Specialty Health Care, Treatment Patterns, and Quality: The Impact of a Mental Health Carve-Out on Care for Depression

Susan H. Busch

Objectives. To assess the effect of a mental health carve-out on treatment patterns and quality of care for outpatient treatment of depression.

Data Sources. Outpatient and pharmaceutical claims from September 1993 through March 1997 for one large managed care organization in the Midwest that carved-out mental health and substance abuse benefits in September 1995.

Research Design. Using the treatment episode as the unit of analysis ($n = 1,747$), changes in treatment patterns associated with the change to a carve-out were evaluated. Logistic regression was used to assess whether in the postperiod a treatment episode was more likely to be treated with (1) an antidepressant and (2) a type and intensity of treatment with proven efficacy. To strengthen confidence in a causal relationship, I search for structural breaks in treatment patterns across a wide range of dates, assuming no a priori knowledge of the timing of the impact of the carve-out.

Results. I find the carve-out to be associated with an increase in the use of drug treatments. Although I find a decrease in the use of guideline-level treatment over the entire study period, there is an increase in the number of episodes treated with guideline-level treatment over what would be the case in the absence of the carve-out.

Conclusions. The increase in the use of drug treatments suggests previous research that excluded these costs may have overestimated the savings attributable to carve-outs. Guideline-level care appeared to increase as a result of carve-out implementation suggesting the use of management and specialization to reduce costs is not antithetical to quality improvement.

Key words. Carve-out, managed care, mental health, quality

Carve-outs, the management of mental health care by firms that are legally and administratively separate from the firm managing general medical care, have become common in both the public and private health care sectors (Grazier and Eselius 1999). Under a mental health carve-out, an employer, health plan, or state Medicaid program generally contracts with a single managed behavioral health organization (MBHO) to administer, manage, and often insure all necessary mental health care to enrollees. The rapid increase in

mental health specialty care has been driven at least in part by the observation that MBHOs reduce mental health spending by as much as 40 percent (Callahan et al. 1995; Frank and McGuire 1997; Goldman, McCulloch, and Sturm 1998; Grazier et al. 1999; Grazier and Eselius 1999; Huskamp 1999; Ma and McGuire 1998; Sturm 1999b). Additional research has shown that this is not a one-time savings and costs continue to decline as the carve-out gains experience (Goldman, McCulloch, and Sturm 1998; Grazier et al. 1999; Ma and McGuire 1998; Sturm 1998).

The implications of this dramatic organizational change are still not completely understood. While MBHOs appear to deliver on promises to contain costs, dramatic reductions in spending raise concerns about the effect of MBHOs on quality of care. If reduced spending is due to increased contracting efficiencies, the use of strict clinical criteria to eliminate unnecessary treatment, or increased use of effective but less costly treatments, concerns about quality should be allayed. Conversely, if reduced spending is due to across-the-board denials, drastic reductions in service provision to those with severe conditions, or a decrease in appropriateness of care, concerns about quality may be warranted.

Empirical evidence assessing the impact of carve-outs on quality of care is limited. Merrick (1998) studied Massachusetts employees discharged from hospitals with major depression and found increases in follow-up care rates and no significant change in 15-day and 30-day readmission rates, suggesting an improvement in quality in the carve-out. Sturm (1999b) examined continuing cost declines in 52 plans of one MBHO and found no correlation between reductions in spending, 30-day follow up rates after hospital discharge, and 30-day readmission rates. The Medical Outcomes Study, a large observational study that followed depressed patients for two years, compared patients whose care was financed fee-for-service with patients whose care was prepaid, a common financing arrangement in mental health carve-outs. These researchers found that patients of psychiatrists in prepaid care were more likely to develop new functional limitations (Wells et al. 1996). While these studies suggest concerns about quality may be unfounded, the narrow indicators considered suggest the need for additional research.

In this paper, I examine outpatient treatment for depression, the most prevalent and costly mental illness (Kessler et al. 1994). The analyses that follow differ from previous research in that I link patient's outpatient and pharmacy claims to construct treatment episodes. Because prescription drugs are an effective and common treatment for mental illness, this allows a more complete view of the treatment received. By examining the receipt of antidepressant drugs, I highlight that drug costs may impact estimates of savings. First, I examine changes in the use of antidepressant medication to assess whether previous estimates of cost savings that exclude the cost of drugs may be overstated. To assess quality, I examine whether patients are more likely to receive treatments consistent with published guidelines post-carve-out implementation. To strengthen confidence in a causal relationship, I search for structural breaks in treatment patterns across a wide range of dates, assuming no a priori knowledge of the timing of the impact of the carve-out.

METHODS

Data

This natural experiment examines the experience of one plan of a large national managed care organization that carved-out mental health and substance abuse care on September 1, 1995. At this time, the plan contracted with an MBHO, hereafter referred to as the vendor, to manage and insure all outpatient and inpatient mental health services for the plan's 26,000 point of service (POS) enrollees.

The vendor was at full risk for the cost of all inpatient and outpatient MHSA (mental health/substance abuse) care, implying a strong financial incentive to limit care and reduce medically unnecessary spending. Linking the cost of inpatient and outpatient care under the same capitated payment encouraged the substitution of relatively inexpensive outpatient services for more costly hospitalization where medically appropriate, thereby generating potential savings. The plan retained responsibility for the costs of prescription drugs. Therefore, the vendor had a strong financial incentive to encourage the use of drug treatments since they did not appear on the vendor's budget. Plan enrollees experienced no explicit changes in insurance benefits over the study period.

The data used for these analyses include pharmacy and outpatient claims for the period September 1993 to August 1995 (two years pre-carve-out)

and September 1995 to March 1997 (one and one-half years post-carve-out). Diagnostic information and dates of service in the claims were used to construct episodes of treatment. Since I do not directly observe symptoms in retrospective claims data, I cannot make my claims-based definition of an episode of treatment correspond directly to an episode of illness. Similar to other research (e.g., Kessler et al. 1980, Wells et al. 1996), I consider an eight-week lapse in treatment to signal the start of a new episode. This resulted in 2,427 episodes of treatment for depression.

To increase the homogeneity of the episodes in the analysis and because information regarding procedures and drugs prescribed in the hospital was not available in the hospital claims data, I omitted episodes of patients with inpatient mental health care during the study period. Removing patients hospitalized eliminated 293 episodes from the analysis. Furthermore, some patients were not continuously enrolled in the plan. To ensure I had information on the entire episode of care, I omitted the episode if the patient was not continuously enrolled both eight weeks prior to the start of the episode, and four months postdiagnosis. This eliminated 343 episodes. Another 44 episodes were eliminated due to incomplete information. This left 1,747 episodes of care for the analysis.

Using information on procedures (CPT4 codes) and drugs prescribed, I describe the composition of treatment for each episode. Ten mutually exclusive treatment categories were determined based on evidence from published guidelines on appropriate care for depression (American Psychiatric Association 1993; Agency for Health Care Policy Research 1993). These treatments and their distribution in the data are listed in Table 1. They include psychotherapy alone, drugs alone, and combination treatments. Episodes not assigned to one of these ten treatment categories were examined and three additional categories were created. These include single visit, office visits (no treatment other than several office visits for unspecified procedures), and no treatment (patient had a depression diagnosis and received neither psychotherapy, general office visits, nor filled a prescription). Because a claim may contain two diagnoses, providers may code a mental health diagnosis when the primary reason for the visit is unrelated to mental health care. With claims data it is impossible to determine whether patients in these three categories neglected to follow-up a mental health care referral, did not fill a prescription, were inappropriately not given further treatment, or the visit was a "rule-out" visit (i.e., the provider ruled out a diagnosis of depression during the visit).

Table 1: Distribution of Initial Treatment Assignment (Percent)

<i>Treatment</i>	<i>Full Sample (N = 1,747)</i>	
	<i>Pre-Carve-Out N(%)</i>	<i>Post-Carve-Out N (%)</i>
1 Psychotherapy alone	322 (30%)	70 (11%)
2 Tricyclic drugs alone	17 (2)	20 (3)
3 SSRI drugs alone	123 (11)	134 (21)
4 Heterocyclic drugs alone	7 (<1)	17 (3)
5 Augmenter strategy (e.g., SSRI+Trazodone)	5 (<1)	6 (<1)
6 Psychotherapy combined with TCA drugs	24 (2)	7 (1)
7 Psychotherapy combined with SSRI drugs	105 (9)	44 (7)
8 Psychotherapy combined with heterocyclics	10 (<1)	3 (<1)
9 Psychotherapy and augmenter strategy	12 (1)	3 (<1)
10 Other drugs alone (includes anxiolytics, lithium, maois, and hypnotics)	12 (1)	7 (1)
11 At least two office visits	130 (12)	77 (12)
12 Single office visit	296 (27)	230 (35)
13 No treatments suggested by guidelines for the care of depression	37 (4)	29 (5)
<i>N</i>	1,100	647

A list of specific drugs included in each category is available from the author.

MEASURES

Receipt of Antidepressant Drug Treatment

Because it is difficult to make specific inferences associating very specific changes in treatment assignment to broad changes in incentives, I created an indicator variable to summarize treatments in a manner that reflects the incentives faced by the vendor. Recall that the MBHO is not at risk for the cost of antidepressant drugs. I estimate a logit model of the effect of the carve-out on the probability an initial treatment assignment includes any antidepressant.

Receipt of Guideline-Level Treatment

The measure of quality of care is a dichotomous variable indicating whether the episode was treated with the recommended number of psychotherapy visits or days of drug therapy according to published guidelines (American Psychiatric Association 1993; Agency for Health Care Policy Research 1993). Claims data do not include information on how many days medication was actually taken; as a proxy I used the number of days of treatment for which a prescription was filled. Following the published guidelines, I considered filling

prescriptions for at least 61 days of a drug approved for the treatment of depression as guideline-level treatment. A prescription is usually for a 30-day supply of medication; therefore 61 days implies the patient filled three prescriptions. Generally, the clinical literature indicates that while patients may show some improvement to antidepressant medication by the end of the first week, full response to acute phase depression may take four to six weeks. Some patients begin treatment on one antidepressant drug and subsequently switch to a new drug due to lack of response or side effects. The group of episodes classified as adhering to guidelines includes patients who receive 61 days of the drug initially prescribed as well as patients who switch to a new drug and subsequently receive 61 days of the new medication.

For patients receiving psychotherapy, receipt of six psychotherapy visits was considered sufficient to be categorized as guideline-level care. Clinical trials indicate that individuals can show partial response to psychotherapy within six weeks (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 (Agency for Health Care Policy Research 1993). 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. For patients assigned to combination treatments, fulfilling guideline recommendations for either the drug treatment or psychotherapy was sufficient to be considered guideline-level.

Explanatory Variables

Based on ICD-9 codes, each episode was assigned one of four specific depression diagnoses. These included Major Depression (296.2x, 296.3x), Neurotic Depression (300.4x), Depression Disorder Not Otherwise Specified (311.xx), and Adjustment Disorder with Depressed Mood (309.xx). Some comorbid medical conditions, such as substance abuse, may complicate treatment for depression. To control for the complexity of treatment, a variable that indicated the number of comorbid conditions known to complicate the treatment of depression present was created (specific ICD-9 codes available upon request from the author).

Several studies have suggested that socioeconomic status may affect treatment patterns either through initial treatment choice or treatment adherence. I include data on the average per capita income in the patient's zip code in 1990 (as a proxy for the patient's own income) and the share of

Table 2: Patient Characteristics

<i>Variable</i>	<i>Full Sample (N = 1,747)</i>		<i>Continuously Enrolled Sample (N = 631)</i>	
	<i>Pre-Carve-Out</i>	<i>Post-Carve-Out</i>	<i>Pre-Carve-Out</i>	<i>Post-Carve-Out</i>
Average age (years)	38	40	39	41
Gender, male (%)	32	31	34	33
Average number of comorbid conditions	.27	.27	.34	.34
Diagnosis				
Major depression (%)	39	30	36	31
Neurotic depression (%)	18	14	20	13
Depression not otherwise specified (%)	35	52	34	52
Adjustment disorder (%)	8	3	10	5
Percent in patient's zip code completing some college (%)	48	46	48	47
Average income in patient's zip code	\$36,409	\$34,862	\$36,197	\$35,351
<i>N</i>	1,100	647	331	300

residents in the patient's zip code completing some college (as a proxy for the patient's own education). To control for secular changes in treatment patterns, a variable indicating the month/year in which the episode began was created. Thus, an episode occurring in the first month of the study was assigned the value 1, in the second month the value 2 and so on. Table 2 presents the mean values for the above variables in both the pre- and post-carve-out samples.

STATISTICAL ANALYSIS

To assess the effect of the carve-out on care for depression, logit models estimated both the presence of antidepressant drug treatment and the receipt of guideline-level care. Due to the absence of an identifiable control group, I conducted a pre/post implementation analysis; in both models the key independent variable indicates whether the episode began in the pre- or post-carve-out period. Additional explanatory variables include age, gender, comorbid conditions, depression diagnosis, a time trend, and average demographic characteristics in the patient's zip code (education, income).

In observational research generally, and pre-post comparisons in particular, causality is often an issue. If the underlying population changed at the time the carve-out was implemented, the carve-out may be correlated with changes without being the cause of these changes. To strengthen my confidence that unobserved variables associated with the carve-out were not confounding my results, I reestimated the model on two subsets of the data. First, I only considered patients continuously enrolled for one year pre- and one-and-one-half years post-carve-out implementation ($n = 631$). This eliminated the possibility that results were due to patients leaving or joining the plan in response to the carve-out. Second, I only considered patients with the diagnosis of major depression (ICD-9 codes 296.2 or 296.3). Chart reviews have indicated that the specificity of these two diagnoses is high, that is, the proportion of true positives is high, while the proportion of false positives is low. Also, clinicians could employ a more ambiguous diagnosis such as "depression NOS" or "neurotic depression." That clinicians designated the diagnosis as either 296.2 or 296.3 indicates a conscious act of volition. Thus, this increased the homogeneity of the symptoms of the population considered. The latter analysis is of particular interest because among patients diagnosed with unipolar depression, these diagnoses are associated with the lowest levels of functioning. Moreover, these are generally the illnesses for which the clinical research underlying practice guidelines was conducted and designed.

Thus, all analyses were performed for four samples: the full sample ($n = 1,747$), episodes in the full sample with a major depression diagnosis ($n = 619$), the continuously enrolled sample ($n = 631$), and episodes in the continuously enrolled sample with a major depression diagnosis ($n = 209$).

One criticism of evaluations that compare outcomes in a pre-intervention to a postintervention period is concern that secular changes in the outcome variable will be incorrectly attributable to the intervention. To more rigorously assess whether the carve-out is associated with a change in treatment patterns independent of secular changes, I searched over the entire study period for statistically significant structural breaks in treatment patterns (Andrews 1993; Piehl et al. 1999). If the most significant break occurred at or near the implementation of the carve-out, I increased confidence that the change was independent of other trends and due to the carve-out. This strategy is also appropriate if the timing of the effect of an intervention is uncertain, perhaps due to anticipatory effects or lags in implementation. Assuming no a priori knowledge of the timing of the impact of an intervention, the researcher may determine the timing of the maximum structural break and whether it is significant (Piehl et al. 1999).

For each month of the study, a “break” variable equal to one if the episode began after that date and zero otherwise was created. For each of these 34 break variables the logit model predicting receipt of guideline-level treatment was estimated. Each regression includes the appropriate corresponding break indicator as an explanatory variable. If there was no significant change in treatment patterns in the given month, the coefficient on the break variable will be insignificant. For each month, a Wald test statistic for the null hypothesis that there is no significant change in the intercept term in that month was calculated. The maximum Wald test statistic over all possible months was determined and compared to an adjusted critical value (Andrews 1993). It is important to note that because the Wald test statistic must still surpass an appropriate critical value to indicate a significant effect, this procedure does not necessarily find a significant break.

To allow for adequate data before the first and after the last potential break variable the first month was chosen such that 15 percent of the data was in the preperiod. Symmetrically, the final month was chosen such that 15 percent of the data was in the postperiod (Piehl et al. 1999).

RESULTS

Search for a Shift in Treatment Patterns

Figure 1 summarizes the results of the search for a structural break in receipt of guideline-level treatment for the full sample. Thirty-four logit models predicting receipt of guideline-level treatment were estimated. Each point on the graph in Figure 1 suggests by how much the odds of receiving guideline-level treatment shift if the break were assumed to occur at the month indicated. Each point represents a distinct estimate of whether a break occurred at that time. Thus, this graph does not represent the general trend in treatments from month-to-month. A priori, if no shift occurs, one would expect both positive and negative values that are close to zero and insignificant. If any true shift occurs, one would expect odds ratios of the greatest magnitude when the change actually occurred. Because as the “true break” is approached, the break variable contains more and more data from the true postperiod, one would expect the months prior (post) to the true break to trend toward (away from) the shift.

The value of the chi-square test statistic of a change in the intercept is noted for select months. As seen in the figure, the maximum value of the Wald test statistic is 8.56 and occurs in November 1995, approximately two months after the carve-out begins. Note that the maximum odds ratio (2.04) occurs in

Figure 1: Receipt of Guideline-Level Treatment: Odds Ratio of Potential Break Variables, Full Sample

** represents significant at $p < .05$

March 1996. The associated chi-square test statistic is 7.26, slightly less than the maximum chi-square value of 8.56.

The critical value for this test at the 5 percent significance level is 8.45 (Andrews 1993). Comparing this value to the maximum value found in this data allowed me to reject the null hypothesis of no change in the intercept in favor of a statistically significant structural break in the data in November 1995. The magnitude of the coefficient indicates an odds ratio of 1.95 or an approximate doubling of the odds of receiving guideline-level treatment. The timing of the break suggests that changes in quality of care attributable to the carve-out were not realized until two months after the carve-out officially began.

Due to the high prevalence of my dependent variables in this population, throughout the remaining results and discussion sections odds ratios are converted to relative risk ratios (Zhang and Yu 1998; Schwartz, Woloshin, and Welch 1999).

Receipt of Antidepressant Drug Treatment

The results of the logit regression estimating the use of antidepressant drugs in the post- compared to the pre-carve-out period are reported in Table 3.

Table 3: Estimation Results: Receipt of Antidepressant Drug Treatment¹

	<i>Full Sample</i>		<i>Continuously Enrolled</i>	
	<i>All Episodes</i>	<i>Only Major Depression</i>	<i>All Episodes</i>	<i>Only Major Depression</i>
Carve-out	1.61** (.351)	2.41** (.897)	1.30 (.468)	3.70** (2.31)
Age	1.08** (.035)	1.10* (.059)	1.07 (.054)	1.13 (.097)
Age squared	.99** (.000)	.99** (.000)	.99* (.000)	.99* (.001)
Number of comorbidities	1.00 (.134)	1.08 (.251)	1.16 (.246)	1.01 (.371)
Diagnosis of major depression	3.20** (1.00)		.37** (.115)	
Diagnosis of neurotic depression	1.53 (.514)		.75 (171)	
Diagnosis of depression NOS	2.8** (.870)		.24** (.114)	
Time (months)	1.00 (.015)	1.00 (.024)	.90 (.070)	.86 (.112)
Time squared	1.00 (.000)	1.00 (.000)	1.00 (.001)	1.00 (.002)
College	.67 (.399)	.48 (.465)	.90 (.887)	.78 (1.450)
Gender	1.03 (.138)	.88 (.191)	1.01 (.232)	.71 (.271)
Income	1.00 (.000)	1.00 (.000)	1.00 (.000)	.99 (.000)
<i>N</i>	1,747	619	631	209

¹ Results are reported as odds ratios.

Robust standard errors in parentheses.

Models account for repeat observations for some enrollees.

Major depression includes ICD9 diagnostic codes 296.2 and 296.3.

**indicates $p < .05$, *indicates $p < .10$

Results in all four regressions suggest the carve-out was associated with an increase in the use of antidepressant drugs. Results for the full sample indicate that in the carve-out regime the relative risk of an episode being treated with drugs is 37 percent greater in the post-carve-out period (odds ratio = 1.61). Episodes in the full sample with a diagnosis of major depression are 66 percent more likely to be treated with a drug treatment after the carve-out (odds ratio = 2.41). Both these results are significant ($p < .05$). In the continuously

enrolled sample, the coefficient on the carve-out variable is positive but not statistically significant. However, this could be a power issue. The power to detect a 10 percent difference for the sample indicated is 70 percent, somewhat less than the 80 percent traditionally suggested.

When the continuously enrolled sample is limited to just those patients with major unipolar depression, the estimated coefficient for the carve-out indicator is positive and significant. The magnitude of the coefficient suggests that in the post-carve-out period, individuals with unipolar depression were 90 percent more likely to be prescribed a drug than in the pre-carve-out period (odds ratio = 3.70). This suggests that the incentive to use drug treatments appeared stronger for patients with more severe forms of depression.

Receipt of Guideline-Level Treatment

The unadjusted proportion of episodes receiving guideline-level treatment in the pre- and post-carve-out period are presented in Table 4. In the year prior to the carve-out, 36 percent of episodes were treated at guideline levels and this declined in the post-carve-out period to 31 percent. While the unadjusted numbers indicate a decline in the receipt of guideline-level treatment, the multivariate analyses indicate the opposite effect. Over time there is a downward trend in the level of guideline treatment. The regression models allow me to control for changes in trends regarding guideline-level treatment provision over time.

Results from the regressions are presented in Table 5. The carve-out variable is determined based on the contract implementation date. These results suggest that, assuming the time trend would continue, episodes were

Table 4: Unadjusted Percentage of Episodes Receiving Guideline-Level Treatment

	<i>Pre-Carve-Out</i>		<i>Post-Carve-Out</i>	
	<i>Year 1</i>	<i>Year 2</i>	<i>Year 1</i>	<i>Year 2</i>
Full Sample				
All episodes	40%	36%	31%	22%
Episodes with most severe diagnoses	46	42	43	39
Continuously Enrolled				
All episodes		40	32	27
Episodes with most severe diagnoses		44	54	49

Table 5: Estimation Results: Receipt of Guideline-Level Treatment¹

	<i>Full Sample</i>		<i>Continuously Enrolled</i>	
	<i>All Episodes</i>	<i>Just Major Depression</i>	<i>All Episodes</i>	<i>Just Major Depression</i>
Carve-out	1.47* (.325)	1.36 (.515)	1.35 (.476)	3.21** (1.89)
Age	1.06* (.032)	1.15** (.058)	.98 (.034)	1.11 (.083)
Age squared	.99** (.000)	.99** (.000)	.99 (.000)	.99* (.001)
Number of comorbidities	1.04 (.138)	1.41 (.315)	1.08 (.214)	1.50 (.530)
Diagnosis of major depression	1.15 (.267)	—	1.07 (.386)	—
Diagnosis of neurotic depression	.79 (.201)	—	.61 (.243)	—
Diagnosis of depression NOS	.57** (.132)	—	.46** (.167)	—
Time (months)	.99 (.015)	.99 (.024)	.88* (.062)	.82 (.098)
Time squared	.99* (.000)	.99 (.000)	1.00 (.001)	1.00* (.002)
College	4.37** (2.65)	3.60 (3.39)	3.78 (3.57)	2.27 (3.65)
Gender	1.07 (.141)	1.02 (.212)	.99 (.210)	.86 (.862)
Income	1.00 (.000)	.99 (.000)	1.00 (.000)	.99 (.000)
<i>N</i>	1,747	619	663	216

¹ Results are reported as odds ratios.

Robust standard errors in parentheses.

Models account for repeat observations for some enrollees.

Major depression includes ICD9 diagnostic codes 296.2 and 296.3.

** indicates $p < .05$, * indicates $p < .10$

more likely to be treated with guideline-level treatment under the carve-out regime than they would have been had the carve-out not been implemented. In the full sample (Table 5, column 1), the carve-out resulted in 25 percent more episodes achieving guideline-level treatment than would have occurred absent the carve-out (odds ratio = 1.47). This result is only statistically significant at the 10 percent level. This result is similar in magnitude and still not significant at the traditional 5 percent level for those episodes diagnosed with more severe types of depression and the continuously enrolled sample.

Episodes with major depression in the continuously enrolled sample were 62 percent more likely to be treated with guideline-level treatment post-carve-out (odds ratio = 3.21, $p < .05$).

To interpret the above results and their implications I refer back to Figure 1. As Figure 1 indicates, there is a change in the receipt of guideline-level treatment at the date of the contract implementation but it is not statistically significant at conventional levels. As previously mentioned, with the adjusted critical value the change at two months post-carve-out implies the carve-out is associated with a statistically significant change in quality; it just occurs two months postimplementation. This suggests the t-test described in the above paragraph would leave the researcher unable to reject the null hypothesis due to the lack of information about the precise timing of the intervention, even though a more careful analysis of the data indicate a significant change.

The results in Figure 1 indicate that not only does the maximum structural break in treatment patterns occur soon after the carve-out is implemented, but also that this is the only break associated with a test statistic large enough to reject the null hypothesis of no change in treatment patterns. This further strengthens the inference that this change is caused by the carve-out and not due to unrelated treatment trends.

DISCUSSION

By 1997, 169 million Americans, or 75 percent of individuals with private health insurance, were in some form of managed mental health care, with 71 million individuals covered under at-risk carve-out contracts (Oss, Drissel, and Clary 1998). This is a significant shift from the prior decade when mental health financing was almost exclusively integrated with general medical care. The widespread adoption of carve-outs suggests that purchasers are confident that this model of benefit management offers advantages over the traditional integrated model.

Research indicates carve-outs are associated with a sharp drop in both specialty inpatient and MHS outpatient spending (Goldman, McCulloch, and Sturm 1998; Grazier et al. 1999; Grazier and Eselius 1999; Huskamp 1999; Ma and McGuire 1998). Given the strong financial incentives in many carve-out contracts, quality assurance can be a critical aspect of the contract between the purchaser or plan and the MBHO. Most MBHO contracts require plans to report some performance indicators such as claims

processing rates, phone response time, and facility/staffing. One survey of Fortune 500 firms that use specialty mental health services found that only 36 percent considered quality assurance; only 15 percent used HEDIS quality measures. These survey results indicate clinical quality of care is still unmeasured by most purchasers and that provisions in existing contracts on their own do not require high quality care (Merrick et al. 1999).

Implementation of the carve-out described here was associated with a significant improvement in the use of treatments at guideline level. Yet, the time trend variables indicate a steady decline in the quality indicator over the full 3.5 years studied—both before and after carve-out implementation. This suggests quality was declining over this time period, but shows a strong shift up at carve-out implementation. This decline in quality is troubling and may or may not be independent of the increase in the presence of carved-out mental health care, and the increase in managed care generally. Market forces associated with managed care may be affecting quality in the area. Providers may contract with several managed care firms and their emphasis on cost control may affect general practice style. The data does not allow me to identify the specific causes of this worrisome trend.

Results presented here also suggest that analyses of the impact of program interventions can be misleading if the precise timing of the intervention (in this case, the carve-out) is not known, a common problem in program evaluation. In the absence of a careful analysis of all the data, this research would have failed to find statistically significant changes in quality due to the lag in the timing of the impact of the intervention. The timing of the break suggests that changes in quality of care attributable to the carve-out were not realized until two months after the official start of the carve-out. This lag in effect is not surprising. The vendor may have assumed responsibility for MHSA care before implementation of all systems was complete, providers may need several months to understand the goals and protocols of the vendor, or the vendor may intentionally implement new systems slowly to ease concerns or complaints during the transition period.

The finding of an increase in drug treatments suggests that prior research has potentially overestimated the savings attributable to MBHO contracts by not including the costs of prescription drugs. The increase in the use of drug treatments found in this study suggests an increase in the aggregate costs of drugs in the postperiod is likely. Average prices for two treatment bundles illustrate this effect. Using the prices attached to the claims, I calculated the average price for two treatment bundles. In the pre-carve-out period the average cost of the treatment psychotherapy alone and Selective Serotonin

Reuptake Inhibitor (SSRI) alone were \$347 and \$257 respectively. Of the \$257 cost of the SSRI alone treatment, \$212 is the cost of the medication, with the rest due to medical management or office visits. If treatment assignment for an episode was switched from psychotherapy alone to SSRI alone, the actual cost savings for the plan is \$90, or 35 percent. Because the vendor was not responsible for the costs of prescription drugs, the savings attributed to the vendor would be \$302—the full cost of psychotherapy alone (\$347) less the cost of medical management or office visits associated with use of an SSRI (\$45). Costs appear to be reduced substantially more than the true savings if drug costs are not considered.

This analysis supports the conclusion that carving-out mental health services can improve the quality of care received. Yet, it is important to carefully consider the limitations of this study when interpreting these results. Perhaps the greatest limitation is the potential lack of generalizability of these results to other carve-out programs. This analysis reports on the effect on quality of a single carve-out vendor for a single managed care plan. The exclusion of prescription drugs from the at-risk contract may have made it financially beneficial for the firm to improve quality. One might expect different effects depending on the design of the pre-carve-out plan, the vendor chosen and the initial quality level.

A second limitation is the use of a proxy for broad changes in outcomes. Ideally, in judging the effect on quality of any intervention one would measure changes in risk-adjusted outcomes. The definition of guidelines used in this analysis is carefully tied to the clinical literature and indicates episodes were treated with an efficacious treatment. In a naturalistic setting the effectiveness of these treatments may deviate from outcomes obtained in clinical trials. Moreover, important welfare changes may not be captured. The effectiveness of all treatments was considered equal, yet a shift to treatments associated with more side effects may be a welfare decline for an individual patient. Psychotherapy treatment in a research setting may have better outcomes than in a naturalistic setting due to more provider training, the act of monitoring providers, or the uniformity of patient characteristics. Other research has considered the effectiveness of treatments for depression by assigning treatment ratings of an expert panel (Frank et al. 1999; Berndt et al. 2002). At present, expert ratings are only available for approximately 65 percent of episodes. Thus, limiting this study to rated treatments would greatly reduce the sample size. Because the treatments to be rated were systematically chosen, limiting these analyses to rated treatments may introduce sample bias, suggesting the use of these ratings is less appropriate for this study.

Econometric issues also suggest some limitations of this study. Of studies of this type, migration of the population and patient selection are of concern. Although not providing decisive evidence, the similarity of the results in the full sample and the continuously enrolled sample reduce concern that the findings are simply an artifact of selection. A larger sample with more detailed patient information would have provided more decisive results. Another limitation is that I did not consider patients who have been hospitalized. Thus, it is possible that declines in inpatient care may have biased the population considered in this analysis. Also, more difficult or treatment resistant cases may be more likely to be hospitalized—exclusion of these patients may miss important quality concerns. A final limitation is the reliance on claims data, which is always subject to potential underreporting or coding problems.

In the carve-out studied, patients diagnosed with depression were approximately 25 percent more likely to receive guideline-level treatment than in the absence of the carve-out. The results presented here are promising in that individuals with mental illness may be less able to advocate for their own interests than individuals with other disorders. Given that, the increase in quality suggests that some management techniques hold promise even for vulnerable populations. For one disease for which clear and decisive evidence about treatments and their effects are available, carving-out mental health care can have a beneficial effect on quality.

ACKNOWLEDGMENTS

Financial support from the National Institute for Mental Health, Harvard University, the Brookings Institution, and the National Bureau of Economic Research is acknowledged. I am grateful to Richard Frank, Haiden Huskamp, Ernst Berndt, David Cutler, Julie Gazmararian, and two anonymous reviewers for helpful comments. Remaining errors are mine alone.

REFERENCES

- Agency for Health Care Policy Research, Depression Guideline Panel. 1993. *Depression in Primary Care* 2. AHCPR publication 93-0551. Rockville, MD: U.S. Department of Health and Human Services.
- American Psychiatric Association. 1993. "Practice Guidelines for Major Depressive Disorder in Adults." *American Journal of Psychiatry* 150 (4, supplement): 1–26.

- Andrews, D. K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61 (4): 821-56.
- Berndt, E., A. Bir, S. H. Busch, R. G. Frank, and S. T. Normand. 2002. "The Medical Treatment of Depression, 1991-1996: Productive Inefficiency, Expected Outcome Variations, and Price Indexes." *Journal of Health Economics* 21 (2002): 373-96.
- Callahan, J. J., D. S. Shepard, R. H. Beinecke, M. J. Larson, and D. Cavanaugh. 1995. "Mental Health/Substance Abuse Treatment in Managed Care: The Massachusetts Medicaid Experience." *Health Affairs* 14 (3): 173-84.
- Frank, R. G., and T. McGuire. 1997. "Savings from a Medicaid Carve-Out for Mental Health and Substance Abuse Services in Massachusetts." *Psychiatric Services* 48 (9): 1147-52.
- Frank, R. G., T. G. McGuire, S. L. Normand, and H. H. Goldman. 1999. "The Value of Mental Health Care at the System Level: The Case of Treating Depression." *Health Affairs* 18 (5): 71-88.
- Goldman, W., J. McCulloch, and R. Sturm. 1998. "Costs and Use of Mental Health Services before and after Managed Care." *Health Affairs* 17 (2): 40-52.
- Grazier, K. L., L. L. Eselius, T. W. Hu, K. K. Shore, and W. A. G'Sell. 1999. "Effects of a Mental Health Carve-Out on Use, Costs, and Prayers: A Four-Year Study." *Journal of Behavioral Health Services and Research* 26 (4): 381-9.
- Grazier, K. L., and L. L. Eselius. 1999. "Mental Health Carve-Outs: Effects and Implications." *Medical Care Research and Review* 56 (2, supplement): 37-59.
- Huskamp, H. 1998. "How a Managed Behavioral Health Care Carve-Out Plan Affected Spending for Episodes of Treatment." *Psychiatric Services* 49 (12): 1559-62.
- . 1999. "Episodes of Mental Health and Substance Abuse Treatment under a Managed Behavioral Health Care Carve-Out." *Inquiry* 36 (2): 147-61.
- Kessler, L. G., D. M. Steinwachs, and J. R. Hankin. 1980. "Episodes of Psychiatric Utilization." *Medical Care* 8 (12): 1219-27.
- Kessler, R. C., K. A. McGonagle, S. Zhao, C. B. Newlson, M. Hughes, S. Eshlerman, J. Wittchen, and K. S. Endler. 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.
- Ma, C. A., and T. G. McGuire. 1998. "Costs and Incentives in a Behavioral Health Carve-Out." *Health Affairs* 17 (2): 53-6.
- Merrick, E. L. 1998. "Treatment of Major Depression before and after Implementation of a Behavioral Health Carve-Out Plan." *Psychiatric Services* 49 (11): 1563-7.
- Merrick, E. L., D. W. Garnick, C. M. Horgan, D. Goldin, D. Hodgkin, and M. Sciegaj. 1999. "Use of Performance Standards in Behavioral Health Carve-Out Contracts among Fortune 500 Firms." *American Journal of Managed Care* 5 (Spec): SP81-90.
- Oss, M., A. Drissel, and J. Clary. 1998. "Managed Behavioral Health Care Market Share in the United States, 1997-1998." *Open Minds*. Gettysburg, PA: Open Minds, Inc.

- Piehl, A., S. Cooper, A. Braga, and D. Kennedy. 1999. "Testing for Structural Breaks in the Evaluation of Programs." Working paper no. 7226. National Bureau of Economic Research Working Paper Series. Gettysburg, PA: Open Minds, Inc.
- Schwartz, L. M., S. Woloshin, and H. G. Welch. 1999. "Misunderstandings about the Effects of Race and Sex on Physicians' Referrals for Cardiac Catheterization." *New England Journal of Medicine* 341 (4): 279–83.
- Sturm, R. 1999a. "Tracking Changes in Behavioral Health Care: How Have Carve-Outs Changed Care?" *Journal of Behavioral Health Services and Research* 26 (4): 360–71.
- . 1999b. "Cost and Quality Trends under Managed Care: Is There a Learning Curve in Behavioral Health Carve-Out Plans?" *Journal of Health Economics* 18 (5): 593–604.
- Wells, K. B., R. Sturm, C. D. Sherbourne, and L. S. Meredith. 1996. *Caring for Depression: A RAND Study*. Cambridge, MA: Harvard University Press.
- Zhang, Jun, and Kai Yu. 1998. "What's the Relative Risk? A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes." *Journal of the American Medical Association* 280 (19): 1690–1.