

Service Penetration by Persons with Severe Mental Illness: How Should It Be Measured?

Paul G. Stiles, JD, PhD

Roger A. Boothroyd, PhD

Kristen Snyder, PhD

Xiang Zong, MSPH

Abstract

As performance indicators and outcomes measures become essential parts of doing business, providers of mental health services are developing and using a number of access measures. One that is being used with increasing frequency is service penetration. However, the lack of standard methods for calculating and reporting service penetration has made the comparison of penetration rates across studies difficult. This article discusses the conceptualization and operationalization of service penetration. In addition, it presents an exploratory study of service penetration using data from the same persons using very different data sources; these data were collected during an evaluation of a Medicaid managed care system in Florida. The article offers recommendations for the use and reporting of service penetration rates.

Introduction

In the ever-changing landscape of public mental health care delivery systems, there is increased reliance on managed care approaches to control health costs, while attempting to improve access to care and quality of care.¹ Determining whether particular managed care approaches live up to the claim of reduced cost and increased access and quality has become essential for state decision makers charged with managing and monitoring managed care.² Yet the development of practical and reliable measures of these concepts has not always been consistent.

Address correspondence to Paul G. Stiles, JD, PhD, Associate Professor, Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida, 13301 Bruce B. Downs Boulevard, Tampa, FL 33612; e-mail: stiles@fmhi.usf.edu.

Roger A. Boothroyd, PhD, is Associate Professor, Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida.

Kristen Snyder, PhD, is Program Director, Leadership Development, College of Education, University of South Florida.

Xiang Zong, MSPH, is a Data Analyst, Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida.

Journal of Behavioral Health Services & Research, 2002, 29(2), 198–207. © 2002 National Council for Community Behavioral Healthcare.

This article focuses on the use and measurement of access to care, one of these three broad concepts that states continue to struggle with.³ *Access* can be defined as whether persons with a mental disorder are able to obtain appropriate services to treat or alleviate the disorder; however, in practice the operationalization of the concept can be much more complex.⁴

A number of access measures have been developed and used as performance indicators and outcomes measures, and they have become essential for providers of mental health services. Measures of process, such as waiting time to schedule office visits and telephone abandonment rates (ie, the number of patients who hang up because they feel they are waiting too long for the provider or insurer to answer their call) are traditional methods of measuring access among mental health care providers; however, recently a shift to the use of results- or outcomes-type indicators has gained prominence.⁵ The performance indicator approach (initial and developmental) adopted by the National Association of State Mental Health Program Directors (NASMHPD) exemplifies this trend. Although NASMHPD includes a process measure in its list of access indicators (average time to first service), it also includes several outcome- or result-oriented indicators (denial of care, consumer perception of access, and service penetration/utilization rates).⁶ The NASMHPD definition of service penetration is the unduplicated number of persons served during a year over the number of persons enrolled (or in the general population).

Increasingly, as the NASMHPD protocol demonstrates, service penetration is being utilized as an indicator of system access. (The concept of penetration is borrowed from the marketing context. *Market penetration*, or what percentage of the sales a particular product commanded in a specified market area, has been used to measure product utilization/access for decades in the business arena. With the expansion of the use of managed care plans in the health care market, the penetration of all health maintenance organizations [HMOs] or a particular HMO into a health care area or market has been a widely used indicator in the health care context. More recently, service penetration has been used to characterize how accessible a particular service or group of services is for a group of persons eligible to receive those services.) In its simplest form, service penetration is the proportion of individuals eligible to receive services who actually do receive a designated service(s) within a specified time frame. That is, the number of eligible persons who used the service(s) divided by the total number of eligible persons. This simple calculation also can be computed for various cohorts or subpopulations such as different age groups or categories of eligibility. However, as various types of adjustments are made to the basic formula to account for factors that may bias the interpretation of the rates (eg, risk adjustment), penetration rates become more difficult to compare.

Measuring Service Penetration

Two relatively recent studies^{7,8} used calculations of service penetration as key variables in measuring service access. The first study⁷ examined the use of mammography, operationalized as the proportion of women who had a mammogram in the past 2 years. The second study,⁸ which examined access to care for low-income uninsured individuals, defined access as whether a person had an ambulatory care visit in the past year. Both studies selected these measures because they are considered standard indicators of access to care; at first glance, they may seem fairly consistent and appropriate. However, one measures service use over a 2-year period, while the other employs a 1-year period. Are the rates obtained by the studies comparable? If they are not directly comparable, can they be adjusted for the differences in service time frame? Which is more useful, the 1-year or the 2-year service rate? Although service penetration is a useful tool, particularly if measured and compared longitudinally, there is unfortunately no standardized or universally agreed on method for estimating penetration rate. This can make it difficult, if not impossible, to compare rates across studies even if the studies are focusing on the same service system. Sources of variation in the calculation of penetration among studies that report penetration rates include the following:

- Differences in service category definitions (eg, what specific services are aggregated into the categories and what units are used to indicate a service contact—minutes, hours, sessions, days)
- Differences in the time of year sampled (eg, impact of seasonal differences)
- Differences in sampling periods (eg, monthly penetration, 6-month penetration, annual penetration)
- Differences in data sources (eg, claims, events, self-report, clinical records)
- Differences in service use thresholds (eg, any contact versus a prespecified level or specific number of contacts)
- Differences in client exposure rates within the same time intervals (eg, eligible for entire sampling period or only part of period)

These sources of variation can produce widely divergent results, making the comparison of studies difficult. For example, two commonly used “types” of penetration are annual penetration and average monthly penetration. Annual penetration is frequently used in research because it is fairly easy to calculate using existing data sets. In its simplest form it is calculated using the following formula:

$$\text{Annual penetration} = \frac{\text{Total persons having any service contact during the year}}{\text{Total persons eligible at some point during the year}}$$

Note that the formula’s sampling period is 1 year, the service use threshold is one contact, and the client exposure rate is eligible at any point in the year. If any of these parameters are changed, very different penetration rates are obtained.

Health care industry administrators use average monthly penetration more because their concern focuses on how many individuals they must provide care for at any point in time. A typical average monthly penetration formula follows:

$$\text{Average monthly penetration} = \frac{\text{Total user months over a 1-year period}}{\text{Total eligible months over a 1-year period}}$$

Even though the service use threshold is the same as the one used in the annual penetration formula (one service contact), the sampling period is 1 month (not annually) and the client exposure rate is 100% (the client is eligible for the entire 1-month period). Each individual month in the year for which data are available might have slightly different penetration rates, but the average monthly rate over the year is captured by the above formula.

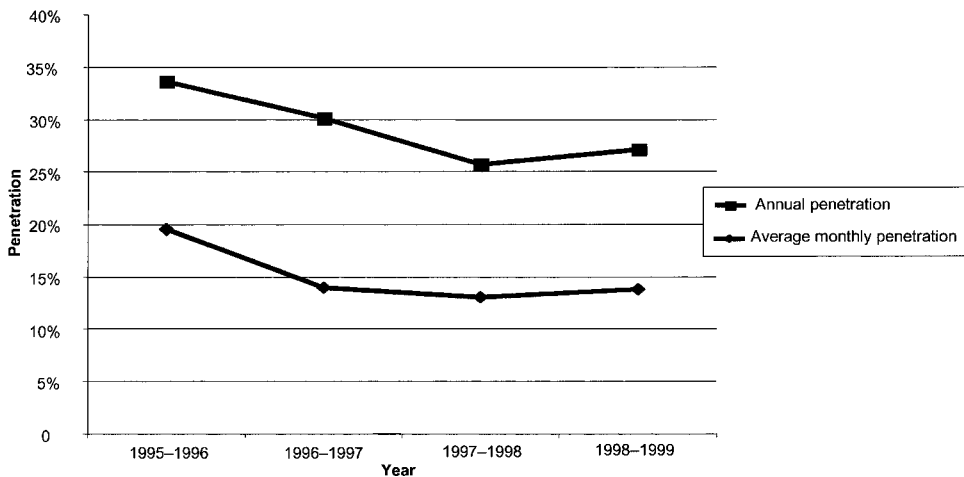
Figure 1 shows the implementation of the above two penetration formulas used with the same data: longitudinal service data for Medicaid recipients in the same eligibility category in a region of Florida. Over the 4-year period, the findings for annual penetration are consistently 12% to 17% higher than the findings for average monthly penetration. This reflects the nature of mental health service use by persons in the selected eligibility group. Over a year’s time 25% to 35% of the group used a mental health service at some point, but in any 1-month period only about 15% of the group used a mental health service on average.

Thus, the sources of variation (especially the method of calculating the penetration rate) can greatly impact the resulting service penetration rate obtained. Establishing standard methods of measuring and calculating service penetration, or at least establishing standard ways of reporting and describing how penetration was calculated, would improve the comparability and usefulness of penetration as a performance indicator.

An Exploratory Study of Penetration

In order to examine the concept of service penetration further, a study was designed using existing data sets collected as part of an evaluation of a Medicaid managed care system in Florida. A unique

Figure 1
Average monthly penetration versus annual penetration



characteristic of the ongoing evaluation is that simultaneous data collection has been implemented on the same study subjects using a variety of research techniques and data sources. This includes self-report survey data collected directly from subjects over the same time period as Medicaid system administrative claims data recorded paid claims for services. These two data sets were used in an exploratory study to examine the effects of different data sources and service use thresholds on penetration rates. The focus of this study is not evaluating the Florida Medicaid managed care system, but rather the concept of service penetration. Other reports have comprehensively described the Florida Medicaid population used in this study.^{9,10}

Data sets

Population-based self-report service utilization data were obtained through intensive interviews with the same individuals at two points in time. An initial sample of 688 adult Medicaid recipients with severe mental illness (as indicated by their receiving Social Security Income [SSI] as a result of mental health disability) participated in the initial wave of interviews; 587 of these individuals participated in the second wave of interviews. Most participants had diagnoses of schizophrenia or major affective disorder. The 688 individuals were identified through random selection of individuals with SSI eligibility in a complete Medicaid enrollment file, and they were originally recruited via a mail survey. Each interview requested a self-report of service utilization in several service categories over the 3-month period prior to the interview date.

Service information from administrative claims and encounter data were compiled for the same 688 persons from a comprehensive Medicaid claims and encounter data system. Only services occurring in the same 3-month time frame (3 months prior to each individual interview date) were included. Self-report service use was restricted to a 3-month period to improve the reliability of subject recall.^{11,12} The data were further aggregated into service categories that matched the self-report categories as closely as possible (eg, the definition of a “mental health case management” service in the self-report was operationalized in the administrative data by selecting the claims with procedures that most closely reflected mental health case management services in the Florida Medicaid system).

Analyses

Several sources of potential variation in service penetration rates were controlled for: service categories were as closely defined as possible; the sampling and seasonal time periods were identical; and, because the individuals in both data sources were the same, the client exposure rates were identical. The two primary sources of variation and the foci of this study were: (1) the source of the data and (2) the service use threshold for defining penetration (ie, penetration = one contact, three contacts, or five contacts). The penetration rate was calculated by determining the proportion of the 688 (or 587 in the second follow-up time period) who “penetrated” the service, using each of the three service use threshold definitions (one, three, or five contacts). In addition to examining the overall service penetration rates derived from each data source, an analysis was conducted to examine the level of agreement between the two data sources in terms of classifying individuals as service users and nonusers in all service categories.

Results

The results of this exploratory study are presented in terms of penetration agreement and individual agreement.

Penetration agreement

The results of the penetration agreement analyses are presented in Table 1. The seven categories of service (and one “overall” category) and sources of the data are indicated in the far left column, while the interview period and various service use thresholds are indicated along the top rows. For some of the services (eg, day treatment) there is remarkable concurrence of penetration rates across both self-report and administrative data and all three service use thresholds. However, for many other service categories, there were substantial differences in penetration rates. Differences across data sources ranged from just a few tenths of a percentage point (ie, day treatment) to over 40 percentage points (ie, evaluation).

To examine the service definition issue further, two definitions of medication use were implemented for self-report rates to compare with the rates for administrative data. “Medication” services were defined in the administrative data as the use of several procedures for medication management and the administration of medication, but not the actual medication use itself (ie, pharmacy information). Self-report medication category 1 included all medication service contacts including pharmacy contacts. Self-report medication category 2 included all medication services except pharmacy. The resulting penetration rates were striking. Rates for self-report medication category 1 were much higher than either self-report medication category 2 or the administrative data rates. Self-report category 2 and the administrative data rates were more comparable.

Individual agreement

The level of agreement across the two data sources (ie, administrative and self-report) also was examined to determine the extent to which the same individuals were consistently identified as service users or nonusers. Table 2 summarizes the number of service users identified by service type and data source for the baseline interviews using the one contact criterion. In addition, the level of agreement between the self-report and administrative data sources assessed using Cohen’s kappa is presented. As can be seen from inspection of this table, consumer self-report service use is consistently higher across each of the service categories compared with the service use rate derived from administrative data. Note that the levels of agreement among the two data sources regarding who used services (as assessed using Cohen’s kappa) are generally poor. Although there are no

Table 1
Penetration rates by service and data set

Service/source	Baseline interview (n = 688)			Follow-up (6-month) interview (n = 587)		
	One use (%)	Three uses (%)	Five uses (%)	One use (%)	Three uses (%)	Five uses (%)
Inpatient						
Self-report	3.4	0.6	0.4	2.0	0.0	0.0
Administrative data	2.2	1.2	1.0	4.1	2.0	1.7
Outpatient						
Self-report	58.9	27.5	13.3	42.5	26.2	11.0
Administrative data	21.8	7.7	3.8	18.7	6.8	3.6
Day treatment						
Self-report	8.2	6.3	5.4	6.5	6.0	5.5
Administrative data	7.0	6.7	6.2	5.8	5.3	5.1
Case management						
Self-report	24.7	15.4	5.4	19.3	11.3	3.5
Administrative data	21.9	19.0	14.2	20.4	16.5	13.3
Evaluation						
Self-report	52.3	27.5	11.7	50.9	26.5	5.7
Administrative data	9.9	0.3	0.0	10.7	0.5	0.0
Medication						
Self-report 1	63.2	38.9	16.8	65.5	50.3	12.9
Self-report 2	28.9	16.6	4.5	23.7	12.4	3.0
Administrative data	38.4	13.1	4.8	33.4	9.7	4.3
Primary care						
Self-report	17.9	9.2	1.0	11.7	4.7	0.8
Administrative data	6.7	0.7	0.1	5.1	1.0	0.2
Overall						
Self-report	73.8	50.9	27.0	63.4	40.1	19.9
Administrative data	54.9	35.8	26.2	51.4	33.4	24.5

Table 2
Number of service users identified by data source and level of agreement

Service	Number of users in both data sources	Number of users in self-report data only	Number of users in administrative data only	Cohen's kappa
Inpatient	4	19	10	.20
Outpatient	86	194	63	.16
Day treatment	26	30	21	.46
Case management	102	66	48	.53
Evaluation	54	303	14	.10
Medication	94	104	169	.12
Primary care	12	108	31	.06
Overall	344	164	34	.29

fixed cutoffs for interpreting kappa coefficients, various guidelines suggest that values below .40 are indicative of poor agreement.^{13,14}

Discussion

Penetration agreement

The pattern of penetration rates across service categories, use thresholds, and data sources demonstrate the great variability that can occur even when using the same method of penetration calculation. Services such as day treatment showed very similar penetration rates across both data sources. This could have resulted because day treatment is fairly well defined by procedure code in the administrative data and by service activity as captured by self-report. More easily defined and recognizable service categories may be reflected more consistently in the data, regardless of data source. However, outpatient care is another fairly well-defined service category in both data sources; yet, there is considerably higher self-report penetration than administrative data penetration for all use thresholds and time periods. Thus, how well a service category is defined is only one factor in consistent penetration rates. The pattern reflected in the outpatient service category may have resulted from a merging of formal (billable) and informal (non-billable) care in the minds of the self-reporting clients; it also may demonstrate the limited capacity of administrative data to reflect all types of mental health services. Similarly, in the evaluation service category, self-reported service use is much higher than that captured by the administrative data. This also might reflect the wide variety of contexts in which an evaluation might occur, but is then not recorded or billed as an evaluation. Thus, the administrative data may not reflect all the evaluation that is occurring. Regardless of the reason for the discrepancies, it is clear that penetration rates calculated for these services using self-report and administrative data resulted in quite disparate rates.

The rates for case management also present an interesting pattern. Case management is another fairly well-defined service category. Yet, although the rates are similar across data sources when using the one-use threshold, there are much higher rates indicated by the administrative data as the use threshold is increased to three and then five uses. It is possible that while the clients who are self-reporting may know they are receiving case management, they may not recall the volume of the service received, which is more precisely captured by the administrative data. Moreover, billable case management services do not necessarily have to be face-to-face services, which is why clients may not be aware of the volume of services they are actually receiving.

In the overall service use category, rates reported for the one-use service threshold are fairly disparate, but there is a noticeable convergence in rates obtained using the two data sources, as the service use threshold is increased from one to five uses. This could again reflect an intensity of service issue. More people self-reporting service use may believe that they received a mental health service, but instead received something less formal or intensive (non-billable). As the intensity of the services increased (ie, five uses), they may be more accurate in their reporting. Of course, this assumes that the administrative data are more reflective of reality, which could very possibly be a faulty assumption.

The two approaches to defining self-report medication use most clearly demonstrate the importance of service category and use threshold definition. It is clear that the inclusion of pharmacy in the medication definition for self-report altered the comparability of the self-report to the administrative data. How a category is defined is critical and should always be included in any report of penetration rates.

Individual agreement

The level of agreement was examined for all service categories, and three service categories in this exploratory study, representing several combinations of high or low penetration agreement and

high or low individual agreement (kappa), were selected for discussion. Inpatient services represents a category where there was fairly high penetration agreement and low individual agreement. The baseline penetration rates were 3.4% for self-report and 2.2% for administrative data. Consistent classifications were observed in both the administrative and self-report sources for 659 of the individuals (95.8%). The initial impression of this finding is that the consistency of classification is extremely high. However, given the low base rate of inpatient utilization, the consistency of classification among only users was examined. Of the 33 individuals who were identified as having used inpatient services in one or both of the data sources, only 4 (12.1%) were classified as having used inpatient services in both data sources. Of the 29 disagreements, 19 (65.5%) were inpatient stays reported by respondents that did not appear in the administrative data while 10 disagreements (34.4%) were inpatient stays found in the administrative data that were not reported by respondents. This perspective suggests a very low level of agreement across data sources regarding the specific identification of users of inpatient services despite relatively similar penetration rates. A subsequent analysis was undertaken to determine the reasons for the lack of agreement across data sources regarding inpatient service use. Several reasons were identified. First a number of respondents who self-reported having inpatient stays reported they occurred in settings that were not Medicaid reimbursable and therefore would not appear in the administrative data. These included stays at Veterans Administration (VA) hospitals and in crisis stabilization units (CSUs). CSUs are intended for brief stays for the purpose of stabilization of a mental health condition, and their cost is not reimbursable in Florida under Medicaid. Similarly, inpatient stays in VA hospitals are not reimbursable under Medicaid and do not appear in the administrative data. A second explanation for these disagreements could relate to establishing the time frames within which inpatient stays were examined. For respondents, accurately determining which hospitalizations occurred within the past 6 months might be challenging. Hospitalizations that happened 7 months previously might easily be recalled as occurring within the past 6 months and thus would result in a disagreement. With the administrative data, delays in filing claims, challenges to claims, and subsequent billing adjustments also might contribute to these disagreements.

Case management services represent another category with high penetration agreement, but with much higher individual agreement. With respect to the utilization of case management services, the penetration rate was 22.9% when based on the administrative data and 24.7% when based on self-report data. Both of these rates are quite similar. When case-by-case comparisons were examined, 574 (83.4%) were consistently classified across both the administrative and self-report data sources as having used (or having not used) these services. This rate of agreement is lower than the 95.8% agreement rate found for inpatient services but is based on a much higher base rate. Of the 216 individuals who were classified as having used case management services in one or both of the data sources, 102 (47.4%) were classified as having used these services in both data sets. This finding is substantially higher than the 12.1% agreement among users of inpatient services but still lower than one would hope. Of the 114 disagreements, 66 (57.9%) were case management services self-reported by respondents that did not appear in the administrative data, while 48 disagreements (42.1%) were case management services appearing in the administrative data that were not reported by respondents. As was the case with the analysis of inpatient services, respondents were more likely to report using services that were not verifiable in the administrative data, although the disagreements regarding case management utilization were more evenly split between the two data sources. Some of these disagreements could result from non-client contact services in which providers bill Medicaid for case management services, while clients might be unaware that they had received any of these services and thus would not report them.

Mental health evaluation services was a category with relatively low penetration agreement and low individual agreement. The penetration rates for mental health evaluations were 52.3% when based on the self-report data and 9.9% when based on the administrative data. Given that the penetration rates are dramatically different across the two data sources, it is not surprising that the rates of disagreement also are very high. Only 371 (53.9%) had the same classification across both data sources, significantly

lower than either the inpatient or case management findings. Of the 371 individuals identified as having had an evaluation in at least one of the data sources, only 54 (14.6%) appeared in both data sources. Of the 317 disagreements, 303 (95.6%) involved individuals self-reporting an evaluation that was not substantiated in the administrative data while only 14 disagreements (4.4%) resulted from claims in the administrative data that were not self-reported by respondents. One explanation for some of these disagreements is likely attributable to the narrow definition permitted when billing for evaluation services in relation to clients receiving “evaluative” services that were not billable under this definition.

Disagreements in service use reported in the two data sources were fairly consistent. In each of the three services discussed, service use was more likely to be self-reported and not verified in the administrative data than identified in the administrative data and not self-reported. These differences are in part understandable and expected given that the administrative data are restricted to services paid by Medicaid, whereas self-reported services could have been paid through any source. These disparate individual agreement results suggest that direct comparison of penetration rates across data collected using these two methods should be conducted with great caution.

Conclusions

This exploratory study of penetration rates was designed to demonstrate the importance of understanding what is being measured when claiming to report a penetration rate. Even when examining data from the same 688 individuals using two collection methods (and using identical penetration calculations), very different results can be obtained. If service penetration is to become a reliable and valid measure of access, standard methods of measurement and calculation must be developed or, at a minimum, it must be clearly and carefully defined when used and reported.

Implications for Behavioral Health Services

As service penetration is used more widely as an indicator of system access, perhaps a more standard methodology will be adopted as the most accepted way to calculate and report service penetration. However, until that time, it is crucial that any researchers or evaluators who use service penetration and report rates in their work provide at least seven pieces of information:

1. A definition of service categories. If the rates reflect “service” penetration, what services are being penetrated: Any mental health service (including pharmacy)? Just case management as reflected by three procedure codes? Only formal services, or are both formal and informal services captured?
2. A statement defining the sampling time period. What is the time frame of the data used: 1 month, 6 months, or 1 year? What is being calculated: monthly penetration or annual penetration? Could there be any influence from seasonal differences?
3. A description of the data source. Are the data self-reported? Administrative records? Claims or encounters?
4. A definition of the service use threshold. What constitutes a penetration of the service or system: one contact, one claim, five claims, a note in the clinical record, or a recollection of service?
5. A description of the population whose penetration rate is being studied. Is it the entire population of potential users of service? Is it a cohort defined by diagnosis/disorder? Is it a subpopulation defined by geographic area?
6. A definition of the client exposure within the time frame defined above. Were the clients eligible to receive services for the entire time frame or just part of it? If only eligible for a part, is that taken into account in the penetration formula denominator?

7. A description of the penetration formula used. Provide the actual formula used to calculate service penetration. Indicate whether it is a simple proportion or an average proportion over a longer time frame.

If the above elements are consistently reported in studies reporting service penetration rates, the comparability and usefulness of service penetration as a performance indicator will be greatly improved. As reporting becomes more consistent, consensus standards may develop for data collection and organization to further enhance the calculation of service penetration. The ultimate goal is to improve the access to, and management and delivery of, high-quality mental health services. Having consistent, reliable, and valid measures of access is essential to moving toward that goal.

References

1. Freeman MA, Trabin T. *The Policy Challenges of Managed Behavioral Healthcare*. Washington, DC: Mental Health Policy Resource Center; 1995.
2. Miller RH. Healthcare organizational change: implications for access to care and its measurement. *Health Services Research*. 1998;33(part II):653–680.
3. Connecticut Medicaid Managed Care Council. *Behavioral Health Subcommittee Meeting Summary* (April 29, 1998). www.cga.state.ct.us/ph/Medicaid/mmcc/bh/minutes/bh0498.htm. Accessed February 20, 2002.
4. Gold M. Beyond coverage and supply: measuring access to healthcare in today's market. *Health Services Research*. 1998;33(part II):625–652.
5. Kramer T, Trabin T. *Performance Indicator Measurement in Behavioral Healthcare*. Portola Valley, CA: Institute for Behavioral Healthcare; 1997.
6. National Association of State Mental Health Program Directors. *Performance Measures for Mental Health Systems, Domain: Access*. Alexandria, VA: NASMHPD; 1997.
7. Makuc DM, Breen N, Fried V. Low income, race, and the use of mammography. *Health Services Research*. 1999;34(part II):229–239.
8. Cunningham PJ. Pressures on safety net access: the level of managed care penetration and uninsurance rates in a community. *Health Services Research*. 1999;34(part II):255–270.
9. Shern D, Robinson P, Boothroyd R, et al. *Evaluation of Florida's Prepaid Mental Health Plan Year 2 Report*. Tampa, FL: Louis de la Parte Florida Mental Health Institute; 1999.
10. Shern D, Robinson P, Stiles P, et al. *Evaluation of Florida's Prepaid Mental Health Plan Year 3 Report*. Tampa, FL: Louis de la Parte Florida Mental Health Institute; 2000.
11. Kashner TM, Suppes T, Rush AJ, et al. Measuring the use of outpatient care among mentally ill individuals: a comparison of self reports and provider records. *Evaluation and Program Planning*. 1999;22:31–39.
12. Wallihan DB, Stump TE, Callahan CM. Accuracy of self-reported health services use and patterns of care among urban older adults. *Medical Care*. 1999;37:662–670.
13. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley & Sons; 1981.
14. Landis J, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.