The Journal of Mental Health Policy and Economics
J. Mental Health Policy Econ. **2**, 59–71 (1999)

The Productivity of Mental Health Care: An Instrumental Variable Approach

Mingshan Lu*

Department of Economics, The University of Calgary, Canada

Abstract

Background: Like many other medical technologies and treatments, there is a lack of reliable evidence on treatment effectiveness of mental health care. Increasingly, data from non-experimental settings are being used to study the effect of treatment. However, as in a number of studies using non-experimental data, a simple regression of outcome on treatment shows a puzzling negative and significant impact of mental health care on the improvement of mental health status, even after including a large number of potential control variables. The central problem in interpreting evidence from real-world or non-experimental settings is, therefore, the potential 'selection bias' problem in observational data set. In other words, the choice/quantity of mental health care may be correlated with other variables, particularly unobserved variables, that influence outcome and this may lead to a bias in the estimate of the effect of care in conventional models.

Aims of the Study: This paper addresses the issue of estimating treatment effects using an observational data set. The information in a mental health data set obtained from two waves of data in Puerto Rico is explored. The results using conventional models—in which the potential selection bias is not controlled—and that from instrumental variable (IV) models—which is what was proposed in this study to correct for the contaminated estimation from conventional models—are compared.

Methods: Treatment effectiveness is estimated in a production function framework. Effectiveness is measured as the improvement in mental health status. To control for the potential selection bias problem, IV approaches are employed. The essence of the IV method is to use one or more instruments, which are observable factors that influence treatment but do not directly affect patient outcomes, to isolate the effect of treatment variation that is independent of unobserved patient characteristics. The data used in this study are the first (1992–1993) and second (1993–1994) wave of the ongoing longitudinal study *Mental Health Care Utilization Among Puerto Ricans*, which includes information for an island-wide probability sample of over 3000 adults living in poor areas of Puerto Rico. The instrumental variables employed in this study are travel distance and health insurance sources.

Results: It is very noticeable that in this study, treatment effects were found to be negative in all conventional models (in some cases, highly significant). However, after the IV method was applied, the estimated marginal effects of treatment became positive. Sensitivity analysis partly supports this conclusion. According to the IV estimation results, treatment is productive for the group in most need of mental health care. However, estimations do not find strong enough evidence to demonstrate treatment effects on other groups with less or no need. The results in this paper also suggest an important impact of the following factors on the probability of improvement in mental health status: baseline mental health status, previous treatment, sex, marital status and education.

Discussion: The IV approach provides a practical way to reduce the selection bias due to the confounding of treatment with unmeasured variables. The limitation of this study is that the instruments explored did not perform well enough in some IV equations, therefore the predictive power remains questionable. The most challenging part of applying the IV approach is on finding 'good' instruments which influence the choice/quantity of treatment yet do not introduce further bias by being directly correlated with treatment outcome.

Conclusions: The results in this paper are supportive of the concerns on the credibility of evaluation results using observation data set when the endogeneity of the treatment variable is not controlled. Unobserved factors contribute to the downward bias in the conventional models. The IV approach is shown to be an appropriate method to reduce the selection bias for the group in most need for mental health care, which is also the group of most policy and treatment concern.

Implications for Health Care Provision and Use: The results of this work have implications for resource allocation in mental health care. Evidence is found that mental health care provided in Puerto Rico is productive, and is most helpful for persons in most need for mental health care. According to what estimated from the IV models, on the margin, receiving formal mental health care significantly increases the probability of obtaining a better mental health outcome by 19.2%, and one unit increase in formal treatment increased the probability of becoming healthier by 6.2% to 8.4%. Consistent with other mental health literature, an individual's baseline mental health status is found to be significantly related to the probability of improvement in mental health status: individuals with previous treatment history are less likely to improve. Among demographic factors included in the production function, being female, married, and high education were found to contribute to a higher probability of improvement.

Implication for Further Research: In order to provide accurate evidence of treatment effectiveness of medical technologies to support decision making, it is important that the selection bias be controlled as rigorously as possible when using information from a non-experimental setting. More data and a longer panel are also needed to provide more valid evidence. Copyright © 1999 John Wiley & Sons, Ltd.

Received 20 January 1999; accepted 9 July 1999.

*Correspondence to: Mingshan Lu, Department of Economics, The University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4.
Mingshan Lu is an assistant professor in the Department of Economics and Community Health Sciences and is an Alberta Heritage Population Health Investigator at The University of Calgary. Lu is also a research fellow at the Institute of Health Economics in Alberta. Tel.: 403-220-5488. Fax: 403-282-5262. Email address: lu@calgary.ca

CCC 1091–4358/99/020059–13$17.50

Introduction

In the United States, mental health care and substance abuse treatment amount for about 10% of all personal health care spending.[1] Under the current payment system, both public and private sectors play important roles in the provision of mental health care. In 1990, out of the $54 billion spent on mental health and substance abuse services in the United States, $22.2 billion (41.1%) came from private insurance, $9.5 billion (17.6%) from Medicaid and $2.2 (4.0%) from Medicare. The rest, $20.1 billion (37.2%), was spent on uninsured patients funded by a combination of local, state and federal programmes.[2]

Like many other medical technologies and treatments, there is a lack of reliable evidence on treatment effectiveness of mental health care. After reviewing 91 published studies, Evers, Van Wijk and Ament[3] concluded that there are few good full economic evaluation studies in the domain of mental health care. Evaluating the effectiveness of mental health services is complicated by difficulties such as the reliability of psychiatric diagnosis, and frequent lack of consensus about the etiology and appropriate treatment for many psychiatric illnesses.[3] As a result, many mental health treatments have not been 'proven' effective.

The purpose of this paper is to estimate the effectiveness of mental health care in a non-experimental data setting in a production function framework using instrumental variable. Effectiveness is measured as the improvement in mental health status. The instrumental variables approach provides a practical way to reduce the selection bias due to unmeasured variables. The data used in this study are the first (1992–1993) and second (1993–1994) wave of the ongoing longitudinal *Mental Health Care Utilization Among Puerto Ricans* study (Margarita Alegria, PI), which includes information for an island-wide probability sample of over 3000 adults living in poor areas of Puerto Rico. Using this data set, the case is made for the use of an instrumental variable approach in analyses of data outside controlled settings. As we will see, a simple logit regression based on the conventional evaluation model in the form of a production function shows a negative and significant impact of mental health care on the improvement of mental health status, even after including a large number of control variables, a result which may due to the potential selection biases existing in an observational data set. After describing the data used in the analyses, the basic regressions and IV method used to analyse the data are introduced and results using each approach are compared. Finally, conclusions are drawn about the strength and implications of the work presented and areas for future research.

Treatment Effectiveness and Instrumental Variables

Strictly speaking, the treatment effect of a medical technology is the difference in an individual's outcomes resulting from receiving the treatment versus an alternative.[4,5] The true treatment effect is not obtainable since the outcomes from the treatment and the alternative cannot be observed for the same individual. Much of the information about treatment effectiveness in health care comes from randomized clinical trials (RCTs), the gold-standard biostatistical approach which identifies the average population treatment effect from the comparisons between the treatment and control group.[4] Randomization of patients to alternative treatments (or alternative quantities of treatments) has of course the enormous advantage of breaking a statistical dependence of treatment on any variable (observed or not) that also influences outcome. However, clinical trials are very costly, and for ethical and practical reasons, cannot address all research questions. Scientific considerations about the nature of illnesses to be included and the strict adherence to treatment protocols may dictate research designs with limited relevance to actual clinical practice, limiting the external validity of RCTs. The treatment research literature stresses this point by making a distinction between the evidence gained from RCTs, and that from actual practice settings. Randomized trials yield evidence about 'efficacy' (the productivity of treatment in laboratory conditions) as distinct from evidence about 'effectiveness' (the productivity of treatment as it is typically applied). Furthermore, RCTs do not completely eliminate the concern for 'selection bias', particularly in experiments with partial compliance.[6,7]

Increasingly, data from non-experimental settings are being used to study the effect of treatment. Many conventional treatment effectiveness studies using observational data follow a production function approach.[8,9] The central problem in interpreting evidence from real-world or non-experimental settings is the confounding of treatment with other variables ('selection bias'), particularly unobserved variables, that influence outcome. A simple pre/post design in which the condition of clients is compared both before and after treatment has one version of this problem. The 'post' observations have had treatment, and have had time pass in relation to the 'pre' observations. If the problem the treatment was designed to deal with tends to abate naturally in time, it is impossible to distinguish a treatment effect from a regression to the mean.[10] In general, a sample of sick patients tend to get better over time.[11,12]

In non-experimental settings, treatment can be confounded with many other variables, some much more difficult to measure than time in treatment. In the *Medical Outcomes Study* conducted by researchers at the RAND Corporation, very extensive measurement of clinical status was relied upon to control for patient condition pre and post treatment.[13] In the study of treatment for depression, however, in spite of including extensive controls in the model, Wells *et al*.[14] estimated a negative relationship between treatment and outcome. The authors interpreted this as an indication of a negative bias introduced by more treatment being confounded with unmeasured severity of the patient's condition. Patients who were more seriously ill (in unmeasured ways) were assigned more treatment, and the treatment variable serves as an unintended proxy for severity.[15]

It is important to note that the direction of bias in an estimated coefficient on a measure of treatment cannot always be anticipated. A correlation of quantity of treatment with unobserved severity will tend to bias the estimate downward, but other factors may also be at work. Patients who are more responsive to treatment and doing better may be more compliant. If these patient factors cannot all be measured and controlled for, the estimate of the effect of treatment will be biased upwards.[17,18] In any case, there is a presumption in non-experimental settings that treatment is correlated with unobserved variables affecting outcome.

There are two methods in econometrics to draw unbiased inferences in non-laboratory study settings. The first type develops a structural model and makes estimates using sophisticated econometrics such as the full information maximum likelihood (FIML) approach. The Heckman style selection model is one such example.[19,20] The drawback of such a method is that the necessary assumptions to estimate structural models may be too strong to make in some cases, especially for data sets in health research. The joint normality distribution hypothesis, for example, is an assumption that researchers are reluctant to make with binary or highly skewed dependent variables.

The second type of econometrics method is the instrumental variable (IV) approach, which is less structural and requires less stringent assumptions. The IV approach, long used in econometrics to deal with an endogeneity problem, was employed by Imbens and Angrist[4,21] to estimate average treatment effect. The essence of the method is to use one or more IVs, which are observable factors that influence treatment but do not directly affect patient outcomes, to isolate the effect of treatment variation that is independent of unobserved patient characteristics. Imbens and Angrist[4] proved that under certain mild conditions,[24] an IV estimator (their study used the special case of a two-stage least-squares estimator) can be used to estimate the so-called local average treatment effects, defined as the average treatment effect for the subpopulation whose treatment participation could be induced to a change by a change in the instrument.[25] Since this group of the population is the major concern of providers of medical treatment, the local average treatment effect can serve as important evidence of treatment effectiveness. The case for using an IV approach was made forcefully by McClellan and colleagues.[26–29] Applications of the IV approach to estimating treatment effects could be found in the work of McClellan, McNeil and Newhouse,[27] Blumberg and Binns,[16] Frank,[30] Stearns,[31] Sturm[18] and Lu and McGuire.[32] In an analysis similar to that conducted here, Frank[30] applied a two-step approach to obtain a less biased estimates of the relationship between use of mental health services and persistence of signs and symptoms of mental distress.

Data

The data were collected for two periods from a community-based sample living in low income areas of Puerto Rico. The first survey was conducted from October 1992 to May 1993. With a response rate of 90.6%, the wave 1 data consisted of 3507 adults. The follow-up survey a year later targeted all eligible wave 1 respondents and achieved a response rate of 92.1%. The wave 2 data consisted of 3263 adults. The available information in both data sets was respondents' demographics, employment, migration status, income, mental health status, health insurance, utilization of mental health services and providers' information.[33,34] There are altogether 3221 respondents who were interviewed in both wave 1 and wave 2.[35] This paper studies the effectiveness of mental health care for these 3221 individuals.

Measuring Utilization of Mental Health Care

Mental health care can be divided into formal care and informal care. Use of formal care refers to receiving help in any of the following settings: office of psychiatrist or psychologist; mental health clinics; physical health facilities and other agencies which employ social workers or counsellors. Informal care refers to consulting or talking to any of the following: a relative; a minister, priest or clergy; persons at a self-help group (e.g., Alcoholics Anonymous); psychic, healer, medium, naturalist, therapist, or astrologer. This study focuses only on evaluating the productivity of the formal care.[36]

Utilization of formal mental health care in the past one year reported in the wave 2 survey is considered as the 'current' utilization, while that reported in the wave 1 survey is considered as the 'previous' utilization. Utilization of treatment is measured by two variables: whether the patient received formal care or not, and quantity of care the patient received (number of visits). 'Current' and 'previous' utilization of formal mental health care are therefore measured using the following four variables:

(i) FORMAL1 and FORMAL2: received any kind of formal mental health care in the past one year or not (0/1), as reported in wave 1 ('previous') and wave 2 ('current');
(ii) LN(VISIT1) and LN(VISIT2): log units of formal care, which is defined as log of number of visits[37] to any type of formal mental health care within the past year, as reported in wave 1 ('previous') and wave 2 ('current').

Out of the 3221 individuals in the data set, 386 (12.0%) reported receiving some kind of formal mental health care in the past year when interviewed in wave 2 (compared with 13.1% reported in wave 1 interview), with an average number of visits of 11.7.

Measuring Outcome of Mental Health Care

Outcome is measured as the change in need for mental health services using wave 1 as a baseline. The need variable (NEED) in this data set has been created by aggregating information on the individual's mental health status from various survey questions. By using diagnostic and impairment criteria, this need variable, developed by the original investigators, is characterized by definite ('1'), possible ('2') or unlikely ('3').[39] The mental health outcome variable

(OUTCOME) is defined, as shown in **Figure 1**, by comparing the need for mental health care of the respondent in wave 1 and wave 2: 1 if need for mental health care decreases in wave 2 compared with wave 1, or if there is no need for mental health care in wave 2 regardless of need in wave 1; 0 otherwise. The number of individuals who fall into each cell are shown in **Figure 1**. Note that 'retaining no need for care' is considered as a desirable outcome ('1'), with the perspective that mental health care may prevent deterioration of mental health, and 'retaining possible need for care' is defined as a undesirable outcome ('0'), because it is not clear whether the individual's mental health status has improved or not. The effect of different definitions of the outcome variable on the estimation results will be tested in the sensitivity analysis.

Independent Variables

Various mental health status measures available in the data set are used to control for the baseline case mix. Besides the need variable, the Psychological Symptom and Dysfunction Scales (PSDS) measures are also used.[41] These include five symptom scales: anxiety (0–48), depression (0–60), psychosocial dysfunction (0–44), cognitive impairment (0–16) and general psychopathology (0–32). Each symptom scale is derived from adding the respondent's answers to each question within that symptom category. Those persons scoring higher than a certain cutoff point are regarded as having severe symptoms. The cutoff point or threshold, established through validity studies of clinical and community cases, is 19 for a high anxiety symptom.[42] Similarly, a score of 24 is for depression, 5 for psychosocial dysfunction, 5 for cognitive impairment and 8 for general psychopathology. Accordingly, five dummy variables are included to control for the baseline severity of mental illness. The Center for Epidemiological Studies—Depression Scale (CES-D) is another criterion for depression measure. In this study a dummy variable is included, which is defined as 1 if the CES-D scale is high. In addition, two self-reported health status measures are used: the first is self-reported physical health status, which scales from 0 to 5 ('5' means the respondent regards himself as in very good physical health); the second is the self-reported mental health status, which scales from 1 to 5 ('5' means the respondent regards himself as in very good mental health).

Besides the case mix variables and treatment variables, the following demographic variables are included: age, sex, marital status, education and family income in all regressions

		Wave 2 (NEED2)		
		"Definite"	"Probably"	"Unlikely"
Wave 1 (NEED1)	"Definite"	215	73	116
	"Probably"	89	91	208
	"Unlikely"	117	218	2094

Figure 1. Definition of outcome (shaded area: outcome = 1)

as independent variables. Table 1 presents a full list of the definitions, means and standard deviations of the independent variables.

Data Analytic Procedures

This section presents the methods of empirical estimation of treatment effectiveness in this data set, drawing from household production theory.[43,44] The basic production functions are in the form of (1) and (2), estimated initially without controlling for the endogeneity of the treatment variables:

$$\text{OUTCOME} = a_{11} + \alpha_{11}H_0 + \beta_{11}\text{FORMAL2} + \gamma_{11}X + \eta_{11} \quad (1)$$

$$\text{OUTCOME} = a_{12} + \alpha_{12}H_0 + \beta_{12}\ln(\text{VISIT2}) + \gamma_{12}X + \eta_{12} \quad (2)$$

Model (1) is estimated using the whole sample (sample size 3221), using FORMAL2 to measure treatment. Considering the fact that only 12.0% of the sample reported any use of formal mental health care and that the units of treatment are highly skewed, model (2) is estimated using the user subset of the sample (sample size 386).[45] In (2), log number of visits is used to measure treatment. The first regression on (1) shows whether receiving mental health care has a positive impact on mental health status improvement, while the second regression estimates the effectiveness of the 'current' mental health care for the treatment sub-population.

Change in mental health status may also partly be a result of previous utilization of care. Previous utilization of care in this study is measured using treatment in the past year at wave 1 interview. Part of the treatment effect of previous utilization may show up between the period of the wave 1 interview and the wave 2 interview, and may therefore influence the treatment outcome. Receiving mental health care in the past may also be an indicator of severity of illness and/or the individual's motivation to deal with the problem. The correlation of previous use and the current use is thus clouded by the possible correlation with unobserved variables. As a result, the basic models are modified with variables measuring previous utilization also controlled for in the production functions.

IV Estimation and Consistency of IV Estimators

Since one of the treatment variables (FORMAL2) is binary, it calls into question the consistency of conventional two-stage-type IV estimation. In this study, the Dubin–McFadden[55] method is used for the IV estimation, which is characterized by a Logit first stage of regressing FORMAL2 on X, FORMAL1 and the IVs; an ordinary least-squares second stage of regressing FORMAL2 on X, FORMAL1, the IVs and the predicted treatment from first stage and a Logit third stage of regressing OUTCOME on X, FORMAL1 and the predicted treatment term from second stage. The

Table 1. Independent variables*

Variable definition	Variable name	Sample size	Mean	Standard deviation
Client characteristics				
Age	AGEMAG1	3221	39.80	13.13
Female	FEMALE	3221	0.60	0.49
Married	MARRIED	3221	0.53	0.50
Years of education	YRSEDUC1	3221	10.48	4.18
Family annual income 92–93	FMINCOM1	3221	13 223.17	13 552.13
Case mix measures				
'NEED' of mental health care	NEED1	3221	2.63	0.70
1—'definite'			12.50%	
2—'possible'			12.00%	
3—'unlikely'			75.40%	
PSDS scales				
Anxiety Scale (1 if symptom scale is >19)	ANXIETY1	3221	0.16	0.37
Psychosocial Dysfunction Scale (1 if symptom scale is >5)	PDYSFUN1	3221	0.24	0.43
General Psychopathology Scale (1 if symptom scale is >8)	GPSYCHO1	3221	0.40	0.49
Cognitive Impairment Scale (1 if symptom scale is >5)	COGNITI1	3221	0.32	0.47
Depression Scale (1 if symptom scale is >24)	DEPRESS1	3221	0.28	0.45
CES-D scale (1 if symptom scale is >16)	CESDD1	3220	0.21	0.41
Self-reported physical health status (0,1, …,5 5 = good)	PHYSHLT1	3221	2.86	1.14
Self-reported mental health status (1, …,5 5 = good)	MHSTATU1	3217	4.11	0.78
Instrumental variables				
Travel distance	DISPSM	3221	8.77	5.14
Travel distance × prob(owning car)	CDISPSM	3221	5.89	3.43
Insurance				
Blue Cross	BLUECROS	3221	0.09	0.28
Medicare	MEDICARE	3221	0.21	0.41
Triple S (a private insurance)	TRIPLES	3221	0.08	0.27
Medicaid	MEDICAID	3221	0.30	0.46
Other insurance	OTHINSUR	3221	0.12	0.32

**Note:* All variables in this table are created using baseline information from the wave 1 survey.

IV estimation on ln(VISIT2) follows a more conventional two-stage IV method which is characterized by a first stage ordinary least squares and a Logit second stage. The consistency of this estimator has been proved in the literature under fairly general conditions.[56]

Instruments for Mental Health Care Utilization

Choice of instruments is crucial in applying IV estimation. A valid instrument must satisfy the following two main requirements: first, the instrument must influence the choice/quantify of mental health care, i.e. with strong enough predictive power; second, the instrument must not be a direct determinant of the outcome (improvement of mental health status). In this study, two types of instrument are employed (definitions, means and standard deviations for the instrumental variables explored in this study are displayed in **Table 1**).

Travel Distance

Distance has been found to affect the decision of seeking treatment and the choice of providers in a number of previous studies of health care utilization.[27,46–48] A series of distance variables have been created to measure the distance from the respondent's residence to the nearest mental health provider by Hodgkin, Alegria and McGuire[49] using the same data set. First, they collected information on the location of mental health providers, using public directories to locate public mental health centers and the telephone company's yellow pages to locate office-based practitioners. Of the island's 78 municipalities, 68 have either a psychiatrist or a mental health center. Hodgkin, Alegria and McGuire[49] assumed that a provider is located at the centroid of its municipality.[50] The distance to the nearest mental health clinic is thus measured as the straight line distance from the centroid of the respondent's census tract to the centroid of the clinic's municipality. Similar variables were created for the distance to the nearest psychiatrist, and to the nearest formal specialty provider, which is defined as the minimum of distance to clinic and distance to psychiatrist.[51]

Travel distance used in this study is the distance to the nearest town with either a psychiatrist or a mental health centre. Both travel distance and an interacted term of travel distance with probability of owning a car (measured at the census tract level) are included as instrumental variables.

Insurance

A large literature in health economics documents the response of providers to incentives in the payment system.[52] Edlund, Wheeler and D'Aunno[53] supported the view that the providers are expected to supply more services to clients bringing in more marginal revenue (i.e., by retaining the client in the treatment longer). Insurance also alters demand-side incentives and thus leads to a different level of utilization.[54] However, this may not apply to the case of mental health care in Puerto Rico given the availability of a large non-insurance-based public mental health system. I will first include dummy variables for five types of insurance in Puerto Rico: Blue Cross, Medicare, Triple S (a private insurance company in Puerto Rico), Medicaid and others. Additionally, a dummy variable showing whether the individual has insurance for mental health is also used as an instrumental variable in some regressions. These IVs will be excluded in some sensitivity analysis to test whether the results are robust.

By definition, instrumental variables should not have a direct impact on the outcome variable. This restriction, also called 'overidentifying restriction', is tested by including each subset of instruments in the last stage regressions to see whether these instruments can properly be excluded from these regressions.

Efficiency of IV Estimators and Standard Errors Correction

The final concern about the IV method is the efficiency of the IV estimator as well as the standard errors produced from three-stage or two-stage methods. It is well known that, in general, an IV estimator is not the most efficient estimator.[56,57] The usual suggested alternative is to estimate the first and second step models using some joint method such as full information maximum likelihood (FIML).[58,59] The FIML estimators, with certain appropriate assumptions, are efficient estimators and yield asymptotically correct estimates of standard errors. However, in the data set used for this study, only 12% of the respondents reported any use of formal mental health care (FORMAL2). As a result, the joint normality assumption of the error terms in the two- or three-stage models involving FORMAL2 is not likely to hold. However, effort has been made to improve the efficiency of IV estimators for the treatment variable ln(VISIT2), which is a continuous variable. Instead of adopting an FIML method, a second-stage probit model is used which includes the residual term from first stage as an independent variable. This IV estimator is more efficient than the standard two-stage IV estimation.[60]

In both IV estimations method used here, the standard errors in the final stage estimation are not correct because in both methods the predicted values are treated as if they are known for purposes of estimation. In this study, the standard errors are corrected in all three-stage or two-stage IV models using the formula provided by Murphy and Topel.[59]

Results

Conventional Models Regression Results

Estimation results on the basic model (1) are presented in the first column of **Table 2** (all the analyses in this study were done using SAS version 6.09). Current treatment, measured by whether the respondent received formal care or not in the previous year, is found to significantly decrease the probability of improving in mental health status. When measuring current treatment using log units of formal treatment in the previous year and estimating basic model on a much smaller sample ($N = 386$), regression results again showed a negative effect (although not significant) of treatment on mental health outcome, presented in the first column of **Table 3**.

An immediate modification of the above basic model is to control for the impact of previous treatment by including the previous treatment variable. Regression results from the revised basic models are reported in the second column of **Table 2** and **Table 3**, respectively. Again, a significant and negative treatment effect was found when current treatment is measured using FORMAL2; a negative but non-significant treatment effect was found when current treatment is measured using ln(VISIT2). No significant effect of previous treatment was found in either production function.

Specification Tests on Treatment Variables

In order to determine whether IV estimations are unnecessary, a Hausman test is conducted to see whether FORMAL2 is not correlated with the unobservable factors.[54] The Hausman chi-square test result is 11.75 (at a confidence level of 99%, the chi-square with one degree of freedom is 6.63). According to this test statistic, one cannot assume that the FORMAL2 variable and unobserved determinants of the outcome are uncorrelated. This indicates that conventional estimates of treatment effect, when treatment is measured using FORMAL2, are misleading. To obtain more accurate results, selection bias must be corrected for. However, a similar test on ln(VISIT2) does not show strong evidence of correlation of this treatment quantity variable with the unobserved determinants among users.

Validity of Instruments: IV Equations and Overidentifying Restriction Tests

The next step is to check whether the instruments satisfy the two main requirements on predictive power and overidentifying restrictions.

Table 2. Mental health care production function (1)
(i) using the full sample (users and non-users) and
(ii) measuring treatment as 'receiving treatment or not'

Independent variables	Probability of improved mental health in wave 2		
	Basic Logit model (1) selection bias not controlled	Basic Logit model (II) selection bias not controlled controlling for previous TX	IV Logit model selection bias controlled
Intercept	–0.39	–0.34	–1.64***
	(0.85)	(0.72)	(2.72)
Current treatment	(–1.55)***	(–1.50)***	
(FORMAL2)	(11.60)	(10.58)	
Instrumented current treatment			2.08*
(FORMAL2_E)			(1.92)
Previous treatment		–0.16	–1.35***
(FORMAL1)		(1.09)	(3.52)
Age	0.01	0.02	–0.02
	(0.35)	(0.38)	(0.58)
Female	0.38***	0.37***	0.33***
	(3.68)	(3.65)	(3.30)
Married	0.35***	0.35***	0.39***
	(3.53)	(3.48)	(3.92)
Years of education	0.05***	0.05***	0.04***
	(3.45)	(3.49)	(2.71)
Family annual income 92–93	–0.00	–0.00	–0.00
	(0.21)	(0.20)	(0.51)
'NEED' of mental health care	0.18**	0.17*	0.41***
	(1.94)	(1.82)	(3.45)
Anxiety Scale	–0.27**	–0.27*	–0.53***
	(1.84)	(1.80)	(3.13)
Psychosocial Dysfunction Scale	0.00	0.02	–0.03
	(0.00)	(0.15)	(0.19)
General Psychopathology Scale	–0.43***	–0.43***	–0.38***
	(3.55)	(3.55)	(3.22)
Cognitive Impairment Scale	–0.19*	–0.19*	–0.25**
	(1.62)	(1.62)	(2.09)
Depression Scale	–0.40***	–0.40***	–0.32**
	(2.76)	(2.77)	(2.28)
CES-D scale	–0.28**	–0.28**	–0.17
	(1.85)	(1.90)	(1.15)
Self-reported physical health status	0.11**	0.11**	0.16***
	(1.97)	(1.98)	(2.89)
Self-reported mental health status	0.19***	0.18***	0.31***
	(2.59)	(2.49)	(3.78)
–2 log *L*	2748.14	2746.97	2855.38

Notes: 1. Absolute values of the *t*-statistics are reported in parentheses below the parameter estimates.
2. Significant levels: *denotes .10, ** = .06, *** = .01.

First, the parameter estimation results on the instruments from the first- and second-stage IV equations, reported in **Table 4**, are examined to see whether the instruments appear to be strongly correlated with the treatment variables. The first two columns are estimation results from the first two stages of Dubin–McFadden methods, with FORMAL2 as the dependent variable. Travel distance and the interaction term of travel distance with households owning a car are significant in the first stage. Those who lived closer to mental health providers are found to be more likely to receive treatment. This could be interpreted as an impact of the time cost in travelling. However, time cost for those with cars is significantly less regardless of the travel distance. This is confirmed in the first-stage results in which the interaction term of travel distance and car ownership is found to have a significant positive impact on receiving treatment. When ln(VISIT2) is the treatment variable to be instrumented, the ordinary least-squares regression results of the first stage are as shown in the third column of **Table 4**. The interaction term of travel distance with owning a car, and the Blue Cross insurance dummy, were found to be significant. Insurance dummies are not significant in most of the instrumental variable equations, which is not surprising given the availability of a large non-insurance-based public mental health system. To further test the predictive power of the instruments, a formal partial *F* test (or LR test) is

Table 3. Mental health care production function (2)
(i) using information from users only (sample size: 386) and
(ii) measuring treatment as log (unit of treatment)

Independent variables	Probability of improved mental health in wave 2			
	Basic Logit model (1) selection bias not controlled	Basic Logit model (II) selection bias not controlled controlling for previous TX	IV Logit model selection bias controlled	IV Probit model selection bias model Newey (1987) model
Intercept	2.21	2.07	1.39	0.83
	(1.43)	(1.33)	(0.80)	(0.80)
Current treatment	–0.10	–0.14		
(ln(VISIT2))	(0.68)	(0.86)		
Instrumented current treatment			0.44	0.28
(ln(VISIT2)_E)			(0.63)	(0.66)
Previous treatment		0.10	–0.10	–0.06
(ln(VISIT1))		(0.64)	(0.34)	(0.37)
Age	–0.15	–0.16	–0.12	–0.08
	(0.95)	(1.03)	(0.77)	(0.80)
Female	–0.20	–0.17	–0.08	–0.04
	(0.57)	(0.47)	(0.22)	(0.17)
Married	0.67*	0.68*	0.85**	0.52**
	(1.80)	(1.80)	(1.96)	(2.07)
Years of education	–0.03	–0.03	–0.03	–0.02
	(0.60)	(0.58)	(0.72)	(0.69)
Family annual income 92–93	–0.00	–0.00	–0.00	–0.00
	(0.39)	(0.34)	(0.53)	(0.57)
'NEED' of mental health care	–0.48	–0.46	–0.57*	–0.36*
	(1.59)	(1.50)	(1.71)	(1.79)
Anxiety Scale	–0.09	–0.12	–0.07	–0.07
	(0.18)	(0.24)	(0.13)	(0.22)
Psychosocial Dysfunction Scale	0.06	–0.02	–0.03	–0.01
	(0.11)	(0.04)	(0.05)	(0.04)
General Psychopathology Scale	–0.94*	–0.91*	–1.08**	–0.66**
	(1.79)	(1.73)	(1.91)	(1.96)
Cognitive Impairment Scale	–0.86**	–0.83**	–0.89**	–0.53**
	(1.83)	(1.75)	(1.85)	(1.85)
Depression Scale	–0.64	–0.61	–0.52	–0.32
	(1.24)	(1.18)	(0.99)	(1.01)
CES-D scale	–0.15	–0.14	–0.28	–0.16
	(0.35)	(0.33)	(0.59)	(0.58)
Self-reported physical health status	0.36	0.36	0.46*	0.27*
	(1.53)	(1.52)	(1.74)	(1.74)
Self-reported mental health status	–0.14	–0.14	–0.19	–0.11
	(0.53)	(0.55)	(0.72)	(0.70)
Residual term from IV equation	n/a	n/a	n/a	(–0.10)
	n/a	n/a	n/a	(1.07)
–2 log *L*	225.93	225.52	225.86	224.47

Notes: 1. Absolute values of the *t*-statistics are reported in parentheses below the parameter estimates.
2. Significant levels: * denotes .10, ** = .05, *** = .01.

conducted to test whether the estimates of the instrumental variable equation are significantly different when the IVs are excluded.[57] From the test statistics reported in **Table 4**, it could be seen that the instruments are shown to be strong predictors of units of visit yet not strong predictors of receiving formal care.

Second, the instrumental variable estimation was repeated by sequentially including a subset of the instruments in the last stage production function estimation to see whether these IVs have significant direct effects on the outcome. All IVs are shown to be excludable from the last stage equations, suggesting there is no evidence for direct correlation between the IVs and treatment outcome.

Treatment Effects Estimated from IV Models

The final step is to estimate the treatment effects using the IV approach. The third column of Table 2 presents the parameter estimation results from the IV model. Interestingly, the treatment variable, measured by a predicted term from

Table 4. Instrumental variables equation

Instrumental variables	Dependent variable: FORMAL2 $N = 3221$ D-M first stage (LOGIT)	FORMAL2 D-M second stage (OLS)	FORMAL2 Standard first stage(OLS)	ln(VISIT2) $N = 386$ Standard first stage (OLS)
Travel distance	−0.14**	−0.00	−0.01*	−0.12
	(1.93)	(0.64)	(−1.76)	(1.14)
Travel distance × prob(owning car)	0.22**	0.00	0.01*	0.19*
	(2.01)	(0.65)	(1.85)	(1.18)
Insurance				
Blue Cross	0.25	0.00	0.02	−0.79*
	(0.88)	(0.44)	(0.80)	(1.29)
Medicare	(−0.04)	(0.00)	(0.00)	(−0.09)
	(0.20)	(0.03)	(0.02)	(0.60)
Triple S (a private insurance)	−0.10	−0.00	−0.00	0.49
	(0.32)	(0.22)	(−0.21)	(1.09)
Medicaid	(0.27)	(0.01)	(0.03)**	(−0.14)
	(1.48)	(0.66)	(1.89)	(0.81)
Other insurance	0.29	0.01	0.02	0.20
	(1.19)	(0.54)	(1.25)	(0.79)
Chi-square/F test	8.76	1.84	1.31	2.63
Null: coefficients on all IVs = 0				
Reject null?	No	No	No	YES
−2 log L (Logit)	1699.35			
Adj R-sq (OLS)		0.28	0.27	0.13

Notes: 1. Only regression results on the instrumental variables are reported.
2. Absolute values of the t-statistics are reported in parentheses below the parameter estimates.
3. Chi-square test results on the excluded IVs for D-M first stage is reported. Chi(7) = 12.02 for 0.10 significant level.
4. Partial F-test results on the excluded IVs are reported. F(7, infinity) = 2.05 for .05 significant level.
5. Significant levels: * denotes .10, ** = .05, *** = .01.

the instrumental variable equation (FORMAL2_E), appear to significantly increase the probability of improvement in mental health status.[62] Estimation results on other independent variables do not change a lot compared with results from basic models. Previous treatment (FORMAL1) is shown to significantly decrease the probability of improvement in mental health status.[63]

The IV estimation results of a standard two-stage IV model using users only are reported in the third column of **Table 3**. Current treatment, measured by the instrumented term ln(VISIT2)_E, is also found to increase the probability of improvement in mental health status yet the effect is not significant. Consistent with what was observed on the overall sample, previous treatment was found to have a negative (but not significant) impact on mental health outcome. Further IV estimation results using a Probit IV model, which provides a more efficient estimator, are reported in the fourth column of **Table 3**. Results from this model remain consistent: positive (not significant) effect of current treatment and negative (not significant) effect of previous treatment was found. Standard errors in all regressions have been corrected using the Murphy–Topel[59] formula.

In order to compare the treatment effects estimated from all the models, average marginal effects are calculated receiving estimation results and sample means in each model and presented in **Table 5**. According to the basic models, receiving formal mental health care significantly reduces the probability of improvement in mental health status by 26.8% to 28.0%, and a one unit increase in formal treatment at the margin reduces the chance of achieving an improved mental health status by 1.8% to 2.4%. On the other hand, according to the results from the IV models, receiving formal mental health care significantly increases the probability of obtaining a better mental health outcome by 19.2%, and one unit increase in formal treatment increased the probability of getting healthier by 6.2% to 8.4%.

Consistent with what has been found in mental health literature, the regression results also suggest significant effects of various factors in the mental health production function. Specifically, an individual's baseline mental health status is found to be important: lower initial 'need' of mental health care, anxiety level, general psychopathology scale, cognitive impairment scale and depression scale and higher self-reported physical health status are found to be related to a higher probability of improving in terms of mental health status in wave 2. Individuals with previous treatment history are less likely to improve. Among demographic factors included in the production function, being female, married, and high education were found to contribute to a higher probability of improvement in mental health status.

Table 5. Marginal treatment effects of mental health care

Estimation models	Treatment variable	
	FORMAL2	ln(VISIT2)
Basic model	×	×
selection bias not controlled	–0.28***	–0.02
	(11.60)	(0.68)
Basic model	×	×
selection bias not controlled	–0.27***	–0.02
controlling for previous treatment	(10.58)	(0.86)
IV Logit model	+	+
selection bias controlled	0.19*	0.08
	(1.36)	(0.79)
IV Probit model	n/a	+
Newey (1987) model		0.06
selection bias controlled		(0.65)

Notes: 1. Average marginal effects are calculated and reported in this table. 2. Absolute values of the corrected *t*-statistics are reported in parentheses. 3. An '×' indicates a NEGATIVE treatment effect, regardless of significant level; a '+' indicates a POSITIVE treatment effect, regardless of significant level.
4. Significant levels: * denotes .10, ** = .05, *** = .01.

Sensitivity Analysis

Many authors have pointed out that caution must be taken when applying the instrumental variable approach to estimate treatment effects. Employing inadequate instruments could introduce another sort of bias in the estimation. Sensitivity analysis should be conducted to test how credible the results are to different statistical models and assumptions.[3,13]

Sensitivity Analysis on Model Specifications

An extensive range of different model specifications are tested in this study. First, different definitions of the outcome variables are used.[64] Second, the sample is divided into three subgroups—most in need, probably in need and no need group—according to baseline need for mental health care and the production function is estimated separately. Furthermore, an alternative model specification which is to include the interaction terms of treatment and need level in the model is also tested. Overall, this set of sensitivity analysis partly supports the main results presented earlier in this section. It should be noted that the IV approach works well on reducing the selection bias for the most in need group, which is the group of most policy and treatment concern. According to the IV estimation results, treatment is productive for the group in most need of mental health care (see **Table 6**).[65] However, estimations do not find strong enough evidence to demonstrate treatment effects on other groups with less or no need.

Sensitivity Analysis on the Choice of Instruments

Since insurance dummies do not perform well enough in some IV equations, tests are done to show how the estimation results change when only distance instruments are included. Furthermore, there is the question of whether the correlation of travel distant instruments with treatment choice is considerately different for urban residents and rural residents. In order to test if the effect is different, the sample is segregated into URBAN and RURAL groups and the estimations run separately. There are two important conclusions. First, after the insurance instruments are excluded, the IV estimation results on the overall group are consistent with the results reported in the text. Second, the distance instruments do not have different correlation with treatment variables for URBAN and RURAL groups. IV estimations on the URBAN and RURAL subgroups do not show any evidence of significant treatment effects. However, this could be a result of small sample size rather than poor instruments.

Conclusions

This paper addresses the issue of estimating treatment effects using an observational data set. The main concern of this paper is the puzzling significant negative effect of treatment found in many previous studies. The information in a mental health data set obtained from two waves of data in Puerto Rico is explored. The results using conventional models—in which the potential selection bias is not controlled—and that from IV models—which is what was proposed in this study to correct for the contaminated estimation from conventional models—are compared.

It is very noticeable that in this study, treatment effects were found to be negative in all conventional models (in some cases, highly significant). However, after the IV method was applied, the estimated average marginal effect of treatment became positive. Sensitivity analysis partly supports this conclusion. The IV approach is shown to be an appropriate method to reduce the selection bias for the group in most need for mental health care, which is also the group of most policy and treatment concern. The results in this paper are supportive of the concerns on the credibility of evaluation results using observation data set when the endogeneity of the treatment variable is not experimentally controlled. Unobserved factors contribute to the downward bias in the basic models, as in the Wells *et al*.[14] study discussed earlier. This study also provides a possible solution for dealing with such endogeneity, which is to explore instruments from the data and break the correlation of the treatment variable with the missing variables in the estimation model by applying the IV approach.

It should be recalled that the IVs explored in this study did not perform well enough in the IV equations and the predictive power remains questionable, which is a main limitation of this study. The most challenging part of applying the IV approach remains on finding 'good' instruments that demonstrate high correlation with the treatment variable yet not directly correlated with treatment outcome. More data and longer panel are needed in future research to provide more valid evidence.

The results of this work have implications for resource allocation in mental health care. Evidence is found that mental health care provided in Puerto Rico is productive, and is most helpful for persons in most need for mental health care. According to what estimated from the IV

Table 6. Sensitivity analysis results: marginal treatment effect for each subgroup

	Treatment variable					
	FORMAL2			ln(VISIT2)		
Estimation models	Most in need group	May in need group	Not in need group	Most in need group	May in need group	Not in need group
Basic model	×	×	×	+	×	×
selection bias not controlled	–0.28***	–0.39***	–0.30***	0.07*	–0.20	–0.12**
	(5.39)	(5.23)	(9.57)	(1.69)	(0.75)	(2.24)
Basic model	×	×	×	+	n/a	×
selection bias not controlled	–0.29***	–0.39***	–0.27***	0.06		–0.14**
controlling for previous	(5.06)	(5.18)	(8.29)	(1.38)		(2.28)
treatment						
IV Logit model selection bias	+	×	×	+	n/a	×
controlled	0.11*	–0.36	–0.42*	0.16		–0.19
	(0.30)	(–0.98)	(1.67)	(1.03)		(1.09)
IV Probit model	n/a	n/a	n/a	+	n/a	×
Newey (1987) model				0.17		–0.15
selection bias controlled				(0.99)		(0.19)

Notes: 1. Average marginal effect is reported in this table.
2. Absolute values of the corrected *t*-statistics are reported in parentheses.
3. An '×' indicates a NEGATIVE impact of treatment, regardless of significant level; a '+' indicates a POSITIVE impact of treatment, regardless of significant level.
4. Significant levels: * denotes .10, ** = .05, *** = .01.

models, on the margin, receiving formal mental health care significantly increases the probability of getting a better mental health outcome by 19.2%, and one unit increase in formal treatment increased the probability of getting healthier by 6.2% to 8.4%. Consistent with other mental health literature, an individual's baseline mental health status is found to be significantly related with the probability of improvement in mental health status: individuals with previous treatment history is less likely to improve. Among demographic factors included in the production function, being female, married, and high education were found to contribute to a higher probability of improvement.

Acknowledgements

I am indebted to Thomas G. McGuire, Kevin Lang and Randall P. Ellis for guidance with this research. Thanks are due to the Center for Evaluation and Sociomedical Research at University of Puerto Rico for providing the data from the project *Mental Health Care Utilization Among Puerto Ricans* study (Margarita Alegria, PI; grant R01 MH42655 from the Division of Epidemiology and Services Research, National Institute of Mental Health). I am also grateful to Margarita Alegria, Chris M. Auld, Jose M. Calderon, Cam Donaldson, Susan Ettner, Dominic Hodgkin, Jospeph P. Newhouse, and two anonymous reviewers for their valuable comments and help.

References

1. Frank RG, McGuire TG. Introduction to the economics of mental health payment systems. In: Levin BL, Petrila J, eds. *Mental health services: a public health perspective*. Oxford: Oxford University Press, 1996, 23–37.
2. Frank RG, McGuire TG, Regier DA, Manderscheid R, Woodward A. Paying for mental health and substance abuse care. *Health Affairs* 1994; Spring (I): 337–342.
3. Evers SMAA, Van Wijk AS, Ament AJHA. Economic evaluation of mental health care interventions: a review. *Health Econ.* 1997; **6**: 161–177.
4. Imbens G, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62** (2): 467–476.
5. McClellan M. Uncertainty, health-care technologies, and health-care choices. *AEA Papers Proc.* 1995; **85** (2): 38–44.
6. Manski CF. Learning about treatment effects from experiments with random assignment of treatments. *J Human Resources* 1996; **31** (4); 709–733.
7. The April 1995 supplement to the journal *Medical Care* contains proceedings of the *Conference on Measuring the Effects of Medical Treatment*. Articles deal with issues of experimental and non-experimental design, and contain important examples of both types of research. Manski[6] explored the inferences that may be drawn from RCTs under assumptions weak enough to yield credible results, and examined various aspects of the potential problems.
8. Howard KI, Kopta SM, Krause MS, Orlinsky DE. The dose–effect relationship in psychotherapy. *Am. Psychol.* 1986; **41** (2): 159–164.
9. Knesper DJ, Belcher BE, Cross JG. Preliminary production functions describing change in mental health status. *Med. Care* 1987; **25** (3): 222–237.
10. Such a study is reported in the June 1995 newsletter of the Foundation for Health Services Research, *Connection*. The California Drug and Alcohol Treatment Assessment study (CALDATA) was based on a survey of 1900 participants in treatment programs. At the start of treatment, clients used more hospital care, committed more crimes, and used more drugs than at the completion of treatment.
11. Welch WP. Regression toward the mean in medical care costs: implications for biased selection in health maintenance organizations. *Med. Care* 1985; **23** (11): 1234–1241.
12. Sturm R, Jackson CA, Merredith LS, Yip W, Manning WG, Rogers WH, Wells KB. Mental health care utilization in prepaid and fee-for service plans among depressed patients in the medical outcomes study. *Health Services Res.* 1995; **3** (2): 319–340.
13. Wells KB, Sturm R, Sherbourne CD, Meredith LS. *Caring for Depression*. Cambridge, Massachusetts: Harvard University Press, 1996.
14. Wells KB, Burnam MA, Rogers W, Hays R, Camp P. The course of

depression in adult outpatients: results from the medical outcomes study. *Arch. Gen. Psychiatry* 1992; **49**: 788–794.
15. The problem of obtaining a result of negative treatment effectiveness is not uniquely encountered in mental health care evaluation. The same puzzle could be found in Blumberg and Binns.[16] Their study of the use of the Swan–Ganz catheter (SGC) in treatment of patients with acute myocardial infarction (AMI) illustrates concern with confounding with time as well as unmeasured severity. Patients who stay longer in a hospital (for whatever reason) have a greater chance of dying in the hospital, so Blumberg and Binns, in a commonly used tactic for dealing with time-related effects, compare 30-day mortality rates for patients exposed or not exposed to the SGC procedure. In spite of many studies calling attention to the correlation of use of SGC and mortality, cardiologists questioned the causality between SGC and mortality because the patients for whom this procedure is prescribed are usually 'sicker' than the typical AMI case, and unless this elevated sickness is controlled for in a study, the observed correlation could be spurious. Blumberg and Binns[16] attempt to circumvent this problem by using an instrument for SGC use. They do not compare mortality for patients with and without the procedure, but compare mortality for patients with a higher or lower risk of being given the procedure, with the hospital's proclivity to administer SGC serving as the instrumental variable. (However, they continue to find an elevated mortality with SGC.)
16. Blumberg MS, Binns GS. Swan–Ganz catheter use and mortality of myocardial infarction patients. *Health Care Financing Rev.* 1994; **15** (4): 91–103.
17. Rosenheck R. Effectiveness and cost of specific treatment elements in a program for homeless mentally ill veterans. *Psychiatric Services* 1995; **46** (11): 1131–1139.
18. Sturm R. Instrumental variable methods for effectiveness research. RAND Working Paper No. 102.
19. Heckman JJ. Varieties of selection bias. *Am. Econ. Rev.* 1990; **80**: 313–318.
20. Heckman JJ, Hotz VJ. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc.* 1989; **84**: 862–874.
21. It has also been shown that IV technique can be fitted into the Rubin causal model and be used for causal inference without assuming constant treatment effects.[23,24]
22. Angrist J, Imbens G, Rubin D. Identification of Casual Effects Using Instrumental Variables. *National Bureau of Economic Research Technical Paper* 1993; No. 136.
23. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 1995; **90** (430): 431–442.
24. The conditions specified by Imbens and Angrist[4] are (i) the instruments are nontrivialy correlated with the treatment variable without correlated with the unobservable factors that affect treatment outcomes (i.e., existence of instruments) and (ii) the instruments affect the treatment variable in a monotonic way. These conditions are mild in the sense that they are satisfied in a wide range of models and circumstances.
25. Their study proved that the TSLS estimator captures a *weighted* average of causal responses to a unit change in treatment, for those whose treatment status is affected by the instrument.
26. McClellan M. Uncertainty, health-care technologies, and health-care choices. *AEA Papers Proc.* 1995; **85** (2): 38–44.
27. McClellan M, McNeil B, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? An instrumental variables analysis. *J. Am. Med. Assoc.* 1994; **272**: 859–866.
28. McClellan M, Newhouse JP. The Marginal Benefits of Medical Technology. *National Bureau of Economic Research Working Paper* 1995.
29. McClellan M, Newhouse JP. The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach. *J. Econometr* 1997; **77** (1): 39–64.
30. Frank RG. Use of mental health services and persistence of emotional distress: an exploratory analysis. *Med. Care* 1988; **26** (12): 1203–1215.
31. Stearns SC. Hospital discharge decisions, health outcomes, and use of unobserved information on case-mix severity. *Health Services Res.* 1991; **26** (1): 27–51.
32. Lu M, McGuire TG. *The Productivity of Outpatient Treatment in Substance Abuse*. Unpublished manuscript, 1999.
33. Alegria M, Pescosolido BA, Santos D, Vera M. Can we conceptualize and measure continuity of care in individual episodes? The case of mental health services in Puerto Rico, *Sociol. Focus* 1997; **30** (20): 113–129.
34. Gibbons R, Alegria M. *Patterns of Mental Health Service Utilization: a Multivariate Probit Analysis*. Unpublished manuscript, 1997.
35. In the follow-up survey (wave 2), efforts were made to interview those individuals who were qualified for the survey but were not interviewed in wave 1. This explains why there are 42 observations in wave 2 but not found in wave 1.
36. Receiving informal care might also contribute to the selection bias of treatment effectiveness of formal care. For example, support from a self-help group may increase compliance. However, in this study, less than 5% of the sample (154 out of 3221) received both formal and informal care. The dummy variable of receiving informal care or not is included in some regressions in the sensitivity analysis, and the regression results do not change much. This suggests the degree of correlation of receiving informal care and the unobservable factors that affects both formal care and treatment outcome is low.
37. The log transformation is used to eliminate the skewness in the distribution of treatment units among users.[38]
38. Manning WG, Newhouse JP, Duan N, Keeler EB, Leibowitz A, Marquis MS. Health insurance and the demand for medical care: evidence from a randomized experiment. *Am. Econ. Rev.* 1987, **77**: 251–277.
39. This need indicator is aggregated using measures of disorders, distress and impairment according to several epidemiologic criteria. To be identified with 'definite' need, a respondent had to endorse positively two or more impairment questions and fulfill criteria for at least a diagnosis of major depression, dysthymia, alcohol abuse, alcohol dependence, or antisocial personality, according to the Composite International Diagnostic Interview and the NIMH Diagnostic Interview Schedule; or score high on five Psychiatric Symptom and Dysfunction Scales (PSDS) sub-scales; or score higher than 22 on the Center for Epidemiological Studies Depression Scale (CES-D); or have used hard core drugs in the previous 30 days. A correspondent with fewer than two positive responses to the impairment questions but who complied with the above indicators, or with two or more positive responses to the impairment questions and who scored high on three or four PSDS sub-scales or scored 16–22 on the CES-D scale is classified with 'probable' need. All others are classified with 'unlikely' need (for details, see [34]).
40. Outcome researchers in the mental health field recognize that outcome is multi-dimensional. The outcome variable defined in this study is, of course, a simplified measure of the effectiveness of mental health care.
41. Note that the 'need' variable is also used to construct the outcome variable. A similar approach has been used in many studies on treatment effectiveness. For examples in mental health literature, see Knesper *et al*[9] and Frank.[30]
42. Robles R, Vera M, Alegria M, Martinez RE, Munoz CA. *Development of a Valid and Reliable Measure of Mental Health Need for the Puerto Rican Population*. Final report submitted to National Institute of Mental Health, Grant No. MH38727, 1986.
43. Auster R, Leveson I, Sarachek D. The production of health, an exploratory study. *J. Human Resources* 1969; **4**: 411–436.
44. Grossman M. *The Demand for Health: a Theoretical and Empirical Investigation*. New York: Columbia University Press, 1972.
45. Since a large percentage of the sample has no treatment, and the distribution of treatment among users is highly skewed, an ordinary least squares regression using treatment units as dependent variable will probably produce inconsistent results.[38] Restricting this part of the model to a subset of the sample sacrifices sample size but ensures the precision of the predictive power of the instrumental variable equation.
46. Ellis RP, Mwabu GM. *The Demand for Outpatient Medical Care in Rural Kenya*. Unpublished working paper, Economics Department, Boston University, 1993.
47. Burgess JF Jr, DeFiore DA. The Effect of distance to VA facilities on the choice and level of utilization of VA outpatient services. *Soc. Sci. Med.* 1994; **39** (1): 95–104.
48. Welch HG, Larson EB, Welch WP. Could distance be a proxy for severity-of-illness? A comparison of hospital costs in distant and local patients. *Health Services Res.* 1993; **28**(4): 441–458.
49. Hodgkin D, Alegria M, McGuire TG. *Travel Distance and Mental Health Care Utilization in Puerto Rico*. Unpublished manuscript, 1997.
50. They divided the city of San Juan into five sub-municipalities, since the provider data are available at this level of detail. In the case of respondents, the survey reports which census tract they live in (census

tract is more detailed than municipality). They therefore defined a respondent's residence as the centroid of the census tract where the respondent lives.

51. This approach sacrifices some accuracy for long trips where road miles would be more accurate than straight line miles. However, available road-mile software could not accommodate the level of detail about respondent location, thereby reducing the accuracy of our measure for short trips. Since most respondents live in municipalities with mental health providers, the method which maximizes accuracy for short trips was chosen. For further details on the definitions and discussion of the distance variable(s), see Hodgkin, Alegria and McGuire.[49]
52. Harrow BS, Ellis RP. Mental Health Providers' Response to the Reimbursement System. In: Frank RG and Manning W, eds. *Economics and Mental Health.* Baltimore and London: Johns Hopkins University Press, 1992.
53. Edlund M, Wheeler JRC, D'Aunno TA. Payment systems and payment incentives in outpatient substance abuse treatment. *Public Budgeting Finance Management* 1992; **4** (1): 107–123.
54. Newhouse JP. *Free for all? Lessons from the RAND Health Insurance Experiment*. Cambridge, Massachusset: Harvard University Press, 1993.
55. Dubin JA, McFadden DL. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* 1984; **52** (2): 345–362.
56. Maddala GS. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.
57. Greene WH. *Econometric Analysis*. 2nd edn. Macmillan, New York, 1993.
58. Leiderman L. Macroeconometric testing of rational expectations and structural neutrality hypotheses for the United States. *J. Monetary Econ.* 1980; 69–82.
59. Murphy K, Topel R. Estimation and inference in two-step econometric models. *J. Bus. Econ. Stat.* 1985; **3** (4): 370–379.
60. Newey WK. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *J. Econometr.* 1987; **36**: 231–250.
61. Hausman JA. Specification tests in econometrics. *Econometrica* 1978; **48**: 696–720.
62. The treatment effect of FORMAL2 is also estimated using a linear probability model. The results from this model, which might be inconsistent, show a negative and non-significant treatment effect. After the marginal effects are calculated, it could be seen that the treatment effectiveness estimated from this IV linear probability model is higher than that estimated from basic models.
63. There are two possible interpretations of the negative effect of previous treatment. One is that previous treatment is also a measure of illness severity. An individual with treatment history is more likely to have a persistent mental illness, and is, therefore, less likely to recover. It should also be noted that the 'previous treatment' used in this study refers to treatment in the previous year of first interview, and that the 'current treatment' refers to treatment in the previous year of follow-up interview. However, there is only a one year interval between the first and the follow-up interview. Therefore, it is possible that an individual who had just completed and discharged from 'previous treatment' would not receive any 'current treatment'. As a result of not being in treatment in the most recent year, it is more likely for the individual to relapse or to not improve in mental health status.
64. Instead of defining retaining no need for mental health care as '1', two new outcome variables are constructed: (a) outcome is defined as '1' if and only if need for mental health care in wave 2 is less that in wave 1; (b) the same as before except outcome is defined as '1' when need for mental health care is 'probably' in both wave 1 and wave 2.
65. All estimation and diagnostic test results from sensitivity analysis which are not presented in this paper are available from the author upon request.

