we call this the OOPPs ratio.

For the SSRI alone bundle, the OOPPs ratio increased only marginally, from 7 percent in 1991 to 9 percent in 1995. For the very low total cost TCA alone bundle, the OOPPs ratio increased more markedly, from 16 percent in 1991 to 23 percent in 1995; in absolute terms, this additional burden borne by the patient was only $12 (from $39 to $51). For the combination treatments involving PT and the SSRIs, however, the increase was much more substantial, both in terms of OOPPs ratio and absolute dollar burden. In the SSRI+PT bundle, for example, the OOPPs ratio rose from 11 percent to 15 percent between 1991 and 1995, with the demand price increasing $40 from $102 to $142; even more dramatically, for the TCA/SSRI+PT bundle the OOPPs ratio rose from 10 percent to 17 percent, while the demand price jumped $61 from $102 in 1991 to $163 in 1995. The changes resulted from increased cost sharing for psychotherapy.

One important implication of these findings is that when considering construction of price indexes for the treatment of a medical condition such as depression, it is imperative that demand prices be distinguished from supply prices, for changes in insurance plan design benefits over time can introduce sharp differences into their time trends.

12.7 Aggregate Price Indexes for the Treatment of Depression

With the prices and quantities of the seven treatment bundles as elementary building blocks, we now move on to the construction of aggregate supply and aggregate demand price indexes. A variety of price indexes can be computed, each having implicit assumptions on the extent of ex ante substitutability among the seven bundles. Since we have discussed these

alternative index number formulas and underlying assumptions elsewhere,[40] and since they are well known in the literature on index numbers,[41] we provide only a very brief discussion here.

Fixed quantity weight indexes such as Laspeyres and the Paasche indexes reflect an assumption of zero ex ante substitutability among the seven treatment bundles; the Laspeyres employs a base-period fixed weight, the Paasche the final-period fixed weight. The Cobb-Douglas employs a fixed expenditure weight (here, equal to the mean expenditure share for each bundle over time), and assumes that the elasticity of substitution among treatment bundles equals unity for all pairs of bundles. When demand curves are downward sloping and bundle prices are all increasing, between any two adjacent time periods the Laspeyres price index is greater than the Paasche.

In general the chained, or sequentially updated, price indexes are viewed as being preferable to the fixed-weight indexes. The Törnqvist discrete approximation to the Divisia, and the Fisher ideal, make no a priori assumptions on the extent of substitutability among treatment bundles. The Törnqvist weights percentage changes in each of the treatment bundles by the arithmetic mean of the expenditure share in the two time periods, whereas the Fisher ideal is the geometric mean of the corresponding Laspeyres and Paasche indexes. A commonly observed empirical pattern is that the Törnqvist and Fisher ideal price indexes are very close to one another, and both tend to be in between the Laspeyres and Paasche price indexes.

With this as background, we now report results of various aggregate price index calculations, first in table 12.7 as viewed from the supply price (i.e., PPI) vantage, and then in table 12.8 from the demand side (i.e., CPI).

The most striking result that immediately emerges from table 12.7 is that, regardless of which index number procedure is employed, over the 1991–95 period the treatment price index for acute phase major depression has hardly changed, remaining at 1.00 or falling slightly to around 0.97. Each of the price indexes reveals an increase from 1991 to 1992, they all fall in 1993, all fall again in 1994 (to between 0.95 and 0.97), and then all increase in 1995. Given some of the data problems in 1995, this last finding should be viewed with caution. Differences among the various fixed and chained indexes are relatively minor.

By comparison, in the bottom panel of table 12.7 we report various PPIs published by the BLS. Over the 1991–95 period, the aggregate PPI for all finished goods increased about 5 percent, and that for antidepressant prescription drugs increased by about 20 percent. From 1992 to 1995, the PPI for psychiatric hospital services increased by about 10 percent, while between 1994 and 1995 the overall health services PPI increased 2.4 per-

40. See Frank, Berndt, and Busch (1999) and Frank, Busch, and Berndt (1998).
41. See, for example, Diewert (1976, 1981).

Table 12.7 **Alternative Aggregate Producer ("Supply") Price Indexes for the Treatment of Acute Phase Major Depression**

Price Index	1991	1992	1993	1994	1995
Fixed weights					
Laspeyres	1.000	1.003	0.975	0.931	0.976
Paasche	1.000	1.027	0.996	0.974	1.000
Cobb-Douglas	1.000	1.018	0.985	0.953	0.992
Chained weights					
Laspeyres	1.000	1.003	0.969	0.938	0.968
Paasche	1.000	1.011	0.976	0.952	0.978
Fisher ideal	1.000	1.007	0.97	0.945	0.973
Törnqvist	1.000	1.007	0.972	0.945	0.972
BLS PPIs[a]					
Aggregate PPI	1.000	1.012	1.025	1.031	1.051
Health services				1.000	1.024
Antidepressants	1.000	1.076	1.134	1.162	1.204
Psychiatric hospitals		1.000	1.024	1.060	109.9

[a]BLS indexes are normalized to appropriate base year.

Table 12.8 **Alternative Aggregate Consumer ("Demand") Price Indexes for the Treatment of Acute Phase Major Depression**

Price Index	1991	1992	1993	1994	1995
Fixed weights					
Laspeyres	1.000	1.028	1.041	1.063	1.266
Paasche	1.000	1.037	1.080	1.118	1.334
Cobb-Douglas	1.000	1.037	1.069	1.110	1.300
Chained weights					
Laspeyres	1.000	1.028	1.048	1.081	1.281
Paasche	1.000	1.032	1.065	1.104	1.317
Fisher ideal	1.000	1.030	1.056	1.092	1.299
Törnqvist	1.000	1.030	1.057	1.093	1.298
BLS CPIs[a]					
All Items	1.000	1.030	1.061	1.088	1.119
Medical Care	1.000	1.075	1.146	1.205	1.266
Prescription drugs	1.000	1.075	1.117	1.155	1.177

[a]BLS indexes are normalized to 1991 base year.

cent. Thus, while our various supply price indexes for the treatment of acute phase depression are either flat or very slightly falling, they all grow considerably less than the various official PPIs. The "real," all-items CPI inflation-adjusted price index for the treatment of acute phase depression has fallen over the 1991–95 time span.

Index number afficionados might notice, however, that the typically observed inequality relationships among the Laspeyres and Paasche price indexes are not present here. For example, for both fixed- and chained-

weight versions, the Paasche price index is larger than the Laspeyres. The reason this occurs is that, as noted in section 12.6 above, in a number of cases changes in treatment bundle supply prices and quantity shares are positively rather than negatively correlated. Because the price index calculations do not take into account factors affecting demand other than price, and because consumers do not face supply prices, there is the appearance of a positive relation between quantity and supply price. We discuss other price-related issues in section 12.8 below, where we deal with hedonic price indexes. Finally, note that as is usually the case, the Fisher ideal and Törnqvist indexes are always in between the chained Paasche and Laspeyres indexes.

We now turn to a discussion of aggregate price indexes viewed from the demand side, or the patients' consumers' OOPPs. As seen in table 12.8, the demand aggregate price indexes have very different time trends than do the supply aggregate price indexes. Each of the price indexes exhibits an increase between each pair of adjacent years, with the 1994–95 price increase being particularly dramatic. By 1995, the fixed-weight price indexes (normalized to 1.000 in 1991) ranged between 1.27 and 1.33, while the chained indexes had a slightly smaller variation, between 1.28 and 1.32. Between 1991 and 1994, the price indexes increased by only about 10 percent, but in 1995 the increase was much larger, averaging around 20 percent. As discussed in section 12.6, we attribute this sharp increase to benefit design changes in 1995 that increased patient copayments and deductibles.

The discrepancy between the PPI and CPI results highlights several important points in interpretation of the data. First is that if the elasticity with respect to the demand price is less than unity (in absolute value), spending must increase. Because cost sharing increased notably and demand estimates suggest an elasticity of less than unity, it is not surprising that the CPI increases over time. Second, and more important, is the welfare interpretation of these changes. It is well known that low levels of demand side cost sharing in medical insurance are associated with welfare losses due to moral hazard (Newhouse et al. 1993). Thus, changes in the CPI for depression may reflect reduced moral hazard in addition to higher prices per unit of "effective care." Research by Manning and Marquis (1989) suggests that 50 percent cost sharing for mental health care was optimal in a fee-for-service context. Because observed rates in our data are quite a bit lower, gains from reduced moral hazard may be significant. Since interpretation of the CPI is unclear, we focus more attention on the PPI.

12.8 Hedonic Price Indexes

An enduring issue in price index measurement is how one should make adjustments for quality change. Quality adjustments could be significant

here, for it is reasonable to expect that the characteristics and attributes of the acute phase of treatment for depression have changed over time, and that changes in the underlying patient population might also have occurred.

Although its use to date in the medical marketplace has been limited, the hedonic price approach has been used in many other contexts to adjust price movements for (not fully priced) quality changes.[42] We now briefly outline our implementation of hedonic price measurement for the treatment of acute phase depression.

We expect that treatment costs for depression are affected by patient characteristics, various attributes of antidepressant medications, the type of treatment given the patient, and the year in which the treatment episode began. Patient characteristics for which we have data include the patient's gender, age, general medical condition (measured by the number of comorbid general medical conditions, as noted on the bottom of table 12.1), the industry in which the plan enrollee is employed (transportation, communication, and utilities; services; or government), and information on the patient's mental health comorbidities. With respect to these mental health comorbidities, we construct dummy variables for whether the patient is also diagnosed with a substance abuse disorder, a panic/anxiety disorder, schizophrenia or bipolar depression, depression not otherwise specified, and whether the current episode involves a recurrence of major depression. For attributes of the various antidepressant medications, we employ data on the half-life of the medication and on the number of side effects that frequently accompany use of the medication.[43] Regarding treatments, we create dummy variables for each of the seven guideline-compatible treatment bundles described earlier. We also create a dummy variable indicating whether at the start of the episode the individual was enrolled in a behavioral health carve-out. Finally, we create dummy variables indicating the year in which the treatment episode began.

The dependent variables in the hedonic regressions are the natural logarithm of the supply price and of the demand price. The explanatory variables are the patient, medication, treatment, and time variables discussed above. The set of omitted dummy variables, and thus the reference case, is that for a female; where the enrollee is employed in the transportation, communication, or utilities industries; has none of the mental health comorbidities noted above; is not being treated for a recurrent episode of acute phase depression; treatment consists of PT alone; and the treatment episode began in 1991. Parameter estimates for the supply price equation

42. For an introductory discussion to the hedonic method, see Griliches (1988), chapters 7 and 8, and Berndt (1991), chapter 4. Applications in the medical context include Trajtenberg (1990) and Berndt, Cockburn, and Griliches (1996).

43. The side effect data are discussed in detail in Berndt, Cockburn, and Griliches (1996). We sum up the 0 (rare) to 4 (common) ratings across the various side effects. Here we also assume that psychotherapy has no side effects.

are in the left-hand panel of table 12.9, while estimates for demand price equation are in the right-hand panel. A number of results are worth mentioning in detail. We begin with the supply price.

First, characteristics of the patient and his or her illness have a statistically significant impact on treatment costs. While age does not appear to affect the supply price, mental health comorbidities such as anxiety/panic, schizophrenia/bipolar depression, and depression not otherwise specified all have a positive and significant impact. Somewhat surprisingly, comorbid substance abuse has a negative coefficient estimate, although not significant. The sum of the patient's other general medical comorbidities is also negative, but not significant. Finally, treatment costs for patients in which the enrollee is employed in the services or government sectors are considerably lower than if the industry involves transportation, communications, or utilities.

Second, attributes of the antidepressant medications also affect supply prices of treatment, even when one controls for type of treatment. Specifically, within the various bundles involving prescription drugs, greater side effects are associated with lower supply prices (lower side effects with higher supply prices), and increased half-life of the drug (mitigating the negative impacts of occasionally forgetting to take the medication) also raises the supply price. Relative to PT alone, treatment with TCA alone or SSRI alone reduces the supply price significantly, while use of SSRI + PT or TCA/SSRI + PT increases supply price.

Somewhat surprisingly, we find that whether at the beginning of the treatment episode the patient was enrolled in a mental health carve-out has no significant impact on supply price. To check whether the carve-out impact was diluted because of differential treatment bundles used by the carve-out, we also estimated a supply price equation in which the treatment bundle dummy variables were deleted; results are given in the second column of parameter estimates in table 12.9. Although the goodness of fit was reduced considerably when the treatment bundle variables were deleted, the carve-out parameter estimate remained insignificant. This may be due in part to the fact that the comparison conditions involved managed indemnity arrangements which tend to negotiate price discounts and shift care patterns.

Of particular interest are the parameter estimates on the 1992–95 yearly dummy variables. In the context of the hedonic price equation, these parameter estimates indicate how the supply price changed relative to 1991, holding constant the various patient characteristics, medication attributes, and the treatment bundle composition. Each of the four parameter estimates is small and insignificantly different from zero; hedonic price indexes computed as antilogarithms of these coefficients are 1.049, 1.004, 0.980, and 1.027 for 1992 through 1995, respectively (1991 = 1.000). These quality-adjusted hedonic price indexes are consistent with traditional aggregate supply price indexes reported in table 12.7 above. Adjusting for

Table 12.9 Parameter Estimates from Hedonic Price Regressions

Variable	Supply Price ("PPI") Equations		Demand Price ("CPI") Equations	
Constant	6.528	6.587	4.614	4.658
	(112.60)	(98.91)	(51.88)	(47.39)
Male	−0.016	−0.005	−0.090	−0.056
	(0.75)	(0.18)	(2.26)	(1.29)
Age	0.003	0.000	0.011	0.008
	(3.38)	(0.09)	(6.71)	(4.19)
Anxiety/panic	0.258	0.350	0.126	0.266
	(5.68)	(7.06)	(0.89)	(1.84)
Schizophrenia/bipolar	0.235	0.393	0.233	0.458
	(9.94)	(14.62)	(5.86)	(10.56)
Other depression	0.128	0.183	0.070	0.125
	(5.32)	(6.16)	(1.75)	(2.61)
Substance abuse	−0.136	−0.093	−0.529	−0.478
	(1.38)	(0.87)	(2.71)	(2.36)
Recurrent	0.080	0.124	0.050	0.132
	(4.04)	(5.32)	(1.44)	(3.42)
Sum medical comorbidities	−0.025	−0.024	−0.029	−0.031
	(2.04)	(1.58)	(1.29)	(1.24)
Government	−0.121	−0.172	−0.858	−0.936
	(3.07)	(3.74)	(17.21)	(16.10)
Services	−0.184	−0.225	−0.991	−1.053
	(4.62)	(4.92)	(21.59)	(19.53)
Sum side effects	−0.084	−0.129	0.025	−0.104
	(3.33)	(10.45)	(0.78)	(6.26)
Halflife	0.001	0.001	0.000	−0.001
	(4.01)	(4.47)	(1.31)	(3.53)
TCA alone	−1.170	n.a.	−1.483	n.a.
	(10.15)		(8.77)	
SSRI alone	−0.892	n.a.	−1.847	n.a.
	(13.50)		(19.18)	
TCA+PT	0.137	n.a.	−0.116	n.a.
	(1.38)		(0.89)	
SSRI+PT	0.195	n.a.	−0.122	n.a.
	(3.21)		(1.41)	
TCA/SSRI	−0.654	n.a.	−1.320	n.a.
	(5.71)		(8.32)	
TCA/SSRI+PT	0.301	n.a.	0.007	n.a.
	(4.15)		(0.07)	
Carve-out	−0.047	0.025	−0.028	0.092
	(0.90)	(0.43)	(0.44)	(1.24)
Start 1992	0.048	0.072	0.108	0.099
	(1.23)	(1.60)	(1.62)	(1.35)
Start 1993	0.004	0.054	0.114	0.094
	(0.11)	(1.21)	(1.71)	(1.30)
Start 1994	−0.021	0.025	0.145	0.097
	(0.54)	(0.58)	(2.09)	(1.31)
Start 1995	0.026	0.059	0.366	0.296
	(0.59)	(1.14)	(4.92)	(3.53)
R^2	0.328	0.088	0.277	0.083
RMSE	0.689	0.802	1.168	1.315
N	5,034	5,034	5,034	5,034

Notes: Absolute value of t-statistic in parentheses, from robust standard error. n.a. = not applicable.

various patient, medication, and treatment changes over time does not appear to affect the central finding that from 1991 to 1995 there was very little change in the supply price for treating acute phase major depression.

We now turn to the demand price hedonic equations, whose parameter estimates appear in the last two columns of table 12.9. In general, the pattern of results from the hedonic demand price equations involving quality attributes is not as strong and dramatic as is that of the hedonic supply price equation. For example, while parameter estimates on the patient's mental health comorbidities are of the same sign, in the demand price equation several of them are no longer statistically significant; attributes of the antidepressant medication are not significant as well. As with the supply equation, the mental health carve-out coefficient is insignificant.

However, in contrast to the supply price equation, for the demand hedonic price equation the parameter estimates on the 1992–95 year dummy variables are positive, and by 1994 and 1995 they become statistically significant. Relative to 1991 = 1.000, holding constant patient characteristics, medication attributes, and treatment composition, the hedonic price indexes for 1992–95 (computed as antilogs of the estimated parameters) are 1.114, 1.121, 1.155, and 1.443. Compared to the traditional aggregate demand price indexes reported in table 12.8, the time trends in these hedonic price indexes are broadly similar. While the traditional aggregate demand price index in table 12.8 averages about 1.30 in 1995, the corresponding hedonic quality-adjusted demand price index is 1.44; if one takes the antilog of the 1995 parameter estimate in table 12.9 minus one times the robust standard error, one obtains a 1995 price index of about 1.34, not much different from that in table 12.8. We conclude, therefore, that traditional and hedonic approaches to price index measurement yield broadly similar results.

12.9 Discussion

Our goal in this research project has been to extend previous research aimed at constructing CPI- and PPI-like medical price indexes that deal with prices of treatment episodes rather than prices of discrete inputs, that are based on transaction rather than list prices, that take quality changes and expected outcomes into account, and that employ more current expenditure weights in the aggregation computations. Although we have made considerable progress in achieving this goal, we believe a number of caveats are in order.

First, the results reported here differ considerably from those reported earlier in Frank, Busch, and Berndt (1998) and Frank, Berndt, and Busch (1999); there we reported substantial supply and demand price declines over the 1991–95 period. Although we are still examining why the differences occurred, we believe that the primary difference is that in the new

data set we bring in many more patients. The evidence suggests that over time the new data set has an increasing share of patients with (1) more complicated conditions, (2) greater severity of illness, and (3) elements of longer term treatments such as continuation phase care. While the older data set may have corresponded with episodes more closely approximating the pristine world of clinical trials, the larger, new data set involves patients with more typical complicated illnesses.

It is therefore of interest to examine how robust are our new data set findings when one relaxes the strict guideline-compatible standards we imposed in defining episodes of care. We have created yet another data set, where we relaxed the definitions of the seven bundles. Bundles which consider drugs alone as treatment remain unchanged at 30 to 180 days of treatment. Definitions of bundles which include psychotherapy have been expanded such that the cost of up to twenty psychotherapy visits is now included in the cost of the episode. Furthermore, for the treatment "psychotherapy alone" we previously allowed for a minimum of five visits. This constraint was relaxed to include episodes treated with three or four visits. Appendix table 12A.1 shows that with the broader definition of standards of care the price indexes are largely unchanged. Specifically, they are rather flat, not falling below 2 percentage points in comparing 1991 and 1995.

It is also useful to comment on implications of our research findings. Using alternative index number procedures, we find that over the 1991–95 time period, the nominal supply price of acute phase treatment for depression is essentially unchanged, while the real supply price (relative to GDP deflator) has declined. Note that this may well create an upward bias in estimated price movements for acute phase treatment of depression. Recall that we showed that the percentage of episodes meeting guideline standards has increased over time. The price indexes reported above do not account for this apparent move toward the production frontier. Improved compliance with standards of care implies greater benefits are being generated over time, thereby understating price reductions of effective care. The demand price index revealed a steady 2 to 3 percent per year increase from 1991 to 1994, which is then followed by a sharp 20 percent increase in 1995. Since the demand price encompasses only consumers'/patients' direct OOPPs, and since the supply price is the sum of consumer plus insurer payments, it follows that from 1991 to 1994 there was steady 2 to 3 percent annual *decrease* in insurers' expenditures, and that in 1995 a very substantial 20 percent *decrease* took place. While the nominal supply price of acute phase treatment for depression was apparently unchanged between 1991 and 1995, the compositional burden shifted from third-party insurer to patient. The change in incidence of payment, in large part due to increased patient copayments and deductibles, likely reflects efforts by insurers to deal with moral hazard via design of mental health benefits and tightening of the size of provider networks used in managed care.

Our research findings based on index number procedures are broadly consistent with those based on simple hedonic price methods, particularly with the supply price. On the demand side, the hedonic price index increases more rapidly up to 1994, and then increases more sharply in 1995 than do the traditional index numbers. Note that what we call "traditional" here, however, represents a substantial departure from current BLS practice, in that we have used as elementary building blocks the transaction prices accompanying alternative treatment bundles, and not the (list) prices of fixed set of inputs. Nevertheless, it is interesting that, as in Berndt, Cockburn, and Griliches (1996), the marginal impact of incorporating hedonic pricing methods is not that significant, once one defines output quantities carefully. Our hedonic research could have gone several steps further, allowing for changing parameters over time and integrating the estimated hedonic price equations with index number procedures, but we leave those nuances for further research.[44]

Finally, it is worth commenting on implications of this research for procedural revisions currently under consideration at the BLS. The BLS is responsible for constructing and publishing PPIs and CPIs for a very large number of medical-related services and commodities. Our research has focused on but one disorder—major depression—albeit one whose output measurement presented considerable challenges. We believe that economies of scope and scale are undoubtedly available, and we have experienced a steep learning curve in working with the MEDSTAT data. We also believe that extension of the treatment episode approach to other illnesses and disorders is feasible. Nevertheless, we are struck by how difficult, time-consuming, and expensive this research has been for us. If a treatment episode approach is to be implemented efficiently at the BLS (or elsewhere, such as at HCFA), considerable care needs to be exercised in choosing a set of illnesses/disorders for which treatment bundles can be well defined, for which quality comparisons can readily be made (facilitated by AHCPR or other published professional guideline standards), and for which claims data can readily be employed.

12.10 Concluding Remarks

We have implemented an approach that employs transaction data from a publicly available retrospective medical claims database of almost half a million lives, annually from 1991 to 1995. Based on a review of the clinical research literature dealing with the acute phase of treatment for depression, we identify seven alternative service bundles that combine varying types and quantities of prescription drugs, medical management, and psychotherapy. We construct data on episodes of treatment and their cost.

44. See Berndt and Griliches (1993) and Berndt, Griliches, and Rappaport (1995).

Because the treatment bundles are viewed by the medical community as being therapeutically similar in terms of ex ante efficacy, our linking of treatment bundles provides an important step toward incorporation of expected medical outcomes. We distinguish a supply price index, similar to the BLS PPI, that represents the total receipts received by providers of medical treatment (from the insurer and the patient), from a demand price index (similar to the CPI) that incorporates only the out-of-pocket payments (via copayments and deductibles) by the patient/consumer. Finally, we employ a variety of traditional aggregate index number formulas, consistent with varying assumptions concerning the ex ante substitutability among the seven treatment bundles, as well as hedonic price procedures in computing price indexes.

For the supply side price index, the various price indexes all point to a flat and essentially unchanging price over the 1991–95 period. Because the BLS's various producer price indexes all indicate a significant increase over the same time period, our results suggest that in real terms, the supply price for treatment of acute phase depression has fallen over time. On the demand side, however, the indexes show a total increase in price of about 10 percent between 1991 and 1994, and then a very sharp 20 percent increase. The usual welfare interpretation of the CPI is not possible in this case. The price increase largely reflects increased cost sharing arrangements which are certain to reduce welfare losses stemming from moral hazard. This is especially the case in MEDSTAT data, which generally cover plans with substantially lower cost sharing for mental health care than is typical in the economy as a whole (20 percent versus 50 percent).

Our research can be extended in a number of ways. Two issues are particularly important. First, although we have experimented with use of both strict and somewhat relaxed guideline criteria to define episodes of care, in each case we have assumed an "all or nothing" treatment. The implicit production function therefore involves a steep step or cliff in which treatment either is or is not effective. An alternative is to consider some subguideline treatment bundles as having a lower probability of being efficacious than are those meeting guideline criteria, and therefore to assign the subguideline episodes a lower quantity weight. Together with a group of clinicians and psychiatric researchers, we are currently initiating an effort to more completely define the specific depression treatment production function.

Second, we have adhered rather closely to the notion of defining output in terms of episodes of acute phase treatment. This involves a somewhat arbitrary delineation of an output that often does not correspond well to the reality of clinical practice. Because depression is often a chronic recurring illness, continuation and maintenance phase care are part of the overall treatment of the disorder. Unfortunately, because the bulk of clinical research has focused on the acute phase of treatment, our links to out-

comes were tied to a somewhat imperfect characterization of care. It is therefore important to more completely specify treatment in this context.

Finally, and perhaps most surprisingly, all of our results point to the fact that given a budget for treatment of depression, much more could be accomplished in 1995 than in 1991. That is, real prices of care have fallen. This runs counter to most public and expert perceptions. The implications for interpreting spending changes are enormous. For years the view has been that spending increases on mental health care (of which depression is 50 percent in private insurance) was due to provision of increasing low benefit services and higher payments to providers. Our results point to a different story where spending increases are due to a larger number of "effective" treatments being provided. These treatments represent a shift toward new treatment technologies that are provided under a new set of organizational arrangements.

Appendix

Table 12A.1 **Aggregate Producer ("Supply") Price Indexes Calculated Using Broader Standard of Care Definitions**

Price Index	1991	1992	1993	1994	1995
Fixed weights					
Laspeyres	1.000	1.015	0.984	0.929	0.997
Paasche	1.000	1.029	1.000	0.968	1.010
Chained weights					
Laspeyres	1.000	1.015	0.981	0.941	0.985
Paasche	1.000	1.017	0.986	0.952	0.993
Fisher ideal	1.000	1.016	0.983	0.946	0.989
Törnqvist	1.000	1.016	0.983	0.946	0.988

References

American Medical Association. 1990. *Current procedural terminology (CPT),* 4th ed. Chicago: American Medical Association.

American Psychiatric Association (APA). 1993. Practice guidelines for major depressive disorder in adults. *American Journal of Psychiatry* 150, no. 4 (suppl.): 1–26.

———. 1994. *Diagnostic and statistical manual of mental disorders,* 4th ed. Washington, D.C.: American Psychiatric Association.

Armknecht, Paul A. 1996. Improving the efficiency of the U.S. CPI in the future. Washington, D.C.: International Monetary Fund. Unpublished manuscript.

Armknecht, Paul A., and Daniel H. Ginsburg. 1992. Improvements in measuring

price changes in consumer services: Past, present and future. In *Output measurement in the service sectors,* ed. Z. Griliches, assisted by Ernst R. Berndt, Timothy Bresnahan, and Marilyn E. Manser, 109–56. NBER Studies in Income and Wealth, vol. 56. Chicago: University of Chicago Press.

Armknecht, Paul A., Brent R. Moulton, and Kenneth J. Stewart. 1994. Improvements to the food at home, shelter and prescription drug indexes in the U.S. Consumer Price Index. CPI Announcement—Version I, 20 October. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

Beach, S. R. H., E. E. Sandeen, and K. D. O'Leary. 1990. *Depression in marriage.* New York: Guilford.

Beck, A. T., A. John Rush, B. F. Shaw, and G. Emergy. 1979. *Cognitive therapy of depression.* New York: Guilford.

Berndt, Ernst R. 1991. *The practice of econometrics: Classic and contemporary.* Reading, Mass.: Addison Wesley.

Berndt, Ernst R., Iain Cockburn, and Zvi Griliches. 1996. Pharmaceutical innovations and market dynamics: Tracking effects on price indexes for antidepressant drugs. *Brookings Papers on Economic Activity, Microeconomics,* 133–88.

Berndt, Ernst R., Richard G. Frank, and Thomas G. McGuire. 1997. Alternative insurance arrangements and the treatment of depression: What are the facts? *American Journal of Managed Care* 3 (2): 243–50.

Berndt, Ernst R., and Zvi Griliches. 1993. Price indexes for microcomputers: An exploratory study. In *Price measurements and their uses,* ed. Murray F. Foss, Marilyn E. Manser, and Allan H. Young. NBER Studies in Income and Wealth, vol. 57. Chicago: University of Chicago Press.

Berndt, Ernst R., Zvi Griliches, and Neal Rappaport. 1995. Econometric estimates of price indexes for personal computers in the 1990's. *Journal of Econometrics* 68 (1): 243–68.

Broadbent, W. Eugene., Dan G. Blazer, Linda K. George, and Chiu Kit Tse. 1990. Depression, disability days and days lost from work in a prospective epidemiologic survey. *Journal of the American Medical Association* 264: 2524–28.

Busch, Susan H., Richard G. Frank, and Ernst R. Berndt. 1996. Effectiveness, efficacy and price indexes for depression: A review of the literature. Boston: Harvard Medical School, Department of Health Care Policy. Unpublished manuscript.

Cardenas, Elaine M. 1996a. The CPI for hospital services: Concepts and procedures. *Monthly Labor Review* 119 (7): 34–42.

———. 1996b. Revision of the CPI hospital services component. *Monthly Labor Review* 119 (12): 40–48.

Catron, Brian, and Bonnie Murphy. 1996. Hospital price inflation: What does the new PPI tell us? *Monthly Labor Review* 119 (7): 124–31.

Croghan, Thomas W., Catherine A. Melfi, Deobrah G. Dobrez, and Thomas J. Kniesner. 1999. Effect of mental health specialty care on antidepressant length of therapy. *Medical Care* 37 (suppl.): AS20–23.

Crown, William E., Robert L. Obenchain, Luella Englehart, Tamra Lair, Don P. Buesching, and Thomas W. Croghan. 1996. Application of sample selection models to outcomes research: The case of evaluating effects of antidepressant therapy on resource utilization. Cambridge, Mass.: The MEDSTAT Group. Unpublished manuscript.

Cutler, David, Mark McClellan, Joseph Newhouse, and Dahlia Remler. 1998. Are medical prices declining? Evidence from heart attack treatments. *Quarterly Journal of Economics* 113 (November): 991–1024.

Diewert, W. Erwin. 1976. Exact and superlative index numbers. *Journal of Econometrics* 4 (2): 115–45.

———. 1981. The economic theory of index numbers: A survey. In *Essays in the theory and measurement of consumer behavior in honour of Sir Richard Stone,* ed. Angus Deaton. Cambridge: Cambridge University Press.

Elkin, Irene, T. Shea, J. T. Watkins, S. D. Imber, S. M. Sotsky, J. F. Collins, D. R. Glass, P. A. Pilkonis, W. R. Leber, J. P. Docherty, S. J. Fiester, and M. B. Parloff. 1989. National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. *Archives of General Psychiatry* 46: 971–82.

Ellis, Randall P., Gregory C. Pope, Lisa I. Iezzoni, John Z. Ayanian, David W. Bates, Helen Burstin, and Arlene S. Ash. 1996. Diagnosis-based risk adjustment for Medicare capitation payments. *Health Care Financing Review* 17 (3): 101–28.

Feldstein, Paul J. 1979. *Health care economics,* 1st ed. New York: Wiley.

———. 1983. *Health care economics,* 2nd ed. New York: Wiley.

———. 1988. *Health care economics,* 3rd ed. New York: Wiley.

Fixler, Dennis, and Mitch Ginsburg. 1997. Health care output and prices in the Producer Price Index. Slides from presentation to the NBER Summer Institute, Franco-American Seminar, Cambridge, Mass., 23 July.

Ford, Ina Kay, and Daniel H. Ginsburg. 1997. Medical care in the CPI. Paper presented to the NBER Summer Institute, Franco-American Seminar, Cambridge, Mass., 23 July.

Frank, Ellen, David J. Kupfer, J. M. Perel, Cleon Cornes, D. B. Jarrett, A. G. Mallinger, Michael E. Thase, A. B. McEachran, and V. J. Grochocinski. 1990. Three-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry* 47: 1093–99.

Frank, Richard G., Ernst R. Berndt, and Susan H. Busch. 1999. Price indexes for the treatment of depression. In *Measuring the prices of medical treatments,* ed. Jack E. Triplett. Washington, D.C.: The Brookings Institution.

Frank, Richard G., Susan H. Busch, and Ernst R. Berndt. 1998. Measuring prices and quantities of treatment for depression. *American Economic Review* 88 (2): 106–11.

Frank, Richard G., Thomas G. McGuire, and Joseph P. Newhouse. 1995. Risk contracts in managed mental health care. *Health Affairs* 14 (3): 50–64.

Getzen, Thomas E. 1992. Medical care price indexes: Theory, construction and empirical analysis of the US series 1927–1990. In *Advances in health economics and health services research,* ed. R. Scheffler and L. F. Rossiter. Greenwich, Conn.: JAI Press.

Gilbert, Milton. 1961. The problem of quality change and index numbers. *Monthly Labor Review* 84 (9): 992–97.

———. 1962. Quality change and index numbers: A reply. *Monthly Labor Review* 85 (5): 544–45.

Goldman, William, Joyce McCulloch, and Roland Sturm. 1998. Costs and use of mental health services before and after managed care. *Health Affairs* 17 (2): 40–52.

Griliches, Zvi. 1962. Quality change and index numbers: A critique. *Monthly Labor Review* 85 (5): 532–44.

———. 1988. *Technology, education, and productivity.* New York: Basil Blackwell.

Griliches, Zvi, and Iain Cockburn. 1994. Generics and new goods in pharmaceutical price indexes. *American Economic Review* 84 (5): 1213–32.

Hays, R. D., K. B. Wells, C. D. Sherbourne, W. Rogers, and K. Spritzer. 1995. Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. *Archives of General Psychiatry* 52 (1): 11–19.

Kanoza, Douglas. 1996. Supplemental sampling in the PPI Pharmaceuticals Index. *Producer Price Indexes Detailed Price Report* (January): 8–10.

Katon, Wayne, Michael von Korff, Elizabeth Lin, Terry Bush, and Johann Ormel. 1992. Adequacy and duration of antidepressant treatment in primary care. *Medical Care* 30:67–76.

Keeler, Emmett B., Kenneth B. Wells, Willard G. Manning, J. David Rumpel, and Janet M. Hanley. 1986. The demand for episodes of mental health services. Santa Monica, Calif.: RAND Report R-3432-NIMH, October.

Keller, Martin B., Gerald L. Klerman, Philip W. Lavori, William Coryell, Jean Endicott, and John Taylor. 1984. Long-term outcome of episodes of major depression. *Journal of the American Medical Association* 252 (6): 788–92.

Keller, Martin B., Philip W. Lavori, Gerald L. Klerman, Nancy C. Andreasen, Jean Endicott, William Coryell, Jan Fawcett, John P. Rice, and M. A. Robert. 1986. Low levels and lack of predictors of somathotherapy and psychotherapy received by depressed patients. *Archives of General Psychiatry* 43 (5): 458–66.

Keller, Martin B., Philip W. Lavori, John Rice, William Coryell, and Robert M. A. Hirschfeld. 1986. The persistent risk of chronicity in recurrent episodes of nonbipolar major depressive disorder: A prospective follow-up. *American Journal of Psychiatry* 143 (1): 24–28.

Kelly, Gregory G. 1997. Improving the PPI samples for prescription pharmaceuticals. *Monthly Labor Review* 120 (10): 10–17.

Kendler, K. S., E. E. Walters, M. C. Neale, Ronald C. Kessler, A. C. Heath, and L. J. Eaves. 1995. The structure of genetic and environmental risk factors for six major psychiatric disorders in women: Phobia, generalized anxiety disorder, panic disorder, bulimia, major depression, and alcoholism. *Archives of General Psychiatry* 52 (5): 374–83.

Kessler, Ronald C., Katherine A. McGonagle, Marvin Swartz, Dan G. Blazer, and Christopher B. Nelson. 1993. Sex and depression in the National Comorbidity Survey I: Lifetime prevalence, chronicity, and recurrence. *Journal of Affective Disorders* 29:85–96.

Kessler, Ronald C., Katherine A. McGonagle, Shanyang Zhao, Christopher B. Nelson, Michael Hughes, Suzann Eshlerman, Hans-Ulrich Wittchen, and Kenneth S. Kendler. 1994. Lifetime and twelve-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry* 51 (1): 8–19.

Kupfer, David J. 1991. Long-term treatment of depression. *Journal of Clinical Psychiatry* 52, no. 5 (suppl.): 28–34.

Kupfer, David J., E. Frank, J. M. Perel, C. Cornes, A. G. Mallinger, M. E. Thase, A. B. McEachran, and V. J. Grochocinski. 1992. Five-year outcomes for maintenance therapies in recurrent depression. *Archives of General Psychiatry* 49:769–73.

Langford, Elizabeth A. 1957. Medical care in the CPI: 1935–1956. *Monthly Labor Review* 80 (9): 1053–58.

Levit, Katharine R., Helen C. Lazenby, Brandley R. Braden, and the National Health Accounts Team. 1998. National health spending trends in 1996. *Health Affairs* 17 (1): 35–51.

Ma, Ching-to Albert, and Thomas G. McGuire. 1998. Costs and incentives in a behavioral health carve-out. *Health Affairs* 17 (2): 53–69.

Manning, Willard G., and M. Susan Marquis. 1989. *Health insurance: The tradeoff between risk pooling and moral hazard.* Publication no. R-3729-NCHSR. Santa Monica, Calif.: RAND Corporation.

Marion Merrell Dow. Annual. *Managed care digest: HMO edition.* Kansas City, Mo. Marion Merrell Dow.

McGuire, Thomas G. 1989. Financing and reimbursement for mental health services. In *The future of mental health services research,* ed. Carl A. Taube, David Mechanic, and Ann A. Hohmann. Rockville, Md.: U.S. Department of Health and Human Services.

Mechanic, David. 1989. The evolution of mental health services and mental health services research. . In *The future of mental health services research,* ed. Carl A. Taube, David Mechanic, and Ann A. Hohmann. Rockville, Md.: U.S. Department of Health and Human Services.

Newhouse, Joseph P. 1989. Measuring medical prices and understanding their effects. *Journal of Health Administration Education* 7 (1): 19–26.

Newhouse, Joseph P., and the Insurance Experiment Group. 1993. *Free for all? Lessons from the RAND health insurance experiment.* Cambridge, Mass.: Harvard University Press.

Papolos, Demitri F., and Janic Papolos. 1992. *Overcoming depression.* Rev. ed. New York: HarperCollins.

Prospective Payment Assessment Commission. 1995. *Report and recommendations to the Congress.* Washington, D.C.: U.S. Government Printing Office.

Schulberg, Herbert C., Marjorie Saul, Maureen McClelland, Mary Ganguli, Wallace Cristy, and Ellen Frank. 1985. Assessing depression in primary medical and psychiatric practices. *Archives of General Psychiatry* 42 (12): 1164–70.

Scitovsky, Anne A. 1964. An index of the cost of medical care—A proposed new approach. In *The economics of health and medical care,* proceedings of the Conference on the Economics of Health and Medical Care, 10–12 May 1962. Ann Arbor: University of Michigan.

———. 1967. Changes in the costs of treatment of selected illnesses, 1951–65. *American Economic Review* 57 (5): 1182–95.

Trajtenberg, Manuel. 1990. *Economic analysis of product innovation: The case of CT scanners.* Cambridge, Mass.: Harvard University Press.

Triplett, Jack E. 1990. The theory of industrial and occupational classifications and related phenomena. In *Proceedings of the Bureau of the Census 1990 Annual Research Conference.* Washington, D.C.: U.S. Department of Commerce.

U.S. Department of Health and Human Services. 1980. *International classification of diseases (ICD-9-CM),* 2d ed. Washington, D.C.: U.S. Government Printing Office.

———. Public Health Service. Agency for Health Care Policy and Research. 1993. *Depression in primary care,* vol. 2. Rockville, Md.: U.S. Department of Health and Human Services.

U.S. Department of Labor. Bureau of Labor Statistics. 1992. *Handbook of methods.* bulletin 2414. Washington, D.C.: U.S. Government Printing Office.

———. 1995. Improvements to CPI procedures: Prescription drugs. Washington, D.C.: U.S. Government Printing Office.

———. N.d. A description of the PPI hospital services initiative. Washington, D.C.: U.S. Government Printing Office.

U.S. Senate Finance Committee. 1996. *Final Report from the Advisory Commission to Study the Consumer Price Index.* Washington, D.C.: U.S. Government Printing Office.

Weiner, Jonathan P., Allen Dobson, Stephanie L. Maxwell, Kevin Coleman, Barbara Starfield, and Gerald Anderson. 1996. Risk-adjusted medicare capitation rates using ambulatory and inpatient diagnoses. *Health Care Financing Review* 17 (3): 77–100.

Wells, Kenneth B., Wayne Katon, William Rogers, and Patti Camp. 1994. Use of minor tranquilizers and antidepressant medications by depressed outpatients: Results from the medical outcomes study. *American Journal of Psychiatry* 151 (5): 694–700.

Wells, Kenneth B., Anita Stewart, R. D. Hays, M. A. Burnam, W. Rogers, M. Daniels, S. Berry, S. Greenfield, and J. Ware. 1989. The functioning and well-being

of depressed patients: Results from the Medical Outcomes Study. *Journal of the American Medical Association* 262 (7): 914–19.
Wells, Kenneth B., Roland Sturm, Cathy D. Sherbourne, and Lisa S. Meredith. 1996. *Caring for depression: A RAND study.* Cambridge, Mass.: Harvard University Press.
Wilde, Michelle I., and Paul Benfield. 1998. Fluoxetine: A pharmacoeconomic review of its use in depression. *Pharmacoeconomics* 13 (5): 543–61.
Wingert, Terence D., John E. Kralewski, Tammie J. Lindquist, and David J. Knutson. 1995. Constructing episodes of care from encounter and claims data: Some methodological issues. *Inquiry* 32: 430–43.

Comment

Darrel A. Regier

This paper by Berndt, Busch, and Frank offers a new and important paradigm for evaluating the costs of providing and obtaining mental health treatment services in a rapidly changing "defacto mental health service system" (Regier et al. 1993). It is an ambitious effort to use the Bureau of Labor Statistics (BLS) Producer Price Index (PPI) and Consumer Price Index (CPI) conventions to describe market changes in the cost of treating major depression over a five-year time frame (1991–1995). The authors attempt to go beyond the usual CPI approach of assessing the average price of a "fixed market basket of goods and services," such as the average cost of a mental health visit, bed-day, or medication/day, to examining the cost of an episode of care that meets American Psychiatric Association treatment-guideline standards for major depressive disorder in adults. (APA 1993). Although such an approach to assessing "the average costs of the complete treatment of individual illnesses" was recommended as early as 1962 by Scitovsky, very few have advanced beyond pricing health care inputs to obtaining the price of health care outcomes.

Within the mental health field, there have been very few published studies that attempt to assess the cost effectiveness of mental disorder treatments by using comparative treatment outcomes in the analysis. A notable exception is the study by Rogers et al. (1993), which evaluated the relative cost of treatment outcomes of depression in different service systems. In this landmark study, psychiatrists in an HMO practice offered ostensibly less expensive treatment for depression with fewer visits and a lower use of antidepressant medications than were provided for comparable patients in fee-for-service settings. However, when outcomes were factored into the cost estimates, the more intensive treatment was shown to be a more cost-effective approach.

Darrel A. Regier is executive director of the American Psychiatric Institute for Research and Education and director of the Office of Research of the American Psychiatric Association.

With this background in mind, the authors are careful to review the epidemiological literature to assess the potential true prevalence of major depressive disorder in contrast to the treated prevalence reported for the MEDSTAT population used in this paper. There is a significant gap between the reported 10.3 percent annual prevalence of major depression cited from the National Comorbidity Survey (Kessler et al. 1994) and the 18,920 episodes of acute care for major depression from 428,168 enrollees during the five-year study period. If every episode represented a different patient, this would represent only 4.4 percent of the population or an average of less than 1 percent of the population receiving such care each year. The obvious discrepancy between ostensible prevalence and treatment rates raises important questions about the clinical significance of the prevalence rates (Regier et al. 1998), barriers to treatment demand, assessments of "medical necessity" for care (Regier et al. 2000), and the cost of failure to provide treatment (Rupp 1995). However, such larger public health issues remain somewhat outside the scope of this important paper.

The authors do take advantage of a large volume of clinical trials literature on treatment efficacy and the synthesis of this information into treatment guideline standards. They are able to use these well-accepted standards to define "treatment bundles" of specific antidepressant medications and psychotherapies, which are considered comparable in effectiveness for either severe or less severe forms of depression. Since only administrative data sets are available for this study, these treatment bundles of known effectiveness (under ideal research conditions as expressed in treatment guidelines) serve as outcome proxies for price comparisons that are attempting to take episode outcomes into account. Seven different bundles included psychotherapy or antidepressant medication treatment either alone or in various combinations for treatment guideline-determined duration and level of intensity or dose.

The authors note that the proportion of guideline-compatible treatment episodes rose substantially over the five years of experience with the treatment system from 35 percent in 1991 to 56 percent in 1995. The quantities and prices of each treatment bundle were assessed with the finding that the average supply price (PPI equivalent) of the treatment bundles decreased slightly over the five-year period even as there were major shifts in the relative proportions of treatment modalities over this time period. In contrast, the consumer demand price (CPI equivalent) increased by more than 22 percent in a single year (1994–1995) as a result of benefit design changes which increased copayments and deductibles for mental health treatment. The net conclusion from these findings is that the improved compliance with treatment guideline standards in the face of supply price reductions implies a gain in production efficiency within the mental health system. However, the increase in consumer costs to control the "moral hazard" that patients will overuse services beyond their most effective level

has other equity implications. The rise in political support for legislation which would mandate parity in benefits for mental and physical disorders has been propelled by the perceived inequity that is reflected in the reported rise in CPI-equivalent costs for treatment of mental disorders (Zuvekas, Banthin, and Selden 1998; National Advisory Mental Health Council 1998).

The authors are to be commended for their creative analytic approaches to mainstreaming health services in general and mental health services in particular into the PPI and CPI framework. The availability of such a well-respected economic framework for evaluating the production of mental health services could be of considerable value to policymakers in their assessment of the relative cost and value of investing greater resources in mental health services, research, and system development.

As a practicing clinical psychiatrist, I am impressed by the care with which the authors have reviewed the clinical treatment literature and have attempted to model the practice guidelines with administrative claims data sets. They also have recognized the limitations of these claims data for capturing the full picture of treatment for major depression, which has both acute treatment requirements for all, and chronic maintenance treatment requirements for many others. However, the shift in treatment styles from a predominance of psychotherapy alone to antidepressant medication alone or in combination during this time frame does have a clear face validity. The shift to the newer SSRI antidepressant medications is also adequately mirrored in this data set. However, before this model is ready for prime time implementation as a component of published PPIs and CPIs from the BLS, it would be important to have a few replications and refinements.

As an epidemiologist with a great deal of interest in the policy applications of such research data, I am impressed by the potential benefits of obtaining mental disorder prevalence rates which would more accurately reflect the need for specific types of treatment (e.g., treatment bundles). The current movement in psychopathology assessment is to obtain more specific measures of clinically significant impairment and symptom levels (routinely used in treatment efficacy studies) in addition to existing diagnostic criteria. If such fine-grained diagnostic criteria could be reflected in epidemiological studies as well as in the diagnostic codes recorded in claims records, the value of the indexes proposed in this paper could obviously be advanced.

Although many other health policy issues are touched on by this paper, including the issue of the optimal cost-sharing for patients seeking mental health services, the most important policy emphasis is the focus on outcome or quality of care rather than simply cost of care. With the current managed care industry competition "race to the bottom" on per-member-per-month costs, it is most important to have cost measures which are

adjusted for quality as determined by the outcome of treatment episodes. It is somewhat unexpected to have economists leading the way by identifying methods which would tilt the industry toward quality competition when clinicians and clinical research investigators have been so slow to respond to this challenge. I hope that the obvious benefits of collaboration between economists, clinical investigators, epidemiologists, and other health services researchers will result in refining the outstanding work represented in this paper to the point where it will become a standard economic index for the field.

References

American Psychiatric Association (APA). 1993. Practice guidelines for major depressive disorder in adults. *American Journal of Psychiatry,* no. 150, no. 4 (suppl.): 1–51.

Kessler, R. C., K. A. McGonagle, S. Zhao, C. B. Nelson, M. Hughes, S. Eshleman, H. U. Wittchen, and K. S. Kendler. 1994. Lifetime and twelve-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry* 51: 8–19.

National Advisory Mental Health Council. 1998. *Parity in financing mental health services: Managed care effects on cost, access, and quality. An interim report to Congress by the National Advisory Mental Health Council.* NIH Publication no. 98–4322. Bethesda, Md.: National Institutes of Health.

Regier, D. A., C. T. Kaelber, D. S. Rae, M. E. Farmer, B. Knauper, R. C. Kessler, and G. S. Norquist. 1998. Limitations of diagnostic criteria and assessment instruments for mental disorders: Implications for research and policy. *Archives of General Psychiatry* 55: 109–15.

Regier, D. A., W. E. Narrow, D. S. Rae, R. W. Manderscheid, B. Z. Locke, and F. K. Goodwin. 1993. The de facto U.S. mental and addictive disorders service system. *Archives of General Psychiatry* 50: 85–94.

Regier, D. A., W. E. Narrow, A. Rupp, D. S. Rae, C. T. Kaelber. 2000. The epidemiology of mental disorder treatment need: Community estimates of medical necessity. In *Unmet need in mental health service delivery,* ed. G. Andrews and S. Henderson. London: Cambridge University Press.

Rogers, W., K. Wells, L. Meredith, R. Sturm, and M. A. Burnam. 1993. Outcomes for adult outpatients with depression under prepaid or fee-for-service financing. *Archives of General Psychiatry* 50: 517–25.

Rupp, A. 1995. The economic consequences of not treating depression. *British Journal of Psychiatry* 166, no. 27 (suppl.): 29–33.

Scitovsky, Anne A. 1962. An index of the cost of medical care: A proposed new approach. In *The economics of health and medical care: Proceedings of the Conference on the Economics of Health and Medical Care,* 10–12 May. Ann Arbor: University of Michigan.

Zuvekas, S. H., J. S. Banthin, and T. M. Selden. 1998. Mental health parity: What are the gaps in coverage? *Journal of Mental Health Policy and Economics* 1: 135–46.