Mingshan Lu
Ching-to Albert Ma

Financial Incentives and Gaming in Alcohol Treatment

This study looks at the effect of performance-based contracting (PBC) on administrative information misreports in substance abuse treatment in Maine. For about 700 alcohol abuse treatment episodes in the period 1990–1995, we constructed clinician report gaming indicators from two data sets: the Maine Addiction Treatment System (MATS) and medical record abstracts. Gaming, in this study, refers to differences in MATS reports and the medical records for an episode. Under PBC, which was implemented in 1992, a provider's financial reward was positively related to treatment outcomes measured by some reports from MATS. We found that the introduction of PBC increased gaming. The data supported the hypotheses that clinicians overstated patient severity at the beginning of treatment episodes, and understated severity at the end.

The concern with cost and quality in the health care market has led to many innovations. In a recent paper, Rosenthal et al. (2004) documented the recent trend of "paying for performance," the use of financial incentives for quality improvements. It is thought that if decision makers can be held accountable, efficiency will improve (Daniels and Sabin 1998; Pawlson and O'Kane 2002; Roper and Cutler 1998; OECD 2002). Accountability requires the availability of information so that appropriate rewards and penalties can be applied. Information gathering on utilization and health outcomes therefore has become a critical component for health care accountability (Smith and York 2004; Dranove et al. 2003; Wedig and Ming 2002; Scanlon 2002).

Although health care plans have been collecting utilization, quality, and performance data for some time, the use of these data for rewards is quite recent. Once this information is used for structuring incentives, can health plans continue to rely on information reported by providers? Clinician gaming, the manipulation of at least some aspects of information to enhance practitioners' own agendas, is a critical issue. As clinicians become more aware of the consequences of accountability and the incentives involved, would gaming be more common? Are information manipulations influenced mainly by financial incentives, or other factors? What determines the extent of gaming?

The literature has recognized the importance of clinician gaming. Many have pointed out that clinicians may misreport information on clients to obtain higher payment from insurers or to improve measured performance (Novack et al.

__Mingshan Lu, Ph.D.,__ is an associate professor in the Department of Economics, University of Calgary. __Ching-to Albert Ma, Ph.D.,__ is a professor in the Department of Economics, Boston University. *This research was funded by a grant (no. R21-AA12886) from the National Institute on Alcohol Abuse and Alcoholism, with additional support from the Alberta Heritage Foundation for Medical Research and the Institute of Health Economics.* *Address correspondence to Prof. Ma at Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215. Email: ma@bu.edu*

Inquiry 43: 34–53 (Spring 2006). © 2006 Excellus Health Plan, Inc.
0046-9580/06/4301–0034
www.inquiryjournal.org

1989; Geron 1991; Morreim 1991; Commons and McGuire 1997; Lu 1999). Studies also have found that clinicians manipulate reports for other reasons: to deal with managed care and provide care the patient needs (e.g., continuation of hospitalization or obtain a long-term placement), to protect a patient (e.g., to avoid jeopardizing the patient's future ability to obtain health or life insurance), to protect themselves against malpractice claims, and to secure a civil commitment for a patient (Rost et al. 1994; Dwyer and Shih 1998; Novack et al. 1989; Sardis 1999; Freeman et al. 1999; Kinghorn 1999; Wynia et al. 2000). Clinician gaming is costly for the health care system. Research showed that after prospective payment was introduced, a significant portion of the Medicare expenditure increase was due to case-mix upcoding ("DRG creep" phenomena; see Carter, Newhouse, and Relles 1990, 1991). Clinician gaming violates distributive justice; the resources a patient wins through her physician's gaming may be lost to others (Morreim 1991; Meskin 2000). Ma and McGuire (1997) showed that when clients and clinicians collude in their report to insurers, such collusion restricts the feasible set of payment schemes and potentially leads to efficiency loss.

In this paper, we examine information manipulation due to financial incentives. It is useful to describe briefly the background of this study (we defer the detailed description to the next section). In 1993, the state of Maine implemented performance-based contracting (PBC). This policy aimed at allocating resources to clinics and centers for substance abuse treatments based on performance (Commons and McGuire 1997; Commons, McGuire, and Riordan 1997). Providers who achieved improved performance in a given fiscal year would be given more resources for the following year. Provider performances were measured by the percentage of clients who reduced their alcohol use, practiced abstinence, found employment, improved social functions, and so on. While PBC was a policy for all substance abuse services, we study only alcohol abuse treatments in this paper.

Beginning in 1989, clinics in the state of Maine receiving any state or federal funds for providing alcohol treatment were required to participate in a standardized information system, the Maine Addiction Treatment System (MATS). Information from MATS formed the foundation for performance-based contracting that started in 1993. We investigate whether the implementation of PBC has affected the veracity of reports in MATS.

To check on veracity, we needed to produce a benchmark for a comparison. This comes from a second data set: abstracts of medical records of clients who were treated by clinicians participating in MATS. For more than 700 episodes of alcohol addiction treatment, we obtained information about clients' treatment outcomes from the medical records. This outcome information then was compared with the corresponding episodes in MATS. In other words, we had two reports on treatment outcomes – both reports made by the same clinician. We found that the implementation of PBC significantly raised the extent of gaming.

In an earlier paper (Lu and Ma 2002), we documented the systematic differences between reports to the Maine information system and medical records in five measures of health and social functioning status: drinking frequency at admission and discharge, employment status at admission and discharge, and termination status. These measures were used in Maine's PBC to construct performance indicators. We found evidence of significant inconsistencies in reports on admission and discharge alcohol use frequencies, but not on employment status at admission and discharge. The evidence of inconsistency on termination status was mixed. These results suggest that clinicians may misreport selectively on measures that are more difficult to verify to avoid embarrassment or financial penalties in potential audits. Our earlier study did not test for specific hypotheses on what determined gaming; neither did it use PBC as an explanatory variable.

In this paper, we explain information misreporting in terms of financial incentives that are introduced by performance-based contracting. When rewards are based on performance, the straightforward reaction perhaps is for clinicians to raise their care quality or quantity so that their clients' health status improves. Nevertheless, the assessment of improvement originates from the reports clinicians submit to the state. By submitting more favorable reports, a clinician may give the impression that performance has improved, even if it is untrue.

Our hypotheses state that positive information manipulation is more common after the implementation of performance-based contracting. The patterns of misreporting are over-reporting of addiction status at the time a client begins a treatment episode and under-reporting of status at the time of episode termination. Our ordered logit and multinomial logit regressions produced results that support these hypotheses.

The patient-physician relationship is complex. When clinicians realize that managed care and incentive mechanisms are potentially interfering with their delivery of care, they may choose to circumvent these restrictions. For example, to gain more resources for clients, clinicians may exaggerate their clients' addiction severity. We control for these other factors that potentially affect report veracity to the extent that our data sets allow.

Our study confirms that health care providers do respond to financial incentives. We also recognize that complex interactions between providers and clients and financial incentives may jointly determine policy outcomes. Data collection and auditing to verify reported information should be considered.

Study Setting: Alcohol Treatment in Maine and Performance-Based Contracting

Performance-based contracting in the state of Maine was intended to be a practical incentive mechanism to improve substance abuse treatment outcomes (Commons and McGuire 1997; Commons, McGuire, and Riordan 1997). Prior to PBC, funding to treatment providers was based on historical levels. Maine perceived a need to change that, and PBC attempted to base part of the funding on treatment outcomes. To prepare for PBC, a standardized information gathering system, the Maine Addiction Treatment System, was initiated in October 1989. Any substance abuse treatment provider in Maine receiving state or federal funds had to participate in MATS, and participation was independent of clients' payment sources.

Following the MATS protocol, a clinician was to report to MATS at the beginning of a treatment episode, and then again at the end. Hence, a record in MATS is based on a treatment episode defined by a clinician. The clinician uses standard forms to collect admission and discharge information such as demographics (age, race, sex, education, marital status), income, employment status, criminal involvement, and health variables (pregnancy, recent medical treatments, etc.). On the clinical side, information about substance abuse severity such as types of alcohol, frequencies, routes of administration, and ages of first use of primary, secondary, and tertiary substances was collected. For services, MATS required information of treatment program and provider, and delivery information such as the number of treatment units and unit cost. At the end of an episode, MATS requested a client's termination status, which could be one of the following: completion of the treatment, referred (further treatment is not appropriate for client at this facility), client discharged without clinic agreement (i.e., client leaves without explanation), noncompliance with rules and regulations and/or client refusal of service/treatment, deceased, incarcerated, moved out of a catchment area, or discharged due to program cut/reduction. MATS requires that "the counselor having the face-to-face contact with the client" complete the forms "either during the session or soon after" (Maine Addiction Treatment System Instruction Manual, 1994, p.2). Nevertheless, it is our understanding that administrative staff at some programs might have completed the MATS forms based on information collected in interviews or in clinical records.

For three years before PBC began, clinicians participated in training and reporting in MATS. After PBC was implemented on July 1, 1992, information in MATS was used to evaluate provider performances in three categories: efficiency, effectiveness, and special populations. Efficiency specifies units of treatment delivered in the contract year. Effectiveness measures changes in client addiction status and social functioning between admission and discharge. PBC uses more than 10 effectiveness measures: alcohol use frequency, employment and employability, criminal involvement, and reduction in problems with family or employers, among others. Special populations concerns service delivery to target populations (women, adolescents, the elderly, and poly alcohol and IV alcohol users).

PBC defined different numbers of indicators and performance standards for different treatment programs, including residential rehabilitation, nonresidential rehabilitation, halfway houses, extended shelters, evaluation, outpatient care, and extended care. A treatment program was deemed to "meet overall standard" if it met minimum performance standards in each performance category (Commons, McGuire, and Riordan 1997).

To illustrate how PBC measured performance, consider two effectiveness measures: abstinence and reduction of use. These are regarded as two main alcohol treatment goals (McLellan and McKay 1998). PBC used admission and discharge alcohol use frequency information in MATS to construct the abstinence and use reduction measures. Abstinence was said to be achieved (recorded as "Yes") if the client did not use any alcohol for 30 days before discharge; otherwise abstinence was not achieved (recorded as "No"). For those clients who did not achieve abstinence at discharge, reduction of alcohol use was said to be achieved (recorded as "Yes") if the client's alcohol use frequency at discharge was less than at admission; otherwise it was not achieved (recorded as "No").

In its fiscal 1993 contracts, Maine's PBC system required 70% of the clients of an outpatient program to achieve abstinence at the time of discharge; otherwise the program was regarded as having failed to meet the abstinence performance standard. This standard was set at different levels for different treatment programs. For example, at least 85% of the residential rehabilitation clients of a program were required to achieve abstinence at the time of discharge. Similarly, to satisfy the reduction of use standard, at least 60% of outpatient clients and 85% of residential rehabilitation clients had to achieve reduction.

Starting from the fiscal year 1993, providers were held accountable for their performance. The contract stated that "allocation of resources for the contract year may be affected by agency performance in the previous year." Good performance in Maine's system would be rewarded by additional block grant funds, technical assistance for developing new services, and retaining surplus contract funds (Commons and McGuire 1997). The focus of Maine's PBC was to assist treatment programs to improve treatment performance. Although low performers rarely were punished by reduced funding, Maine's Office of Substance Abuse (OSA) added special conditions to new contracts to these programs, and switched some to fee-for-service contracts and paid only for services that were delivered (Commons and McGuire 1997).[1] The effect of these conditions was that the financial incentives under PBC were perceived to be real.

Previous studies have shown that effectiveness, as defined by the MATS measures used in PBC, has improved since PBC was implemented (Commons, McGuire, and Riordan 1997). However, Lu (1999) showed that an alcohol relapse indicator, which was constructed from MATS but not used by PBC as a performance measure, did not improve after PBC was implemented. This result supports the hypothesis that PBC encouraged clinicians to simply report a better treatment outcome although clients might not have improved. Using MATS only, Lu (1999) performed an indirect check on clinician gaming.

In this paper, we perform a direct check on clinician gaming by employing a second data source. This is a set of abstracts that Boston University researchers collected directly from clinical records under the supervision of OSA representatives in summer 1996. A scrambling algorithm allowed us to link the medical record abstract and MATS data sets, and our study on gaming is based on the comparison between them.

The medical record abstract data were collected with a number of criteria. First, each client in the medical record abstract data had alcohol abuse as the primary problem at admission; each received outpatient treatments, and had no a priori treatment episodes one year before. Second, the medical record abstract data set was limited to 10 large clinics. The actual data points in the medical record abstract data were obtained by sampling 100 episodes evenly distributed across each fiscal year. This resulted in 988 episodes in the medical record abstract data, covering the period from October 1990 to June 1995. Details of the data collection have been reported in Lu and Ma (2002).

The medical record abstract data set contains information on clients, their visits, and clinicians' judgments. First, the abstract has information of a client's employment status at admission and discharge, as well as use frequency and termination status. These were coded in categories identical to those in MATS. Second, we have information on episode admission and discharge dates. For each episode, the abstract data include the number of taken visits, their exact dates, whether appointments have been kept, the title of the responsible clinician, and the type of treatment in each visit. Finally, we have a number of outcome measures in the medical abstract:

whether abstinence was a stated goal and whether it was achieved at discharge, whether relapse occurred after the previous visit, and whether there was reduction of use. The medical abstract data set also contain clinicians' private judgments on clients' progress in each attended visit toward abstinence or other treatment goals.

Financial Incentives and Information Manipulation

The conceptual framework that underlies our analysis comes from a theory of gaming. Our maintained hypothesis is that clinicians may manipulate the system to achieve their goals. Whereas a simple fee-for-service allows much flexibility and minimal interference on the physician-patient relationship, managed care seeks to impose controls to achieve objectives deemed necessary by health plans or payers. In our study setting, the implementation of PBC can be regarded as containing inducements for cost and quality efficiencies. Against PBC, clinicians may find ways to game the system.

As we have described, PBC intends to set the funding amount to a provider according to the provider's performance. Various indicators have been set up to measure performance, and MATS is the system through which performance information is available. Information manipulation may be used by clinicians to game the system. Among the indicators that PBC uses to evaluate provider performance, reduction in alcohol use and abstinence are commonly regarded as ultimate goals of alcohol abuse treatments. It is expected therefore that PBC puts some emphasis on these indicators. The data used to construct reduction and abstinence indicators are alcohol use at admission and at discharge of an episode; both are available from MATS and the medical abstract data sets.

In our earlier investigation (Lu and Ma 2002), we found evidence that the admission and discharge alcohol use information in MATS and the medical abstract data sets are *significantly inconsistent*. We also found that other variables such as employment status at admission and discharge are *significantly consistent* across the two data sets. Thus, we already know that alcohol use information at admission and discharge was manipulated. How is the manipulation related to incentives?

Consider the performance indicator reduction in alcohol use after an episode of treatment. A reduction in alcohol use can be verified by comparing the use at admission to use at discharge. By reporting to MATS a higher amount of a client's alcohol use at admission, a lower amount at discharge, or both, a clinician makes the PBC evaluation of his performance more favorable. For example, even if a client has not experienced any reduction in alcohol use in a treatment episode, a provider may report to MATS a reduction (by increasing the use amount at admission relative to the actual value or reducing it at discharge). Likewise, to produce an abstinence result, a clinician simply reports an alcohol use of zero at the end of the episode when in fact a client continues to drink.

The implementation of PBC within the sample period allows us to examine the impact of financial incentives on gaming.[2] Our hypotheses are summarized as:

- ☐ Hypothesis 1: Under performance-based contracting, a clinician has a financial incentive to report a higher amount of a client's alcohol use at admission of a treatment episode.
- ☐ Hypothesis 2: Under performance-based contracting, a clinician has a financial incentive to report a lower amount of a client's alcohol use at discharge of a treatment episode.

According to these hypotheses, the time dummy for the implementation of PBC will be positively associated with over-reporting of alcohol use at admission and with under-reporting of alcohol use at discharge. Over-reporting and under-reporting of use are obtained by comparing the information in the MATS data against the medical record abstract data.

Although in this project we focus on financial incentives and information manipulation, many other factors may affect clinicians' reporting practics. Some clinicians may adhere more or less stringently to some medical protocols, and may report information according to these protocols more or less routinely. Clinicians may have a variety of attitudes toward managed care and information solicitation by payers. These differences in attitudes may lead to different reporting practices.

Perhaps most significantly, clinicians interact

Table 1. Clients' characteristics

Clients' characteristics	Percent (n = 988)	Mean	S.D.
Age		31.76	11.63
Male	73.91		
Marital status			
Married	21.32		
Divorced/widowed/separated	32.18		
Single/never married	46.50		
Education (years)		12.37	2.22
Employment			
Full time	28.63		
Not full time	71.27		
Household income (last 30 days) ($)		856.21	847.60
Alcohol use frequency at admission			
At least once per month but less than four days per week	60.20		
At least four days per week	39.80		
Severity of alcohol abuse			
Casual/experimental user	5.89		
Lifestyle-involved user	21.73		
Lifestyle-dependent user	38.78		
Dysfunctional user	19.70		
Undetermined	13.91		
With legal involvement at time of admission	53.50		
Concurrent psychiatric problem	12.59		
Number of prior treatment episodes			
No prior treatment episodes	50.25		
One prior treatment episode	27.92		
Two or more prior treatment episodes	21.83		
Primary payer status			
OSA	26.90		
Medicaid	22.84		
Self-pay	23.65		
Privately insured	18.78		
Other	7.82		
Admitted after PBC was implemented	63.53		
Discharged after PBC was implemented	70.36		
Termination status			
Completed treatment	35.73		
Referred	8.6		
Without clinic agreement	36.34		
Died	.2		
Incarcerated	.71		
Moved/couldn't attend	4.96		
Noncompliance/refused treatment	11.23		
Discharged due to program cut/reduction	.51		
Unknown reason	1.72		
Progress		.18	.22

Note: Other than the "Progress" variable, information reported in MATS was used. Percentages are reported for binary variables; means and standard deviations are reported for continuous variables.

with patients in complex ways. A clinician may report information to a payer to help a patient. For example, to gain more resources or enhance the chance of treatment approval, a clinician may exaggerate a client's substance abuse severity at the beginning of an episode. To help a client re-enter the labor force, a clinician may report a less severe condition at discharge. On the other hand, to ensure care authorization by managed care plans, a clinician may exaggerate a client's addiction severity before referral.

Our data allow us to test for the financial incentives. Performance-based contracting was a well-defined policy implemented in the middle of our sample period. MATS and medical record abstract comparisons produced the discrepancies; the PBC time dummy identified the financial incentives for explaining the discrepancies. Other factors may have caused the discrepancies between the MATS and medical record abstract data sets, but we have less reliable information to identify them. To the extent possible, we include relevant variables to control for other factors that determine gaming. Our primary concern remains the financial incentives and its empirical identification by the PBC implementation time dummy.

Empirical Identification of Gaming under Performance-Based Contracting

Descriptive Statistics

The main characteristics of the 988 clients in our sample are presented in Table 1. The clients had an average age at admission of 32. Seven-

Table 2. Cross frequency table of alcohol use frequency at admission, MATS vs. Record abstract data

Admission alcohol use: MATS	Admission alcohol use: Record abstract data										
	None in past 30 days	Once per month	2–3 days per month	Once per week	2–3 days per week	4–6 days per week	Once daily	2–3 times daily	>3 times daily	Missing	**Total number**
None in past 30 days	***0***	0	0	0	0	0	0	0	0	0	0
Once per month	21	***21***	23	9	3	1	1	3	2	51	135
2–3 days per month	12	10	***24***	13	7	3	2	0	2	45	118
Once per week	8	4	10	***20***	21	7	1	1	4	29	105
2–3 days per week	30	8	16	13	***58***	18	10	3	5	74	235
4–6 days per week	8	6	5	4	19	***29***	7	5	9	33	125
Once daily	13	0	7	4	5	11	***31***	7	8	32	118
2–3 times daily	8	0	0	2	1	1	7	***18***	7	8	52
>3 times daily	7	2	5	2	2	5	9	14	***34***	17	97
Missing	0	0	0	0	1	0	0	0	0	2	3
Total number	107	51	90	67	117	75	68	51	71	291	988

Note: Numbers in bold italicized diagonal indicate consistent reports.

ty-four percent of them were male, but only 21% were married. The clients achieved high school education (12 years) on average. Less than one-third were fully employed at time of admission, with an average household income in the 30-day period before admission lower than $900.

A client's alcohol use frequency was the measure for gaming. In both MATS and medical abstract data, this frequency is coded in nine categories: not drinking in the past 30 days, drinking once per month, two to three days per month, once per week, two to three days per week, four to six days per week, once daily, two to three times daily, or more than three times daily. About 60% of our sample drank at least once per month but less than four days per week; the rest drank at least four days per week. A summary of admission alcohol use frequencies is in Table 1, while the cross-frequency distributions of admission and discharge frequencies in the MATS and medical abstract data are in Tables 2 and 3. As seen in Table 1, about 60% of the clients used alcohol more than once a month but less than four days a week at admission.

Table 3. Cross frequency table of alcohol use frequency at discharge, MATS vs. Record abstract data

Discharge alcohol use: MATS	Discharge alcohol use: Record abstract data										
	None in past 30 days	Once per month	2–3 days per month	Once per week	2–3 days per week	4–6 days per week	Once daily	2–3 times daily	>3 times daily	Missing	**Total number**
None in past 30 days	***465***	7	7	6	5	2	0	0	0	91	583
Once per month	10	***12***	6	3	0	0	1	0	0	30	62
2–3 days per month	11	2	***6***	4	2	2	0	0	1	30	58
Once per week	8	0	1	***7***	10	1	3	0	2	35	67
2–3 days per week	10	3	3	3	***9***	2	3	2	0	47	82
4–6 days per week	2	0	2	3	5	***9***	1	0	0	25	47
Once daily	6	1	0	1	1	1	***5***	0	1	26	42
2–3 times daily	0	0	1	1	1	3	2	***5***	1	7	21
>3 times daily	1	1	1	0	0	0	0	2	***4***	13	22
Missing	1	0	0	0	0	0	0	0	0	3	4
Total number	514	26	27	28	33	20	15	9	9	307	988

Note: Numbers in bold italicized diagonal indicate consistent reports.

The MATS data make available a counselor-assessed alcohol abuse severity measure. Shown in Table 1, about 5% of our sample was assessed as casual or experimental users, 20% as lifestyle-involved users, 40% as lifestyle-dependent users, and 20% as dysfunctional users. Other client information consists of a client's legal involvement, concurrent psychiatric problem at admission, and number of prior treatments. More than half of the clients in our sample had legal involvement, and more than 10% had concurrent psychiatric problems. About half of our sample had no prior treatment episode, about a quarter had one prior treatment episode, and the rest had two or more.

Less than 20% of the clients had private insurance. Medicaid, OSA, and clients' own resources each supported roughly 25% of all clients. Our understanding is that many clients who reported to pay for treatment with their own resource (classified as "self-pay") would rely partly on public support. More than 60% of our sample were admitted after PBC was introduced; about two-thirds were discharged after PBC was introduced.

The clients' treatment completion rate was low, at just over 30%. Incomplete treatments were due to a variety of reasons. More than one third left a treatment program without any explanation ("without clinic agreement"); another 11% refused treatment. The rest terminated treatment because they were referred, deceased, incarcerated, moved out of a catchment area, or discharged due to a program cut/reduction.

The variables in the medical abstract for hypothesis testing were alcohol use at admission and discharge. Besides the alcohol use information, the medical record abstract contained a clinician's judgment on a client's treatment progression throughout the treatment episode. At the end of each attended visit, the clinician reported her view on the client's health improvement: a client's progress toward abstinence would be rated as bad, fair, or good. This visit-based measure can be used to infer a client's chance of achieving abstinence at discharge. For our regressions, we used a summary of these progress reports. A client's progress indicator was defined as the percentage of good progress reports in all attended visits. As shown in Table 1, on average a client in our sample received about 18% good progress reports.

Measuring Gaming

All clients had two reports on their admission and discharge alcohol use frequencies: one from MATS, the other from the record abstract. These two reports were compared to construct gaming indicators. The joint distribution of the admission alcohol use frequencies reported in the MATS and record abstract data is presented in Table 2 (see also Lu and Ma 2002). The entries on the italicized diagonal of the table indicate consistent reports on admission alcohol use frequencies. Below the diagonal, admission alcohol use frequencies in MATS were higher than medical record abstract data; above the diagonal, alcohol use frequencies in MATS were lower. Similarly, Table 3 presents the joint distribution of the discharge alcohol frequencies in the two data sets. A casual inspection of Tables 2 and 3 already suggests inconsistencies between the MATS and medical abstract data sets.

For testing the two hypotheses on the effect of PBC on gaming, we define two gaming-indicator variables, gaming on admission alcohol use frequency G_1 and gaming on discharge alcohol use frequency G_2, as follows:

		Admission alcohol use
$G_1 =$	0	MATS < medical record abstract
	1	MATS = medical record abstract
	2	MATS > medical record abstract
		Discharge alcohol use
$G_2 =$	0	MATS > medical record abstract
	1	MATS = medical record abstract
	2	MATS < medical record abstract

The definitions G_1 and G_2 are asymmetric. Indicator G_1 was defined to be a higher number if MATS reported a higher quantity of alcohol use than the medical record abstract (corresponding to the entries below the diagonal in Table 2), while the opposite was true for G_2 (the entries above the diagonal in Table 3). The asymmetric treatment of the two indicators makes for easier exposition. Hypothesis 1 says that PBC leads to over-reports of alcohol use at admission; hypothesis 2, under-reports at discharge. So with G_1 and G_2 as dependent variables, hypotheses 1 and 2 say that the estimated coefficients of the PBC time dummy will be statistically significant and positive.

Admission and discharge drinking frequencies were missing for about 30% of the sample in the record abstract data. This was due to the nature of medical records (handwritten notes in free formats) as well as a cautious data collection methodology. Our research team was instructed to report the information as unavailable when the uncertainty about accuracy of information was judged to be significant. In our regressions, data points with missing reports were deleted, reducing our valid sample size to a little less than 700. In an earlier study, multiple imputation methods were employed to address the problem of missing values in the record abstract data set. The results on report inconsistencies were found to be robust (Lu and Ma 2002).

Empirical Specifications

We used the standard ordered logit model to test hypotheses 1 and 2. Under each hypothesis, clinicians react to the financial incentives under PBC when making their reporting decisions. The dependent variables were gaming on admission alcohol use frequency (G_1) or gaming on discharge alcohol use frequency (G_2). Therefore, the gaming variables G_1 and G_2 are naturally ordered: the higher are G_1 and G_2, the more favorable the performance reports under PBC's evaluation system. The following is the equation that relates the dependent variables to a set of independent variables in the ordered logit model:

$$
\begin{aligned}
\Pr[G_i = j] &= \Pr\left(\kappa_{j-1} < \sum_l \beta_l X_l < \kappa_j\right) \\
&= \frac{1}{1+\exp\left(-\kappa_j + \sum_l \beta_l X_l\right)} \\
&\quad - \frac{1}{1+\exp\left(-\kappa_{j-1} + \sum_l \beta_l X_l\right)}, \quad j = 0, 1, 2.
\end{aligned} \tag{1}
$$

In all regressions, we included the following independent variables (X): client demographical characteristics such as age, sex (male or female), marital status (single, married, or divorced/widowed/separated), education level, employment status (full time or unemployed), household income in the 30-day period before admission. Also included were measures of the client's case mix: dummy variables indicating whether the client had psychiatric problems or legal involvement; alcohol use severity assessment dummies indicating whether the client was a casual or experimental user, a lifestyle-involved user, a lifestyle-dependent user, or a dysfunctional user; and the number of prior treatment episodes.

The key variable for identifying gaming was the PBC time dummy. As noted earlier, performance-based contracting was implemented in 1993. The PBC dummy for a treatment episode that began after 1993 was assigned a value of one; it was given a value of zero if the episode began before 1993. Hypotheses 1 and 2 will be rejected if the estimated coefficients of the PBC dummy are either insignificant or negative.

As we have discussed, the divergence between MATS and the medical abstracts may well be due to complex patient-clinician interactions besides financial incentives. To control for these, we included variables that may influence such interactions; these are payment source (whether the primary payer was Medicaid, private insurance, self-pay, or the state Office of Substance Abuse), and termination status (whether treatment was complete or not at time of discharge, and whether the client was referred or not). Clients with private insurance may be subject to more strict utilization review and treatment approval. In order to help clients with private insurance get approval for further treatment, clinicians may misreport.

Treatment outcomes may determine whether a clinician needs to game on the discharge frequency. If a client has achieved abstinence, there is no need to misreport for financial reason. To control for treatment outcomes, we included the progress indicator and the number of visits in all regressions on G_2. If a client received many positive progress reports during the treatment program, the client likely achieved a good treatment outcome. Number of visits is one of the most important inputs in the treatment production function, and has been shown to be positively related with treatment outcomes (Lu and McGuire 2002).

Gaming also may be affected by how an entire clinic reacts to financial and other concerns, which may be due to unobserved agency factors such as management style. So we estimated models with and without agency fixed effects.

While our hypotheses naturally point to an ordered logit regression, we also considered a multinomial model. Besides financial incentives, many other factors may affect a clinician's reporting practice. These complex motives for clinician gaming may not suggest an ordinal structure of the gaming measures. For example, a clinician may exaggerate a client's drinking frequency at the beginning of an episode to gain more resources or enhance the chance of treatment approval – a direction of misreporting consistent with financial incentives. However, if a client needs an insurance company's approval for continuation of treatment, a clinician may exaggerate the client's discharge condition to help the client – a direction of gaming opposite to financial incentives. The direction of clinician gaming reflects the aggregate result of financial and other incentives.

We were unable to assess the relative magnitude between financial and other motives. We lacked details to identify complex clinician-patient interactions. Nevertheless, we used a multinomial logit specification as a sensitivity check:

$$\Pr[G_i = j] = \frac{\exp\left(\sum_l X_{il}\beta_{lj}\right)}{\sum_{j=0}^{2} \exp\left(\sum_l X_{il}\beta_{lj}\right)}, \quad j = 0, 1, 2. \quad (2)$$

The same set of independent variables as in model 1 was included in X. We ran regressions on admission and discharge gaming indicators, with and without agency fixed effects. The regression results of ordered and multinomial logit are presented next.

Estimation Results

Table 4 presents the estimation results of the ordered logit model 1 on gaming in admission alcohol use frequency. In the second and third columns, we report the estimated coefficients and z-statistics from the model with agency fixed effects; the fourth and fifth columns show results from the model without fixed effects.

In the fixed-effects model, the estimated coefficient on the PBC dummy was significant and positive. This supports hypothesis 1: PBC significantly encourages clinicians to exaggerate a client's drinking frequency at time of admission. This, however, does not hold in the model without fixed effects. In order to check which model fits our data better, we conducted a likelihood ratio (LR) test on the null hypothesis that the agency dummies are jointly zero in the fixed-effects model. The LR test result was 18.22, rejecting the null hypothesis at a significance level of .10. We concluded that the agency fixed-effects model was more reliable. Using results from the fixed-effects model, we calculated the marginal effect of PBC on clinician gaming. After the implementation of PBC, a clinician's report on an average client's admission drinking frequency in MATS was 9% more likely to indicate higher frequency than that in the clinical record.

The only client demographic variable that was significant in either set of regressions was marital status: clinicians were more likely to exaggerate the admission alcohol use frequency of clients who were married than those who were single. Our data did not allow us to further investigate this finding. In any case, the significant level of the marital status variable was quite marginal.

Clinician-assessed severity levels were significantly correlated with clinician gaming. Compared with casual/experimental and lifestyle-involved users, lifestyle-dependent or dysfunctional users were more likely to receive an exaggerated report on their admission drinking frequencies. When making a gaming decision on admission drinking frequency, a clinician likely takes into account the client's drinking severity. Apparently, gaming is more common with clients who have a more severe alcohol abuse problem. A possible explanation is that clinicians want to ensure that proper resources are allocated to those who need more treatment – that is, those with more severe drinking problems.

None of the variables indicating payment sources, termination status, or referral at discharge was significant. Our data set was probably too small to detect gaming due to altruistic concerns and complex clinician-patient interactions. Finally and importantly, several agency dummies had highly significant coefficients in the fixed-effects model. These results indicate that there were significant variations across agencies in terms of how clinicians reacted to both financial and altruistic incentives in their gaming in Maine's system.

In Table 5, we report the ordered logit estimation results on gaming on discharge alcohol use frequencies. For models with and without fixed effects, the estimated coefficients of the PBC

Table 4. Ordered logit model 1 regression results: gaming on admission alcohol use frequency

	With agency fixed effects		Without agency fixed effects	
	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics
Demographic variables				
Age	−.006	[.008]	−.002	[.008]
Sex	−.275	[.189]	−.248	[.186]
Married	.419	[.230]*	.442	[.226]*
Divorced/widowed/separated	.227	[.204]	.272	[.201]
Education	.005	[.038]	.036	[.035]
Full-time employed at admission	−.054	[.262]	−.001	[.258]
Unemployed at admission	.087	[.233]	.007	[.229]
Household income (last 30 days)	.000	[.000]	.000	[.000]
Client case-mix variables				
With legal involvement at time of admission	.064	[.169]	.008	[.163]
Concurrent psychiatric problem	−.330	[.250]	−.105	[.232]
Casual/experimental user	.293	[.393]	.585	[.370]
Lifestyle-involved user	.416	[.261]	.477	[.252]*
Lifestyle-dependent user	.593	[.245]**	.670	[.235]***
Dysfunctional user	.865	[.285]***	.946	[.269]***
One prior treatment episode	.209	[.177]	.233	[.173]
Two or more prior treatment episodes	.255	[.208]	.282	[.205]
PBC	.367	[.167]**	.227	[.159]
Payment sources				
OSA	.186	[.314]	.147	[.306]
Medicaid	−.357	[.333]	−.172	[.315]
Self-pay	−.010	[.332]	.254	[.310]
Privately insured	.063	[.343]	.110	[.333]
Termination status (treatment completed or not)	.260	[.166]	.270	[.163]*
Referral	−.168	[.234]	−.165	[.198]
Agency fixed effect				
Agency 7	.409	[.364]		
Agency 10	1.139	[.319]***		
Agency 13	−.186	[.412]		
Agency 16	.477	[.323]		
Agency 19	.444	[.317]		
Agency 25	.824	[.358]**		
Agency 34	.680	[.313]**		
Agency 36	.599	[.304]**		
Agency 41	.273	[.341]		
Constant	−.071	[.715]	−2.116	[.643]
Observations	694		694	
Log L	−726.16		−735.22	

*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

dummy support our hypothesis 2: the introduction of PBC significantly encouraged clinicians to report a lower drinking frequency at time of discharge. We calculated the marginal effect of PBC using the estimation results from the fixed-effects model; with the implementation of PBC, a clinician's report on an average client's discharge drinking frequency was 4% more likely to be lower.

As expected, the progress variable, which is

Table 5. Ordered logit model 1 regression results: gaming on discharge alcohol use frequency

	With agency fixed effects		Without agency fixed effects	
	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics
Demographic variables				
Age	.003	[.011]	.002	[.010]
Sex	.111	[.242]	.135	[.239]
Married	.044	[.304]	.092	[.301]
Divorced/widowed/separated	.212	[.265]	.256	[.261]
Education	−.097	[.049]**	−.077	[.046]*
Full-time employed at admission	.244	[.347]	.294	[.343]
Unemployed at admission	−.245	[.310]	−.234	[.305]
Household income (last 30 days)	.000	[.000]	.000	[.000]
Client case-mix variables				
With legal involvement at time of admission	.081	[.216]	−.043	[.210]
Concurrent psychiatric problem	.945	[.350]***	.879	[.343]**
Casual/experimental user	1.284	[.466]***	.953	[.431]**
Lifestyle-involved user	.421	[.335]	.289	[.322]
Lifestyle-dependent user	.192	[.317]	.203	[.307]
Dysfunctional user	.157	[.365]	.155	[.349]
One prior treatment episode	−.167	[.230]	−.226	[.225]
Two or more prior treatment episodes	−.402	[.268]	−.433	[.263]*
PBC	.461	[.220]**	.463	[.211]**
Payment sources				
OSA	.086	[.410]	.100	[.401]
Medicaid	−.228	[.438]	−.144	[.418]
Self-pay	−.274	[.428]	−.058	[.404]
Privately insured	−.163	[.441]	−.172	[.432]
Termination status (treatment completed or not)	−.047	[.220]	.024	[.216]
Referral	.289	[.322]	.199	[.259]
Indicator of progress toward abstinence	−1.330	[.475]***	−1.018	[.441]**
Number of visits	−.002	[.006]	−.001	[.006]
Agency fixed effect				
Agency 7	−.387	[.421]		
Agency 10	−.579	[.408]		
Agency 13	−.564	[.552]		
Agency 16	.679	[.432]		
Agency 19	−.780	[.417]*		
Agency 25	.228	[.473]		
Agency 34	−.043	[.398]		
Agency 36	.129	[.420]		
Agency 41	−.255	[.405]		
Constant	−3.065	[.895]	−2.620	[.808]
Observations	680		680	
Log L	−453.47		−461.24	

*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

a proxy for a client's actual treatment outcomes, was significantly negative. If a client achieved a higher percentage of positive progress reports during treatment, there was less need for the clinician to game on the discharge alcohol use frequency. However, the number of visits, which is another indicator of actual treatment outcomes, was insignificant in both models. None of the demographic variables was significant except education level; clinicians were less likely to under-report the discharge drinking frequencies of clients with higher education. Similar to results on admission frequency, the severity of clients' drinking affects gaming; clinicians were more likely to report a lower discharge drinking frequency on clients with less severe drinking problems. As in Table 4, none of the insurance source, termination status, or referral variables was significant.

Regression results on the multinomial logit model 2 are reported in Tables 6 and 7. The coefficients of the PBC dummy were all positive and many were significant. We calculated the marginal effects using results in the fixed-effects model; after PBC, clinicians' reports in MATS on an average client's admission drinking frequency were 8% more likely to indicate higher frequency than the clinical records; for discharge drinking frequency, 2% were more likely to show lower frequency.

The effects of client case mix on clinician gaming remained consistent with results in Tables 4 and 5. When making the decision on gaming, clinicians took into account potential impacts on a client, avoiding jeopardizing current and future treatment for needy clients. Similar to results in the order logit model, clinicians had a higher tendency to game on admission alcohol use frequency than on discharge. A client's actual treatment outcomes, proxied by the progress indicator, were found to affect gaming on discharge alcohol use frequency as well.

We conducted various sensitivity analyses.[3] First, we checked whether results were sensitive to the gaming definition. Our gaming indicators mentioned earlier were ordinal: they refer to whether admission or discharge alcohol use frequencies in MATS were lower, equal to, or higher than those in the medical record abstract. As a check, we changed to a more cardinal measure, defined in terms of differences:

G_1^C = admission use frequency in MATS − admission use frequency in medical record abstract data;

G_2^C = discharge use frequency in medical record abstract data − discharge use frequency in MATS.

Since alcohol use frequencies were recorded in nine categories, the gaming indicators in the new definitions were categorical variables ranging from −8 to 8. They measured the extent of gaming. Ordered logit regression results using the cardinal gaming measures are reported in Tables 8 and 9.[4] The main results on the impacts of PBC and factors such as insurance and the progress indicator on clinician gaming remained unchanged.

Second, we checked whether PBC's effect on gaming varied according to clients' situations. Did factors such as a client's severity and insurance sources affect clinicians' gaming decisions, and did PBC change these effects? Clinicians may be more sympathetic to clients with severe drug use problems, and thus may manipulate reports to help these clients. Now, PBC can potentially change these manipulations since it brings in financial incentives. We tested this hypothesis by including in all regressions the interaction terms of PBC and severity variables. Likewise, clinicians' gaming decisions may depend on clients' payment sources. For example, clinicians may manipulate information more readily when the payer is a fee-for-service insurance company, but less so when the payer is a managed care firm. PBC, again, potentially changes these incentives. To test for this effect, we combined the insurance categories of OSA, Medicaid, and self-pay: the redefined insurance sources became either privately insured or others. Then we included in all regressions an interaction term of PBC and the privately insured. None of these interaction terms in either of the two sets of regressions turned out to be significant. Our sample probably was not big enough to detect secondary effects of PBC. (For brevity, we did not include a table for this and the following results.)

In our fixed-effects models, some agency dummies were significant. To test explicitly the hypothesis that PBC has different effects on the gaming behaviors of clinicians at different clinics, we added in all regressions interaction terms of PBC and agency dummies. Some of these interaction terms were shown to be significant, supporting our hypothesis.

Finally, we tested whether our results on clinician gaming were robust in different subsamples. About 11% of our sample was recorded at time of admission as not drinking in the past 30 days. These "non-users" were clients who had a history of drinking problems and were seeking treatment to prevent relapse, or just were referred from treatment programs such as residential reha-

Table 6. Multinomial logit model 2 regression results: gaming on admission alcohol use frequency

	With agency fixed effects				Without agency fixed effects			
	$G_1 = 1$		$G_1 = 2$		$G_1 = 1$		$G_1 = 2$	
	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics
Demographic variables								
Age	.006	[.012]	−.005	[.012]	.01	[.011]	.002	[.011]
Sex	−.518	[.281]*	−.432	[.278]	−.551	[.275]**	−.414	[.272]
Married	−.068	[.332]	.460	[.317]	−.058	[.322]	.490	[.308]
Divorced/widowed/separated	−.342	[.297]	.122	[.288]	−.280	[.291]	.222	[.281]
Education	.018	[.053]	.015	[.052]	.032	[.049]	.051	[.048]
Full-time employed at admission	.343	[.383]	.039	[.372]	.285	[.373]	.054	[.361]
Unemployed at admission	.056	[.344]	.172	[.333]	−.094	[.333]	.012	[.322]
Household income (last 30 days)	.000	[.000]*	.000	[.000]	.000	[.000]**	.000	[.000]
Client case-mix variables								
With legal involvement at time of admission	−.552	[.244]**	−.090	[.236]	−.635	[.234]***	−.160	[.227]
Concurrent psychiatric problem	.103	[.359]	−.414	[.363]	.065	[.338]	−.174	[.338]
Casual/experimental user	−.381	[.539]	.236	[.523]	−.194	[.505]	.692	[.479]
Lifestyle-involved user	−.151	[.357]	.482	[.360]	−.047	[.340]	.604	[.344]*
Lifestyle-dependent user	.179	[.335]	.768	[.343]**	.319	[.319]	.914	[.329]***
Dysfunctional user	.528	[.404]	1.164	[.402]***	.578	[.380]	1.310	[.381]***
One prior treatment episode	.019	[.252]	.257	[.246]	.079	[.246]	.282	[.239]
Two or more prior treatment episodes	.083	[.309]	.358	[.295]	.045	[.301]	.342	[.287]
PBC	.309	[.242]	.507	[.236]**	.158	[.225]	.294	[.219]
Payment sources								
OSA	.285	[.454]	.347	[.441]	.286	[.439]	.286	[.426]
Medicaid	−.551	[.489]	−.504	[.479]	−.452	[.451]	−.262	[.440]
Self-pay	−.073	[.493]	.006	[.477]	.100	[.448]	.356	[.435]
Privately insured	−.041	[.506]	.054	[.490]	−.121	[.480]	.098	[.466]
Termination status (treatment completed or not)	.089	[.239]	.335	[.230]	.182	[.234]	.371	[.225]*
Referral	.087	[.340]	−.167	[.331]	.155	[.286]	−.151	[.281]
Agency fixed effect								
Agency 7	−.018	[.481]	.590	[.503]				
Agency 10	.064	[.465]	1.465	[.458]***				
Agency 13	−.619	[.581]	−.240	[.577]				
Agency 16	−.847	[.454]*	.521	[.434]				
Agency 19	−.335	[.438]	.574	[.446]				
Agency 25	.835	[.554]	1.504	[.565]***				
Agency 34	.088	[.458]	.957	[.463]**				
Agency 36	−.292	[.424]	.764	[.427]*				
Agency 41	.330	[.478]	.462	[.525]				
Constant	.577	[1.020]	−1.150	[1.025]	.257	[.912]	−1.228	[.893]
Observations	694				694			
Log L	−704.45				−722.08			

*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

Table 7. Multinomial logit model 2 regression results: gaming on discharge alcohol use frequency

	With agency fixed effects				Without agency fixed effects			
	$G_2 = 1$		$G_2 = 2$		$G_2 = 1$		$G_2 = 2$	
	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics
Demographic variables								
Age	−.014	[.015]	.015	[.019]	−.019	[.014]	.011	[.018]
Sex	.116	[.330]	.086	[.437]	.139	[.316]	.105	[.422]
Married	−.238	[.415]	.080	[.554]	−.134	[.403]	.154	[.530]
Divorced/widowed/separated	.554	[.377]	.518	[.509]	.513	[.365]	.485	[.497]
Education	−.116	[.069]*	−.152	[.090]*	−.108	[.065]*	−.147	[.086]*
Full-time employed at admission	−.196	[.513]	1.039	[.686]	−.298	[.496]	.948	[.496]
Unemployed at admission	−.007	[.441]	−.441	[.612]	−.059	[.430]	−.457	[.592]
Household income (last 30 days)	.000	[.000]	.000	[.000]	.000	[.000]	.000	[.000]
Client case-mix variables								
With legal involvement at time of admission	.365	[.302]	−.206	[.413]	.276	[.294]	−.281	[.402]
Concurrent psychiatric problem	−.040	[.459]	1.092	[.534]**	−.093	[.429]	1.040	[.498]**
Casual/experimental user	1.388	[.877]	2.760	[1.054]***	.973	[.840]	2.446	[1.002]**
Lifestyle-involved user	.599	[.480]	1.288	[.670]*	.433	[.453]	1.154	[.634]*
Lifestyle-dependent user	.347	[.431]	.447	[.601]	.416	[.410]	.587	[.567]
Dysfunctional user	.068	[.491]	.562	[.672]	.204	[.458]	.782	[.634]
One prior treatment episode	−.114	[.338]	−.154	[.446]	−.271	[.321]	−.281	[.425]
Two or more prior treatment episodes	−.267	[.368]	−.799	[.504]	−.387	[.346]	−.850	[.478]*
PBC	.351	[.288]	.757	[.396]*	.440	[.278]	.878	[.384]**
Payment sources								
OSA	−.833	[.882]	.620	[1.171]	−.789	[.831]	.749	[1.131]
Medicaid	−1.855	[.857]**	.219	[1.137]	−1.761	[.812]**	.275	[1.105]
Self-pay	−2.167	[.890]**	−.014	[1.185]	−1.722	[.823]**	.436	[1.130]
Privately insured	−1.226	[.889]	−.225	[1.190]	−1.341	[.844]	−.17	[1.153]
Termination status (treatment completed or not)	1.616	[.378]***	−.802	[.596]	1.688	[.369]***	−.712	[.584]
Referral	.184	[.509]	.404	[.671]	.288	[.459]	.375	[.627]
Indicator of progress toward abstinence	.223	[.681]	−4.365	[1.235]***	.661	[.659]	−3.982	[1.187]***
Number of visits	.033	[.014]**	−.007	[.019]	.035	[.014]***	−.006	[.019]
Agency fixed effect								
Agency 7	−.275	[.571]	−1.046	[.828]				
Agency 10	−.763	[.554]	−.801	[.720]				
Agency 16	.911	[.763]	1.069	[.927]				
Agency 19	−1.349	[.523]***	−1.007	[.689]				
Agency 25	.908	[.775]	.780	[1.020]				
Agency 34	−.398	[.572]	−.006	[.730]				
Agency 36	−.237	[.570]	.145	[.721]				
Agency 41	−.349	[.558]	−.479	[.747]				
Constant	3.436	[1.408]**	.771	[1.885]	3.044	[1.324]**	.258	[1.803]
Observations	680				680			
Log L	−352.33				−363.84			

*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

Table 8. Ordered logit model 1 regression results: gaming on admission alcohol use frequency (using cardinal gaming measure G_1^C)

	With agency fixed effects		Without agency fixed effects	
	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics
Demographic variables				
Age	−.002	[.008]	.002	[.007]
Sex	−.245	[.179]	−.235	[.177]
Married	.369	[.216]*	.416	[.214]*
Divorced/widowed/separated	.074	[.194]	.129	[.192]
Education	.013	[.036]	.043	[.033]
Full-time employed at admission	−.002	[.247]	.043	[.245]
Unemployed at admission	.094	[.223]	.04	[.219]
Household income (last 30 days)	.000	[.000]	.000	[.000]
Client case-mix variables				
With legal involvement at time of admission	.001	[.160]	−.044	[.156]
Concurrent psychiatric problem	−.361	[.244]	−.18	[.228]
Casual/experimental user	.188	[.367]	.446	[.345]
Lifestyle-involved user	.446	[.248]*	.492	[.239]**
Lifestyle-dependent user	.629	[.233]***	.679	[.223]***
Dysfunctional user	.879	[.269]***	.903	[.253]***
One prior treatment episode	.183	[.167]	.199	[.164]
Two or more prior treatment episodes	.339	[.201]*	.367	[.199]*
PBC	.375	[.160]**	.251	[.153]**
Payment sources				
OSA	.12	[.303]	.136	[.295]
Medicaid	−.407	[.322]	−.191	[.304]
Self-pay	.006	[.321]	.284	[.299]
Privately insured	.051	[.331]	.15	[.321]
Termination status (treatment completed or not)	.228	[.157]	.226	[.155]
Referral	−.115	[.229]	−.19	[.191]
Agency fixed effect				
Agency 7	.391	[.348]		
Agency 10	.924	[.293]***		
Agency 13	−.33	[.389]		
Agency 16	.438	[.300]		
Agency 19	.412	[.312]		
Agency 25	.597	[.337]*		
Agency 34	.641	[.297]**		
Agency 36	.56	[.296]*		
Agency 41	.256	[.329]		
Constant	−4.714	[.979]	−4.468	[.9309]
Observations	694		694	
Log L	−1,445.66		−1,453.63	

*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

bilitation or from the legal system (and therefore had no access to alcohol in the past month). In addition, only 37% of the clients completed treatment at time of discharge. Was gaming unduly affected by these subsamples? We re-ran the estimations using the sample with these "non-users" excluded, and the subsample of those who completed treatment. For both sub-

Table 9. Ordered logit model 1 regression results: gaming on discharge alcohol use frequency (using cardinal gaming measure G_2^C)

	With agency fixed effects		Without agency fixed effects	
	Estimated coefficient	Z-statistics	Estimated coefficient	Z-statistics
Demographic variables				
Age	.003	[.011]	.001	[.010]
Sex	.049	[.239]	.081	[.238]
Married	−.014	[.300]	.053	[.299]
Divorced/widowed/separated	.201	[.264]	.252	[.260]
Education	−.098	[.049]**	−.077	[.046]*
Full-time employed at admission	.238	[.343]	.289	[.338]
Unemployed at admission	−.259	[.305]	−.247	[.300]
Household income (last 30 days)	.000	[.000]	.000	[.000]
Client case-mix variables				
With legal involvement at time of admission	.114	[.217]	−.019	[.210]
Concurrent psychiatric problem	.957	[.345]***	.9	[.338]***
Casual/experimental user	1.203	[.458]***	.908	[.425]**
Lifestyle-involved user	.421	[.331]	.297	[.320]
Lifestyle-dependent user	.208	[.314]	.215	[.305]
Dysfunctional user	.161	[.361]	.151	[.347]
One prior treatment episode	−.168	[.229]	−.222	[.225]
Two or more prior treatment episodes	−.43	[.265]	−.459	[.262]*
PBC	.436	[.218]**	.433	[.209]**
Payment sources				
OSA	.08	[.412]	.095	[.403]
Medicaid	−.245	[.437]	−.163	[.418]
Self-pay	−.315	[.429]	−.085	[.405]
Privately insured	−.176	[.443]	−.189	[.434]
Termination status (treatment completed or not)	−.029	[.220]	.04	[.216]
Referral	.292	[.319]	.203	[.258]
Indicator of progress toward abstinence	−1.371	[.475]***	−1.052	[.439]**
Number of visits	−.003	[.006]	−.002	[.006]
Agency fixed effect				
Agency 7	−.365	[.422]		
Agency 10	−.494	[.402]		
Agency 13	−.574	[.551]		
Agency 16	.658	[.429]		
Agency 19	−.774	[.419]*		
Agency 25	.258	[.472]		
Agency 34	−.073	[.401]		
Agency 36	.138	[.420]		
Agency 41	−.253	[.404]		
Constant	−7.810	[1.336]	−7.331	[1.281]
Observations	680		680	
Log L	−705.63		−713.06	

*** Significant at 1% level.
** Significant at 5% level.
* Significant at 10% level.

samples, the impact of PBC on clinician gaming remained significant.[5]

Concluding Remarks

We presented two data sets on client information in alcohol treatment episodes. Each of the two data sets originated from the same clinician. One set consisted of reports required by the Maine Addiction Treatment System for administrative and funding assessments; the other contained abstracts from actual medical records. Performance-based contracting was a policy implemented by the state of Maine around the middle of the period for our data sets. PBC used MATS data to assess program performance, which then determined future funding.

There were discrepancies in the two data sets. We hypothesized that the information inconsistencies were due to clinicians misreporting information to MATS to game the system. The implementation of PBC resulted in a financial incentive for clinicians to misreport information to MATS. Good performance recorded in MATS could be due to over-reporting of a client's alcohol use at admission, or under-reporting at discharge. We identified over-reporting and under-reporting by comparing MATS data against medical record abstracts. We identified gaming incentives by the time dummy of the implementation of PBC.

We found strong empirical evidence for our hypotheses. PBC was found to have a significant and positive effect on clinicians over-reporting client alcohol use at admission and under-reporting at discharge. Because performance evaluation comes from assessing alcohol use reduction and abstinence, these misreporting practices help boost treatment performance where clients actually have not improved.

We were able to directly test gaming because of our unique data sets. The comparison between the MATS and medical abstract data was a straightforward way to identify gaming. The identification of financial incentives on gaming was due to the implementation of PBC. Our results call for attention on provider reactions against incentive mechanisms: information manipulation should not be ruled out when such actions have financial consequences.

Our study suggests two policy implications. First, auditing should be used more often when regulatory authorities must rely on information supplied by providers for financial and funding decisions. Auditing may deter gaming, and gives more reliability to the veracity of reports. Second, establishing a gold standard should be considered whenever it is feasible. In our case, the analysis was possible precisely because we were able to compare the administrative reports against an appropriate standard, namely the medical record abstracts. Having an independent and reliable data source for validating the reliability of administrative data may seem obvious, but appears to have received less emphasis. Data collection methods should consider obtaining the same information in more than one way.

Notes

The authors thank the National Institute on Alcohol Abuse and Alcoholism for its support. Mingshan Lu also thanks the Alberta Heritage Foundation for Medical Research and the Institute of Health Economics. Views expressed in this paper do not reflect those of the funding agencies. At various stages of this project, Priyodorshi Banerjee, Bing Li, Lee McGuire, Jitian Sheu, and Jenny X. Zhang provided excellent research assistance. For their comments and suggestions, the authors thank Hisen-Ming Lien, Thomas McGuire, and many seminar participants. Thanks are also due to one referee and the editor who gave us very helpful suggestions.

1 More specifically, "each program which did not meet the performance criteria was asked to meet with OSA staff to review the performance data and additional data OSA had prepared (fiscal data – projected versus actual unit costs, recidivism and early dropout rates, time in treatment, and measures of client difficulty). OSA and program staff discussed possible actions to be taken to improve performance .... OSA has chosen to contract with certain low-efficiency performers on a fee-for-service basis, ensuring that OSA only pays for services that are actually delivered. Providers with low effectiveness or special population performance have special conditions which address specific indicators included in their 1994 contracts. Finally, in the case of some low overall performers, OSA has only renewed the program's contract for a period of six months" (Commons and McGuire 1997).

2 Like most studies using observational data, there is

no control group. However, our data span less than five years. During this period, Maine's substance abuse treatment system (including treatment population and practice) remained relatively stable. Furthermore, clinicians in Maine participated in training and reporting in MATS as early as October 1989, at least one year before the time of the first episode in our sample. During the first year of implementation of MATS, extensive training and consultation sessions were given by the Maine Office of Substance Abuse to help clinicians learn how to use the new reporting system. We regard the implementation of PBC to be a significant change for the sample period. We were unaware of another significant change in the sample period that would completely confound the effect of PBC.

3 The regression results of all sensitivity analyses are available from the authors upon request.

4 Our sample is too small to obtain multinomial logit regression results on cardinal gaming measures.

5 The subsample with non-users excluded is too small to obtain results of multinomial logit model without fixed effects on gaming on discharge alcohol frequency. The subsample of those who completed treatment is too small to obtain results of the multinomial logit model with or without fixed effects on gaming on discharge alcohol frequency.

References

Carter, M.G., J.P. Newhouse, and D.A. Relles. 1990. How Much Change in the Case Mix Index is DRG Creep? *Journal of Health Economics* 9(4): 411–428.

———. 1991. Has DRG Creep Crept Up? Decomposing the Case Mix Index Change between 1987 and 1988. Working paper. Santa Monica, Calif.: RAND.

Commons, M., and T.G. McGuire. 1997. Some Economics of Performance-Based Contracting for Substance-Abuse Services. In *Treating Alcohol Abusers Effectively*, J.A. Egertson, D.M. Fox, and A.I. Leshner, eds. Malden, Mass.: Blackwell.

Commons, M., T.G. McGuire, and M.H. Riordan. 1997. Performance Contracting for Substance Abuse Treatment. *Health Services Research* 32(5): 631–650.

Daniels, N., and J. Sabin. 1998. The Ethics of Accountability in Managed Care Reform. *Health Affairs* 17(5): 50–64.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite. 2003. Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy* 111(3): 555–588.

Dwyer, J., and A. Shih. 1998. The Ethics of Tailoring the Patient's Chart. *Psychiatric Services* 49(10): 1309–1312.

Freeman, V.G., S.S. Rathore, K.P. Weinfurt, K.A. Schulman, and D.P. Sulmasy. 1999. Lying for Patients. *Archives of International Medicine* 159: 2263–2270.

Geron, S.M. 1991. Regulating the Behaviour of Nursing Homes through Positive Incentives: An Analysis of Illinois' Quality Incentive Program (QUIP). *Gerontologist* 31(3): 292–301.

Kinghorn, W. 1999. Should Doctors Ever Lie on Behalf of Patients? *Journal of the American Medical Association* 282: 1674.

Lu, M. 1999. Separating the "True Effect" from "Gaming" in Incentive-Based Contracts in Health Care. *Journal of Economics and Management Strategy* 8: 383–432.

Lu, M., and C.A. Ma. 2002. Consistency in Performance Evaluation Reports and Medical Records. *Journal of Mental Health Policy and Economics* 5(4): 141–152.

Lu, M., and T.G. McGuire. 2002. The Productivity of Outpatient Treatment for Substance Abuse. *Journal of Human Resources* 37(2): 309–335.

Ma, C.A., and T.G. McGuire. 1997. Optimal Health Insurance and Provider Payment. *American Economic Review* 87(4): 685–704.

Maine Addiction Treatment System Instruction Manual. 1994. Augusta, Me.: Office of Substance Abuse.

McLellan, A.T., and J.R. McKay. 1998. Components of Successful Treatment Programs: Lessons from the Research Literature. In *Principles of Addiction Medicine*, 2nd edition, A.W. Graham and T.K. Schultz, eds. Chevy Chase, Md.: American Society of Addiction Medicine.

Meskin, L.H. 2000. The Noble Lie. *Journal of American Dental Association* 131: 556–560.

Morreim, H.E. 1991. Gaming the System: Dodging the Rules, Ruling the Dodgers. *Archives of Internal Medicine* 151: 443–447.

Novack, D.H., B.J. Detering, R. Arnold, L. Forrow, M. Ladinsky and J.C. Pezzullo. 1989. Physicians' Attitudes toward Using Deception to Resolve Difficult Ethical Problems. *The Journal of the American Medical Association* 261:2980–2985.

Organisation for Economic Co-operation and Development (OECD). 2002. *Measuring Up: Improving Health System Performance in OECD Countries*. Ottawa: Health Canada.

Pawlson, L.G., and M.E. O'Kane. 2002. Professionalism, Regulation, and the Market: Impact on Accountability for Quality of Care. *Health Affairs* 21(3): 200–207.

Roper, W.L., and C.M. Cutler. 1998. Health Plan Accountability and Reporting: Issues and Challenges. *Health Affairs* 17(2): 152–155.

Rosenthal, M.B., R. Fernandopulle, H.R. Song, and B. Landon. 2004. Paying for Quality: Providers' Incentives for Quality Improvement. *Health Affairs* 23(2): 127–141.

Rost, K., R. Smith, D.B. Matthews, and B. Guise. 1994. The Deliberate Misdiagnosis of Major Depression in Primary Care. *Archives of Family Medicine* 3: 333–337.

Sardis, J.B. 1999. Pills, Policies, and Patients. *Health Affairs* 18(5): 156–162.

Scanlon, D.P. 2002. The Impact of Health Plan Report Cards on Managed Care Enrollment. *Journal of Health Economics* 21(1): 19–41.

Smith, P.C., and N. York. 2004. Quality Incentives: The Case of U.K. General Practitioners. *Health Affairs* 23(3): 112–118.

Wedig, G.J., and T. Ming. 2002. The Effect of Report Cards on Consumer Choice in the Health Insurance Market. *Journal of Health Economics* 21(6): 1031–1048.

Wynia, M.K., D.S. Cummins, J.B. VanGeest, and I.B. Wilson. 2000. Physician Manipulation of Reimbursement Rules for Patients. *Journal of the American Medical Association* 283: 1858–1865.