AN EXAMINATION OF THE ACCURACY OF MEDICAID CLAIMS DATA, A STATE MANAGEMENT INFORMATION SYSTEM, AND COMMUNITY MENTAL HEALTH CENTER CLINICAL RECORDS IN HAWAI'I

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

MAY 2004

By

Julie A. Slay

Dissertation Committee:

A. Michael Wylie, Chairperson
John J. Steffen
Paul G. Stiles
Kelly M. Vitousek
Eldon L. Wegner
Acknowledgements

I would like to thank my committee, Dr. Michael Wylie, Dr. Eldon Wegner, Dr. John Steffen, Dr. Paul Stiles and Dr. Kelly Vitousek for their critical feedback and willingness to support this research and my career. I would like to give special thanks to Dr. Wylie for his patience, support and guidance throughout this process and Dr. Bob Anders and Dr. Anthony Marsella for their support early in the development of this research.

Without cooperation from the Adult Mental Health Division at the Hawai‘i Department of Health this research would not have been possible. I would also like to acknowledge several colleagues that provided their knowledge and unrelenting support. These include: Dr. Roger Boothroyd, Dr. Jason DeViva, Lilia Espinoza, Dr. Edward Suarez, Dr. Charlotte Sykora and April White.
Abstract

A variety of administrative data sets have been used to evaluate mental health outcomes, service utilization patterns, and quality of care. Data sets most commonly used are claims data, management information systems data, and information collected directly from medical records or clinical charts. When these sets of data are used to answer research questions, they are typically used outside the scope of the intended use of these data systems. More and more often researchers are using Medicaid health insurance claims to understand mental health services because of the breadth and depth of information contained in this federally supported database. Therefore, the accuracy of the information contained in this set of data needs to be determined in order to make confident recommendations and conclusions based on research utilizing this large administrative database. This study examined the reliability, or agreement, of Medicaid Insurance claims data through a comparison to a state management information system and clinical chart data of mental health service utilization at community mental health centers in the state of Hawai‘i. Six major categories of mental health services were compared as was primary diagnosis across the three data sets. Analyses did not confirm that data sets were reliable when examined against each other. Characteristics of the administrative data affected the ability to compare it to data extracted from charts. In general, compared to other literature, rates of agreement among collected chart data, claims data and the state MIS were low. Service categories that were more unique and required aspects of care not present in other services such as drug administration or group therapy had better rates of agreement. Service utilization patterns also differed depending on the data source examined, although the differences were not statistically significant. However, from a
practical perspective if administrators use one data set exclusively over another without understanding discrepancies within systems, over- and underestimation of service utilization could occur. Findings suggest that administrative databases and chart data are not equivalent and a process to ensure reliable and accurate data must be established before such data are used for policy analysis.
# TABLE OF CONTENTS

Acknowledgements ........................................................................................................ iii
Abstract ............................................................................................................................ iv
LIST OF TABLES ............................................................................................................. viii
LIST OF FIGURES ........................................................................................................... ix
CHAPTER ONE ................................................................................................................ 1
  Introduction .................................................................................................................... 1
CHAPTER TWO ................................................................................................................ 5
  Literature Review ......................................................................................................... 5
    Definitions .................................................................................................................... 5
    Methodological Issues When Using Administrative Data ........................................ 6
      Access and Generalizability ...................................................................................... 6
      Data Confidentiality ................................................................................................. 9
      Statistical Issues ..................................................................................................... 10
    Reliability of Administrative Data ........................................................................... 15
      Summary of Reliability Studies .............................................................................. 28
    Applicability of Administrative Data ....................................................................... 28
      Summary of Applied Administrative Data Studies ................................................. 31
    Rationale and Research Questions ............................................................................. 31
CHAPTER THREE ............................................................................................................ 34
  Methods ....................................................................................................................... 34
    Design ........................................................................................................................ 34
    Preliminary Study ...................................................................................................... 34
    Sources of Data ......................................................................................................... 35
      Chart Review Data .................................................................................................. 35
      Medicaid Claims ................................................................................................... 37
      MFASIS .................................................................................................................. 38
    Procedures ................................................................................................................ 39
      Crosswalk Table ..................................................................................................... 39
      Eligibility ................................................................................................................ 41
      Data Cleaning and Preparation ............................................................................. 41
      Agreement Analyses ............................................................................................... 44
    Results ....................................................................................................................... 47
      Data Cleaning and Preparation ............................................................................. 47
      Agreement Analyses ............................................................................................... 50
    Drug Administration ................................................................................................. 50
    Evaluation/Assessment ............................................................................................. 52
    Family Therapy ......................................................................................................... 53
    Group Therapy .......................................................................................................... 54
    Medication Check ..................................................................................................... 55
    Therapy/Case Management ....................................................................................... 56
    Agreement of Diagnostic Codes ............................................................................. 58
    Service Utilization ..................................................................................................... 59
    Funding Source ........................................................................................................ 62
CHAPTER FIVE ................................................................................................................ 65
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Review of Reliability Studies</td>
<td>17</td>
</tr>
<tr>
<td>2. Data Cleaning and Preparation</td>
<td>43</td>
</tr>
<tr>
<td>3. Gender and Diagnostic Category of Participants by Source</td>
<td>49</td>
</tr>
<tr>
<td>4. Mean Age of Participants</td>
<td>50</td>
</tr>
<tr>
<td>5. Agreement Rates of Drug Administration</td>
<td>51</td>
</tr>
<tr>
<td>6. Agreement Rates of Evaluation/Assessment</td>
<td>53</td>
</tr>
<tr>
<td>7. Agreement Rates of Family Therapy</td>
<td>54</td>
</tr>
<tr>
<td>8. Agreement Rates of Group Therapy</td>
<td>55</td>
</tr>
<tr>
<td>9. Agreement Rates for Medication Checks</td>
<td>56</td>
</tr>
<tr>
<td>10. Agreement Rates of Therapy/Case Management</td>
<td>58</td>
</tr>
<tr>
<td>11. Agreement Rates of Diagnostic Codes</td>
<td>58</td>
</tr>
<tr>
<td>12. ANOVA for Evaluation/Assessment</td>
<td>60</td>
</tr>
<tr>
<td>13. ANOVA for Medication Check</td>
<td>61</td>
</tr>
<tr>
<td>14. Agreement Rates for Service Categories by Funding Source</td>
<td>64</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mean Number of Service Episodes by Service Category</td>
<td>59</td>
</tr>
<tr>
<td>2.</td>
<td>Total Number of Service Episodes by Service Category</td>
<td>62</td>
</tr>
</tbody>
</table>
CHAPTER ONE

Introduction

In March 2001, a report from the Medi-Cal Policy Institute reported that the $44 million dollar managed care data system used to track persons using California's Medicaid managed care program could not be “used to make sound policy decisions” because data were inaccurate and incomplete (Medi-Cal Policy Institute, 2001). Despite reports such as this, researchers more often are using large banks of data to examine various questions about health care policy (Drake & Johnson-Reid, 1999). These inquiries include epidemiological studies, utilization reviews, and outcome analyses. A variety of different data sets have been used including insurance claims or administrative data, medical records or charts, and survey data (e.g., Fine, Keogh, Cretin, Orlando, & Gould, 2003; Garnick, Hendricks, & Comstock, 1994; Unützer, Simon, Pabiniak, Bond, & Katon, 2000). Claims data are computerized records of services that have been submitted to an agency that reimburses the provider of services. There are also computerized provider records, often referred to as management information systems, which contain a record of all services provided regardless of billing status. Medical records, or chart data, are often seen as the gold standard, the paper record that contains all services provided by that health care provider and may provide more detail than computerized systems (Drake & Johnson-Reid, 1999).

Chart data may be seen as the most accurate representation of true service provision, since, typically, information is entered directly by the person providing service, whereas computer recording may involve transcription by another person specifically charged with data entry. Further, administrative data systems tend to focus
predominantly on billable services, while charts often provide a complete record of services with less influence by billing requirements. There are limitations and advantages for all three types of data and it is important to consider the information lost by using just one source. However, including all three when examining utilization of services or outcome can be time-consuming and costly.

Sets that contain claims often only include services that are billable, therefore all provided services might not be reported in anticipation of a denial or unpaid claim. Management information systems (MIS) should include all provided services because they are typically developed to assess outcome, utilization and quality management information, and include billing processes as one part of this larger system. However, this data source is susceptible to data entry error on the part of the data entry personnel. Historically, clinicians have eschewed electronic recording of information, preferring the traditional pen and chart method of recording.

The advantages of using claims and information systems data arguably outweigh the advantages of using chart data, which take large amounts of time and are costly to review and aggregate compared to existing computerized and archived data. Administrative data, which include both claims and MIS data sets, usually contain information about a large number of persons with a variety of illnesses that tend to be representative of patient populations under study. Outcomes can be measured easily by mortality or relapse through examination of these data sets. Having such a large amount of data lends greater power and generalizability to studies using these data.

There has been a change in the health care system in the United States in the last 30 years. With the rising cost in health care, along with the changes in federally funded
insurance programs such as Medicare and Medicaid, Garnick et al. (1994) argued that the quality of health care and health care utilization will have to be measured and that data systems, such as claims or MIS databases, must be able to support the kind of questions that will be asked. In 1997, Robinson and Tataryn argued that computerized information systems are becoming more common because of the growing need for accountability, outcome evaluation, and program planning. Accuracy of the data must be established if analyses from those data will successfully inform policy makers or researchers about a phenomenon occurring within a health care system. Drake and Johnson-Reid (1999) echo that and assert that researchers should inform policy decisions and outcomes through analyses of such data. The Medi-Cal Policy Institute study found that because of the lack of consistent and reliable data their system could not provide accurate information about service use and ultimately cost. Though many authors have used these large sets of data compiled by governments or private health care providers to answer a variety of questions, there is concern about whether these data sets are reliable and valid enough to inform researchers about the true nature of health care. More importantly, are they valid enough to be used to make policy decisions about health care?

This study examined the reliability, or agreement, among three sources of data for a group of adults with severe and persistent mental illness. Hawai’i Medicaid claims data, MIS data from the state of Hawai’i’s Adult Mental Health Division, and chart data collected at community mental health centers in the state were analyzed. Service utilization patterns were estimated using all three sources of data, and a comparison among these data sets demonstrated how different provision of services appear when
approximated by each source of data. The impact of these differences on research and policy are discussed.
CHAPTER TWO

Literature Review

There have been a variety of commentaries about the use of large administrative data sets ranging from supportive to cautionary perspectives. These commentaries focus on the methodological and statistical advantages of using large administrative data sets, as well as the potential for limited generalizability, and raise concerns over the validity of outcomes using these data. This literature review will define what administrative data are and then review a number of major considerations when using administrative data. These considerations are organized by and include the follow areas: access and generalizability, data confidentiality and statistical issues. A review of studies that examine the reliability of administrative data follows and finally and a discussion of studies that use administrative data for the purposes of examining service utilization and outcomes in mental health services is presented.

Definitions

Administrative data are typically large computerized databases that collect a variety of variables to record the service history of clients. A typical administrative database used in research on service utilization is health insurance claims data that include information about the patient, his/her condition (e.g., diagnosis), and medical encounters (e.g., service records). Claims data are bills submitted by providers for specific services rendered. Claims data provide a plethora of information about the system of health care within an agency or funding body and about the population served. The location, date, extent of service, and the provider are all contained within the set as are participant descriptors. Diagnoses, symptoms, tests, and procedures administered are further examples of items
included. There are clearly advantages and disadvantages when using these data sets, as
there are with any data collection procedure. The following section will review those
advantages and disadvantages along with some general concerns that must be considered
when using administrative data.

Methodological Issues When Using Administrative Data

Access and Generalizability

Despite the advantages of administrative data, at times these data are not easy to
use or access. Typically, these sets are developed by agencies for inter-agency use only,
and outsiders using these data must learn the specific system in which the data are
housed. Potvin and Champagne (1986) suggested than an interdisciplinary team
knowledgeable in statistics, measurement theory, and advancements in technology is
necessary to examine and process these data files because of the varied and complicated
nature of using computerized databases. Different administrative files can be held in a
variety of database programs and formats and when manipulating these data, it is
important to have the expertise available in order to identify properly and analyze key
variables. However, after experts have established a protocol to translate and code data in
a way that is meaningful for the researcher, new analyses can be conducted easily (Drake
& Johnson-Reid, 1999) and immediately with little cost. Understanding potential uses for
data requires system development. A lack of understanding of how data may be used and
what the data represent can greatly limit the “usability” of such data for researchers and
administrators (e.g., Medi-Cal Policy Institute, 2001).

Identifying where a computerized file is located within an organization and
determining whether or not access can be granted are important considerations in the
decision to use administrative data as well. Which organization has the desired data and the merging of multiple files must be primary concerns when using administrative files. If access cannot be granted or if usage is limited by the agency from which the data are obtained, analysis and utility could be restricted.

Each agency has its own criteria for persons it serves; therefore these data sets include only those who are eligible for services under that agency. Intra-agency analyses would have strong validity for its participants because of the homogeneity of persons covered by that provider; however, this can greatly affect a researcher’s ability to compare data across agencies or apply results to other populations not served by that agency. If data were taken from only one agency, generalizability to the larger population is threatened, thus limiting the utility of the research (Potvin & Champagne, 1986). For example, Bright, Avorn, and Everitt (1989) warn that, though pharmacological data from Medicaid claims can be helpful in examining compliance with medication regimen and patient condition, state restrictions and regulations on pharmacies can influence the use of prescriptions state by state, thus requiring more careful interpretation of prescription use when analyzing claims across states. In the case of Bright et al., the state is the agency, and because regulations differ across states, a meaningful comparison is limited.

On the other hand, some agencies with administrative data, like the federal Medicaid health insurance program, receive information from every state about the utilization of health care by Medicaid recipients. Each state collects data in order to bill Medicaid. Thus, submitted claims by states to Medicaid create a database reflecting a geographically and a culturally diverse population. However, Medicaid data represent those who are disabled and poor. Though ethnic diversity may be present, biases are
inherent because of the homogeneity of socioeconomic status. Other databases have other inherent biases in the population served; therefore the persons contained within a set may not be representative of the general population, threatening the external validity of studies using administrative data (Bright et al., 1989). Hence, it may not be appropriate to make a prediction for an individual or group based on a sample sharing few common characteristics. Ray and Griffin (1989) echo this concern while suggesting that the population characteristics of Medicaid are both a limitation and strength. A sample of Medicaid participants may be diverse in race and ethnicity, age, and illness. Though diversity in these variables may lend external validity, it may also produce confounded effects.

Aggregation of data from many persons who may not be receiving care under the same conditions can also restrict conclusions that are made. Melfi, Holleman, Arthur, and Katz (1995) reviewed four different approaches in predicting outcome with Medicare claims data. They examined two measures, 30-day mortality and length of hospital stay, to assess outcome after patients underwent total knee replacement surgery (TKR). They found that the number of diagnoses given to a patient best predicted outcome. In other words, the number of comorbid conditions explained the outcome of TKR surgery better than other predictive models. Additionally, these other predictive models were found to be highly susceptible to inaccurate coding of data used for analyses. Thus, the aggregation of data points may not accurately represent the conditions under study, resulting in unreliable information about the result of medical procedures. Melfi et al. found one model in which the diagnoses provided strong predictive validity for a medical procedure. However, consideration should be given to the fact that administrative data
may not be accurate even though its use has many advantages, such as the potential for homogeneity of subjects, ease of use and low cost when compared to other types of experimental designs (Rosko, 1988).

Data Confidentiality

Administrative data contain a number of identifying variables necessary to process claims and track service use by individual participants. A major drawback to using these files lies in the need to protect the confidentiality of persons whose records are within the data file. Several researchers have recognized the value of using large sets of data but were concerned about confidentiality of persons within the data file (Bright et al., 1989; McDougall, Adair-Bischoff, & Grant, 1995). Medicaid and Medicare data have unique identifiers, but also maintain participant names and social security numbers, making it possible to merge sets of administrative data when a researcher lacks the unique identifier. Unfortunately, these files become more susceptible to violations of confidentiality. However, by using unique identifiers alone, confidentiality can be protected. In addition, this strategy may help in the elimination of attrition and participant refusal (Lohr, 1990).

The Health Insurance Portability and Accountability Act of 1996 (HIPAA), enacted by the federal government, required the Department of Health and Human Services to establish national standards for electronic health care transactions and national identifiers for providers, health plans, and employers. HIPAA addresses the security and privacy of health data from any health care organization. The purpose of the Act was to improve the efficiency and effectiveness of health care systems by encouraging the use of administrative data. Center for Medicaid and Medicare Services (CMS) has developed
standards for reporting and exchanging electronic data to avoid the problems linking data that do not share common operational definitions or linkages. HIPAA has the potential for standardizing and creating systems that may be easier to use for analyses; however, standard guidelines and practices must consider the potential uses for this data in order to ensure a database that can inform on service utilization, outcomes, and policy.

**Statistical Issues**

Despite risks to confidentiality, a major advantage of administrative data lies in the number of variables that can be used to match participants across sets and the ability to draw a random sample from population databases (e.g., Medicaid and Medicare) void of non-responder and volunteer bias because files include all persons who have received treatment (Potvin & Champagne, 1986). The possibility of poor memory on the part of the participant in accurately reporting health service episodes is also eliminated. The large number of subjects lends greater power to statistical analyses as well (Lohr, 1990; Motheral & Fairman, 1997).

The combination of data sets can be advantageous as well. For example, both Medicare and Medicaid typically insure elderly persons; thus, the state may bill Medicare before Medicaid, then bill Medicaid for uncovered deductibles or any unpaid services. For studies using those aged 65 and over, Medicare data can be linked with Medicaid in order to capture more accurate picture of use and cost of services for the poor and/or disabled elder (Bright et al., 1989; Whittle, Steinberg, Anderson, & Herbert, 1991).

Comparing findings from multiple, independent databases can establish validity and credibility to the use of administrative databases to inform policy. Drake and Johnson-Reid (1999) reference several studies that report similar rates of recidivism of
child abuse and neglect. Each study used independent data systems in different geographic regions but similar outcomes were found. Replication of this type speaks to the ability of administrative data to produce results that can be considered unbiased despite various ways of collecting information.

Using archived service data collected for purposes other than evaluating outcome or utilization has some limitations that may weaken the ability to evaluate outcomes or utilization patterns. There is little control over subject selection – all patients are users and have received services in some way. Therefore, no control group can be selected.

Randomized clinical trials (RCT) have become the gold standard used for outcomes research because of the ability to control variables that would otherwise confound results (Gilbody, House & Sheldon., 2002; Poses, Smith, McClish & Anthony, 1995). RCTs are not immune to other methodological problems and have their own set of advantages and disadvantages. Threats to external validity (e.g., exclusionary criteria and atypical experimental settings), in addition to the high cost and ethical concerns when withholding treatment, are serious disadvantages of RCTs that at times may outweigh benefits of randomization and control. These disadvantages make administrative data attractive because they can be comprehensive, inexpensive and easy to use, and they contain a large number of cases providing more power in studies utilizing such data (Motheral & Fairman, 1997). Administrative data can also be an effective tool in longitudinal designs because data cover extended periods of time, making this method more flexible than RCTs (Drake & Johnson-Reid, 1999; Motheral & Fairman, 1997; Quam, Ellis, Venus, Clouse, Taylor & Leatherman, 1993). Records of patients in a given population independent of treatment location or practitioner over time have some characteristics of,
if not randomization, then at least "arbitrariness," and examining the effects of services or medications over long periods of time can give the researcher flexibility in how they track patients – either by date of service or by onset of treatment, for example.

However, there are many threats to internal validity when utilizing claims data. Primarily, correlations found in the data do not necessarily explain the phenomenon being studied. Unreliable diagnostic information, missing data, lack of information regarding the severity of illness, and referral bias have been argued to confound results (Motheral & Fairman, 1997; Rosko, 1988; Sena & Pashko, 1993). Because of the nature of administrative data and lack of researcher control over data collection, the potential for error at data entry and data management may increase. Such error is difficult to account for and the researcher is dependent on consultation with data managers for information about the process of data handling. However, Drake and Johnson-Reid (1999) assert that despite the potential for error, administrative data can produce results that are replicable in other unrelated data sets.

Problems may not occur at the source of data entry or even data management, but with the actual data. An analysis of a Medicaid-managed care organization in California details the variety of errors that can be introduced into a data system, subsequently creating a large yet unusable source of data (Medi-Cal Policy Institute, 2001). Analysts found that providers were not submitting claims because of a lack of incentive under capitated plans that guarantee a flat payment for each member regardless of whether that member utilized a service. Also, there is a lack of standardization of how to report services. For example, there is no standard procedure for assigning a unique identifier so that information from multiple systems cannot be linked. Also, some providers use
updated procedure codes that have not yet been incorporated into the system so that
provided services may not be captured accurately. Without built-in checks for quality
management such data can contain errors that may bias findings (Motheral & Fairman,
1997) or render the data useless.

Despite such potential for error, administrative data can be used to examine more
rare events because of the large number of persons these data represent. It has been
recommended that when using claims data, groups be made as similar as possible with
identification of a specific time frame and diagnosis or treatment used to indicate a
serious condition. The ability to aggregate particular services and diagnoses when using
administrative data lends stronger predictive validity for the outcome of certain
conditions that might be more serious or rare (Drake & Johnson-Reid, 1999; Lohr, 1990).
Finally, an examination of the accuracy of the data being used is necessary to provide
some control over possible confounds (Motheral & Fairman, 1997).

Even with these limitations, a refined database system reflecting services across
hospitals and caregivers could aid in securing better continuity and quality of care.
Examination of trends in patient care and service provision across an entire region could
take place, taking advantage of a system that collects detailed information from a variety
of places and persons (McDougall et al., 1995).

Stiles, Bahl, Bernstein, Halman, Van Harrison and Standiford (2000) examined the
performance of a managed care organization (MCO) on standardized Health Plan
Employer Data and Information Set (HEDIS) measures. These measures allow
employers, policy makers, consumers, and other interested parties to examine managed
care companies on a variety of indicators including: effectiveness of care, access and
availability of care, patient satisfaction, health plan stability, utilization, and cost. The MCO doctors argued that the statistics reported using MCO data were not reflective of the quality of care that patients received. The authors augmented the MCO data with information from medical center administrative data to approximate a more accurate picture of service delivery. Augmented data improved the outcome measures and were more aligned with provider estimates. They suggested that a trustworthy data set must be used when compiling information from administrative data that will be used by stakeholders. Accreditation alone is important to various institutions and it is imperative that data used to determine accreditation eligibility be consistent and accurate. Often this requires that a more thorough examination of how data are collected and maintained take place.

Stiles et al. report that the managed care organization they were studying did not allow them access to system data. They also found that there were no shared identifiers in their databases for patients, sites, and providers. Therefore, they could not link MCO data with clinical data gathered at area medical facilities. Consequently, they sampled from those enrolled in the MCO and augmented MCO claims data with chart data gathered from area medical facilities. They focused on cervical cancer screenings and related services. By attaching a new identifier, they linked data from the medical facilities to MCO data. They found a number of discrepancies in the data (e.g., incorrect coding, duplicated systems of data within the organization), which required them to make corrections. This study illustrates the complex nature of organizing data from different sources to inform researchers and stakeholders about the true nature of health care provision. First, knowledge about each system and how they might be linked through the
creation of a unique identifier was essential in collecting the appropriate data. Second, agencies may not collect all the necessary data that is important to answer the research questions and linking various data sets may be necessary. In this study, the MCO had to augment data in order to depict accurate service utilization. Thirdly, accuracy of the data must be established before any conclusions can be made. Despite such concerns, studies continue to be published that use administrative data alone to assess quality indicators for HEDIS (Druss, Miller, Rosenheck, Shih & Bost, 2002).

Reliability of Administrative Data

Ultimately, traditional psychometric properties (e.g., reliability and validity) of administrative files are necessary to establish utility of those files. If reliability and validity cannot be established, researchers will be unable to rely on the results or conclusions based on these data. Kerlinger (1986) defines reliability in three ways: dependability, accuracy, and “the relative absence of errors of measurement in a measuring instrument” (p. 405). Therefore, if a measure produces similar information each time the object or behavior is measured, with little error, the measure or instrument is considered to have good reliability. Validity refers to whether the measure used actually measures the concept of interest. The validity of any measure requires the presence of good reliability.

There are a variety of ways to examine validity; however, when examining administrative data, the concept of reliability has been more important than validity for a variety of reasons. First, the accuracy of each administrative file will influence how confident the researcher can be about the nature of health care delivery and utilization such that, if service information is similar across several administrative files, more
confident conclusions can be made about service provision. Second, it is more difficult to understand the experience of a particular service, or whether that category of service is experienced similarly across patients. Though this is an important issue particularly when trying to examine clinical outcomes, the need for a data set that contains the most accurate picture of service provision must be established before validity can be measured.

In research examining the accuracy of administrative data, reliability has been measured by analyzing error in a variety of ways, including examining error that occurs at times of raw data collection (coding error) and error that occurs when raw data are transformed to a computer file (Melfi et al., 1995; Potvin & Champagne, 1986; Sena & Pashko, 1993). Changes in coding can increase the likelihood of error as well. Garnick et al. (1994) discuss how a change in Medicaid billing codes resulted in a number of previously different surgical procedures being bundled under one code. The Medi-Cal Policy Institute (2001) study found that because the data system did not capture newer codes, information was lost when providers used new and revised diagnostic and procedure codes. Changes like this endanger both the ability to code accurately and the validity of service utilization data.

Several researchers emphasize the importance of establishing reliability among administrative data sets (Calle, Saturno, Parra, Rodenas, Perez, San Eustaquio, & Aguinaga, 2001; Demlo, Campbell, & Brown, 1978; Federspiel, Ray, & Schaffner, 1976; Roos, Roos, Cageorge & Nicol, 1982). Reliability, or agreement, has been examined using a variety of methods and a sample of studies are summarized in Table 1.
Table 1: Review of Reliability Studies

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Type of Administrative Data</th>
<th>Units of Analysis</th>
<th>Rate of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calle et al. (2001)</td>
<td>Charts and Hospital Records</td>
<td>Diagnosis</td>
<td>58.4%</td>
</tr>
<tr>
<td>Dalrymple et al. (1994)</td>
<td>Mental Health Clinic Records and Hospital Records</td>
<td>Date of Discharge, Diagnosis, Unique Identifiers</td>
<td>69.5% - 82.6%, 67.7% - 84.4%, 84% - 97.4%</td>
</tr>
<tr>
<td>Demlo et al. (1978)</td>
<td>Medicare to Hospital Records and Charts</td>
<td>Admission Date, Discharge Date, Diagnosis</td>
<td>99%, 99%, 57% - 65%</td>
</tr>
<tr>
<td>Federspiel et al. (1976)</td>
<td>Medicaid Data, Pharmacy and Physician Claims</td>
<td>Demographic Data</td>
<td>Reported as consistent, but no data provided</td>
</tr>
<tr>
<td>Fisher at al. (1992)</td>
<td>Hospital Records and Abstracted Hospital Records</td>
<td>Diagnosis</td>
<td>58% - 97%</td>
</tr>
<tr>
<td>Parente et al. (1995)</td>
<td>Claims and Charts</td>
<td>Diagnosis</td>
<td>74% - 96%</td>
</tr>
<tr>
<td>Robinson &amp; Tataryn (1997)</td>
<td>Manitoba MIS and Charts</td>
<td>Demographic Data, Admission Date, Discharge Date, Number of Contacts, Diagnosis</td>
<td>46.5% - 100%, 67.1%, 68.8%, 96%, 76.4% - 82.9%</td>
</tr>
<tr>
<td>Roos et al. (1982)</td>
<td>Manitoba Health Services Claims and Hospital Records</td>
<td>Diagnosis</td>
<td>59% - 95%</td>
</tr>
<tr>
<td>Roos et al. (1979)</td>
<td>Manitoba Health Services Claims and Charts</td>
<td>Service Episodes</td>
<td>7% more episodes in claims than records</td>
</tr>
<tr>
<td>Schwartz et al. (1980)</td>
<td>Medicaid Claims and Hospital, Private and Public Clinic Charts</td>
<td>Diagnosis</td>
<td>54%, 71%, 63%</td>
</tr>
<tr>
<td>Studney &amp; Hakstian (1981)</td>
<td>British Columbia Claims Data and Charts</td>
<td>Diagnosis</td>
<td>60%</td>
</tr>
<tr>
<td>Walkup et al. (2000)</td>
<td>Medicaid Claims and MIS</td>
<td>Diagnosis</td>
<td>100%</td>
</tr>
<tr>
<td>Worth &amp; Mytinger (1996)</td>
<td>Hospital Claims and Hospital Records and Physician Records</td>
<td>Diagnosis</td>
<td>57%, 43%</td>
</tr>
</tbody>
</table>
One approach is to compare data recorded in different locations at a similar point in time. Roos et al. (1982) compared different provider claims for the same surgery, that is, compared the diagnosis reported from a surgeon, anesthesiologist and assisting physicians for a treated patient. Another method is to examine data located in charts to data located in a database. This approach is more time consuming and costly, but can find discrepancies more readily when there is reason to believe that there are errors in a database.

Data that are used often in these reliability studies are service episode or diagnostic information. Diagnostic information is typically recorded using 5-digit diagnostic codes based on the International Classification of Diseases, 9th Revision (ICD-9; World Health Organization, 1996). Studies vary in the length of the code used to examine accuracy (e.g., 3 digits vs. 5 digits) for billing purposes. Often, data regarding an outpatient service procedure are coded using the American Medical Association’s Current Procedural Terminology system (CPT) in which procedures are classified into numerical codes. Studies can then examine if a certain procedure was performed on a certain date for a particular patient. CPT codes can be used to examine reliability using the two approaches described above. On the following pages, reliability studies will be reviewed that use Medicare/Medicaid claims data, Canadian systems of data, and private health insurance administrative data.

Early studies examining reliability of Medicare and Medicaid claims predicted some of the phenomena seen in recent literature. Federspiel et al. (1976) found that Medicaid records for the state of Tennessee were reliable and valid by examining payment files and missing information such as sex, age, and race. Very few cases had
missing information (.8 percent). They also compared pharmacy claims to physician claims for the date of service on which a drug would have been prescribed. They found a lag time between pharmacy claims and physician claims, but concluded that the number of claims for both providers was consistent. Even though Federspiel et al. failed to report rates of agreement, Federspiel et al. asserted that Medicaid records are appropriate for data analysis and should be considered before new automated systems of data collection are developed. The evidence for this statement, at that time, was not as compelling as more recent studies, but certainly visionary.

Walkup, Boyer, and Kellerman (2000) examined Medicaid claims data for reliability of diagnostic information as well as the ability to track outpatient mental health follow-up after discharge from a psychiatric hospital. Walkup et al. examined diagnostic information from New York State’s Medicaid Management Information System (MMIS) for persons with a primary diagnosis of a major mental illness who were discharged from a psychiatric hospital during a one-month period in 1992. They identified 69 persons meeting these criteria and found 100% agreement for both primary and secondary diagnoses between Medicaid claims and medical records. They also found that 20% of discharges had immediate outpatient follow-up. The low number of persons with outpatient follow-up as indicated by Medicaid claims was explained by the number of persons who were discharged to practitioners that did not bill Medicaid (e.g., group homes). This study exemplifies the need to be cautious when interpreting claims data in that results are dependent on submitted or recorded data only. These records may not adequately explain a phenomenon because of missing information.
Although Walkup et al. found perfect agreement among diagnoses, this is rarely the case in other data systems. Demlo, Campbell, and Brown (1978) examined a sample of Medicare patients who had been hospitalized for medical reasons. Items selected for reliability analyses included date of admission, discharge, and diagnoses using Medicare records and data abstracted from hospital records and patient charts, if necessary, to determine diagnosis. Demlo et al. found strong agreement for admission/discharge dates (99.2% - 99.7%) but poor agreement for principal diagnosis (57.2% - 65.2%). They proposed that errors began with the provider, in this case the hospital, and strongly recommended improvement in the way data are compiled at the provider level. Specific recommendations included instituting training programs and activities to improve the quality of data and avoidance of coding errors.

Studney and Hakstian (1981) similarly report poor agreement for chart diagnosis compared to diagnosis from claims data. They studied data from the Medical Service Plan of British Columbia. Raters blind to the intent of the study were trained to determine how similar the diagnoses were in chart records and on billing cards where diagnosis and chief complaint were recorded. They found that in 40% of the cases reviewed, raters judged the two forms of data to be dissimilar. Studney and Hakstian cautioned against the use of claims diagnoses without independent validation first.

A case study using the Manitoba Health Service Commission (MHSC) found 7 percent more episodes located in claims than in medical records with minor discrepancies in other analyses (Roos, Nicol, Johnson, & Roos, 1979). They too recommended that some form of validation, such as manual checks of administrative data, is vital to ensure reliable data in data banks.
Roos et al. (1982) examined the reliability of the MHSC database by examining the relationship between surgery episodes and diagnoses (e.g., hysterectomy and cholecystectomy). They used a method to assess reliability by comparing different data points within the MHSC system. In this case, they examined correspondence among surgeons’ health care claims and hospital claims as well as agreement among claims from surgeons, anesthetists, and assisting physicians. Rates of agreement for all analyses were strong (59% - 95%). Claims data and hospital records showed strong reliability for diagnosis (95%) among tested medical conditions for gallbladder patients using the MHSC data bank and hospital records. Inter-rater agreement by two physicians on diagnosis based on chart reviews provided the lowest agreement rate (59%) primarily due to the vague nature of the disease under study (gastrointestinal disorders). They concluded that the MHSC database is a reliable one but they also emphasized the importance of computerized assessment of reliability as an alternative to review of medical records. If different data sources are demonstrated to be consistent with one another, one data source or administrative file can be presumed an accurate representation of services. An important aspect of this study was also the attempt to validate the original diagnosis by examining inter- and intra- physician agreement. Methodologically, to establish validity it is necessary to have reliability. The inter-rater reliability methods of Roos et al. provided a methodology to ascertain whether the diagnoses used were true indicators of the phenomenon researched.

In a similar study, Fisher et al. (1992) examined accuracy of diagnoses. The authors had raters review data from hospital records and, based on their review, determine a diagnosis. These “reabstracted” diagnoses were compared with hospital record diagnoses
from 1977 to 1985. Agreement rates of 58% to 97% were established. Fisher et al. concluded that though there is variability in diagnostic accuracy among different medical conditions, more improvement needs to be made to ensure the diagnosis under study is a reliable and valid one.

Fisher et al. echoed the concern of Roos et al. (1982) about validity of diagnostic information. Fisher et al. discussed the need to establish validity to ensure that the diagnosis coded in fact is present in the patient and that all diagnoses the patient has are in the medical record and administrative data. To date, there are few validation studies that examine the validity of diagnosis.

Robinson and Tataryn (1997) examined the accuracy of data contained in Manitoba's state database (Manitoba Mental Health Management Information System; MHMIS) by comparing computerized data with client charts. Their rationale for examining reliability between these two types of information is that an accurate and complete data set is important and necessary when developing informative health policy or conducting research. Variables examined included demographic data, diagnosis, contact data, and the dates that cases were opened or closed. Analyses found agreement ranged from as low as 46.5% to 100% for demographic data. Percent agreement was 67.1% for exact agreement for open date for cases and 68.9% for close date. For the total number of contacts, agreement was as high as 96% over the course of one year. Agreement for primary diagnosis at the 5-digit level was 76.4% but increased to 82.9% when considering only the first three digits of the 5-digit code. They determined that diagnostic categories fared more poorly because of a lack of updated diagnoses in the MHMIS. They also performed general linear modeling and found no significant
differences in agreement between client charts and MHMIS on demographics, primary diagnosis, or contact data. The authors concluded that MHMIS is a reliable and representative database that can be used as a method to follow patient care and outcome.

Dalrymple, Lahti, Hutchinson, and O'Doherty (1994) argued that it is important to have a reliable data system so that health managers and researchers can follow the progression of any individual or group of people over the course of disease manifestation and examine outcome based on the sequence of care. This study attempted to integrate two systems of data in the province of Ontario, Canada – community mental health centers and inpatient psychiatric care. Dalrymple et al. expressed concern over confidentiality issues when using sensitive data such as mental health treatment episodes and inherent difficulties of linking confidential databases. They reported that hospital data were considered proprietary and were not shared with community health centers. In order to link sets they developed a unique identifier (to maintain confidentiality) consisting of initials, date of birth, and gender and the new identifier was used to merge data sets from hospital and community settings across two districts within one province.

Agreement between data sets was explored using percent agreement calculations. The date of the most recent hospital discharge and patient diagnosis were found to have the lowest agreement (69.5% and 67.7% respectively) of items studied, although agreement of hospital discharge rose to 82.6% when a time period of plus or minus one week was allowed. When diagnoses were clustered into categories based on similarity, agreement increased to 84.4%. In a follow-up (2 years later), the authors again examined agreement among the community and hospital settings. The different data sets were examined by analyzing reliability of variables comprising the unique identifiers in each set. Data were
found to be more accurate at hospitals than community centers when compared to clinical
records. Community error (14%) was five times larger than hospital error (2.6%) on the
accuracy of unique identifiers. The authors suggested that workers in hospitals are better
trained than those in community centers and have fewer people entering data, minimizing
the amount of error in recording. They conclude that as more health care reform takes
place, the importance of having reliable information across settings will increase. Shared
definitions of service must be developed now so that policy makers can make informed
decisions in the future. They argue that researchers cannot determine accurate treatment
histories until linked and reliable databases are established.

A more recent study conducted in the UK (Fine et al., 2003) found hospital medical
records to be largely discrepant with death records. The authors compared a hospital
database shared by 10 cardiac surgery centers to data extracted from charts. Authors were
interested in creating a risk-adjusted mortality score for preoperative patients before
cardiac surgery. Researchers were concerned that the data systems could not support the
development of a score because of missing data. They found the average percent of
missing information from each hospital was 24.9% whereas only 1.2% of data were
missing from charts. The authors then implemented a system to validate and monitor the
data system including feedback on performance of each surgery center. Rates of missing
information for the hospital database were significantly reduced to 9.3% but charts
continued to have significantly less missing information (4.0%). Findings suggested that
without an examination of the accuracy and completeness of administrative data, chart
information may be the least error-laden source of data for outcome analysis.
Recently, the Health Care Financing Administration developed the National Claims History data set to track Medicare Part A and B claims. In doing so, they created a set of data that can be used for examinations of cost, quality, and effectiveness. Parente, Weiner, Garnick, Richards, Fowles, Lawthers, Chandler and Palmer (1995) reviewed this data set and describe the development and nature of the database. They also include a short description of the reliability analyses used to examine the accuracy of the data set. The project compared diagnostic categories of the computerized claims file with chart data and found a range of 74% to 96% agreement on the presence or absence of 6 diagnostic categories with Kappa values of .27 to .67. An additional chart review was conducted to explore disagreements and once clinical records were reviewed, nurses, who served as raters, were able to determine that the diagnosis found in claims was being treated, but the chart may not have included the diagnosis when the service was provided. They concluded that these types of data can reach good concordance with medical charts, thus supporting the notion that these data can be used to examine continuity of care provided by health care providers.

Kosecoff et al. (1987) found Medicare Part B claims data to be a useful source when evaluating appropriate care in areas of coronary angiography, carotid endarterectomy, and upper gastrointestinal tract endoscopy. They reviewed medical records and Medicare claims in five states and found that of the 4,988 claims only 4.8% were found to have errors. Kosecoff et al. concluded that Medicare Part B claims are accurate sources of information and can be used for health care services research. They asserted that medical record review is also a valid approach when examining cost and appropriateness of medical procedures. However, the use of computerized claims data, if
found to be as informative as medical record review, is easier and less costly both in time and money.

Worth and Mytinger (1996) examined reliability of hospital claims and physician-generated claims submitted from July 1986 to June 1987 to a private insurance company as part of a goal to establish a large archive of data within this insurance company. They were concerned about the accuracy of diagnostic data contained within this set of data so they conducted a study to examine the accuracy of this administrative data. A stratified random sample (by patient age and gender) was taken from 169,586 claims, producing a sample of 186 claims. Hospitals and physicians were asked to report diagnostic information based on chart data for the selected set of claims. These data were used to determine data accuracy. At the hospital level, claims had different diagnoses compared to hospital report in 4% of the cases; however, errors were typographical in nature. Thirty-eight percent of the claims had “unprecise” codes, meaning the primary diagnosis was not the same in both claims and hospital report, but the correct code was listed elsewhere. In these cases, chronic conditions were typically present and a variety of codes were listed rather than a single diagnosis.

Errors in physician claims included typographical errors (12%), confusion of medical terminology (8%), coding of infectious disease rather than coding of exposure to that disease (17%), and “unprecise” codes (19%). Only 43% had identical primary diagnostic codes in both databases. To correct errors, a system based on a diagnostic algorithm was developed for those diseases (diagnostic codes) that were unlikely to be seen and treated. The algorithm used age and gender to flag diagnoses that should not be seen given those two factors. When suspected errors were found by this system,
physicians were sent a letter and asked to correct/verify the diagnosis. After these corrective actions, 94% of physician claims were shown to be accurate in diagnostic coding. The results suggest that within a computerized system, regular checks of data accuracy can lead to improvement if appropriate protocols are developed, but if these are not in place, administrative data may not be dependable.

As mentioned previously, comparing chart data and claims data is a popular method of establishing reliability and there has been a long history of examination of diagnostic information in charts and claims data. Not all studies have recommended the use of administrative data, not because of coding errors, but because of possible intentional error. Schwartz et al. (1980) used chart data and Medicaid claims data from 29 free-standing psychiatric clinics, a sample of private practitioners serving Medicaid patients, and 6 outpatient psychiatric clinics in general hospitals to evaluate New York Medicaid claims data. They found that 29% of chart diagnoses of private practitioners, 37% of chart diagnoses in the freestanding outpatient clinics, and 46% of diagnoses from outpatient clinics of general hospitals did not match Medicaid claims. It was hypothesized that the higher percentage of error in the outpatient hospital setting was a result of staff giving less severe diagnoses to patients. There was consistent discrepancy between the outpatient diagnosis and diagnosis given while the person was an inpatient. Such errors may be even greater in environments where the illness reported may influence reimbursement. For example, in certain acute care settings, particular psychiatric diagnoses will not qualify the inpatient stay as reimbursable so practitioners may report a more severe or inaccurate diagnosis in order to get paid. Reporter bias in the original source of data (e.g. medical charts) may directly affect the accuracy and validity
of administrative claims, jeopardizing the quality of research and policy recommendations based on claims data (see also Sena & Pashko, 1993).

Summary of Reliability Studies

The studies reviewed above show a range of reliability, or rates of agreement, among different administrative data sets. The range of agreement for diagnostic accuracy of clinical charts compared to Medicaid claims data can be quite large when comparing studies (54% agreement, Schwartz et al., 1980, to 100%, Walkup, Boyer, & Kellerman, 2000). Agreement of service provision and service dates also vary greatly (50% - 99.7%; Robinson & Tataryn, 1997) depending on how large the time period is around the date of service examined. A review of the possible sources of error and an examination of past studies examining agreement among different data sets describe conditions that warrant concern when relying exclusively on administrative data to understand epidemiological or service information about a given population.

Applicability of Administrative Data

It is clear that the research examining the reliability of diagnostic information using claims data is inconclusive. Additionally, researchers suggest the importance of checking data regularly. Worth and Myrtinger (1996) showed that once errors are corrected, more reliable estimates can be made. Even though there are limitations to using such data, researchers have begun to examine mental health care outcomes, cost, utilization, and the effects of programmatic changes with claims data. Several examples of such use of administrative data are reviewed below.

Leslie and Rosenheck (1999) performed an analysis of pre-existing data to explore whether medical services were being used as a replacement for mental health services.
This study examined the patterns of privately insured adult inpatients between 1993 and 1995 and found inpatient costs decreased over the period as a result of a lower number of inpatient days used. Those with a diagnosis of mild/moderate depression showed the largest decrease in utilization of inpatient stays whereas individuals with schizophrenia showed the smallest reduction in use of inpatient days. They found that managed care plans had lower costs relative to other types of plans studied; however, utilization rates were similar across the insurance categories studied and the authors concluded that plans were not substituting medical services for mental health services in order to cut costs.

Gilbody, House, and Sheldon (2002) reviewed mental health outcomes research of published studies using administrative data. They concluded that if clinical databases are used for outcomes research a great deal of time and money must be invested to ensure accurate and valid data. Ultimately, they suggest that findings from current studies based on pre-existing databases should not be considered on par with data collected from a randomized clinical trial nor should it be used to inform mental health policy without consideration of the many limitations that accompany administrative data.

Evaluating changes in policy can be examined using administrative data even when the data contain a degree of inconsistency. Weiner, Powe, Steinwachs, and Dent (1990) examined changes in utilization after implementation of managed care of Medicaid benefits in Iowa. They used a statewide claims database to examine rates of substance abuse treatment before and after the change in program management over a three-year period. They concluded that managed care did not affect accessibility of substance abuse treatment. Even this conclusion was established with an admitted lack of valid diagnostic information and with the potential for admission rates to be exaggerated due to changes
in boundaries of the region under managed care. In establishing the conclusion, the
authors argued that error is expected to be distributed randomly and not affect results of
the variables under study. This analysis demonstrates the flexibility of administrative data
and ability to compare provider patterns among different mental health care plans.

Another study exploring the effects of managed mental health and substance abuse
treatment using state claims data found total expenditures decreasing and the number of
users in outpatient clinics increasing during the first year of managed care
implementation (Beinecke & Perlman, 1997). By the second year, costs, utilization, and
the number of persons on medication prescribed at outpatient clinics had increased along
with the number of outpatient users but inpatient stays declined. Clearly there was a shift
in treatment from inpatient to outpatient centers. The finding of increased cost was
unexpected as was the change in prescribing of medications for these patients. An
important advantage to administrative data is that trends can be examined and followed
over time to track changes in patients, provision of care, and the costs of service.

Johnson and McFarland (1994) compared treatment rates among persons with
schizophrenia and bipolar disorders in a health maintenance organization (HMO) to data
from the Epidemiological Catchment Area (ECA) survey. They used five different data
sets from the HMO including pharmacy data, encounter data, medical data, eligibility
data, and utilization review data. They validated diagnostic data by reviewing charts to
ensure criteria for the diagnoses were met. Of those with schizophrenia or bipolar
disorder, data in charts were found to correspond to administrative data. When comparing
treatment rates it was found that the HMO treatment rates were lower than those reported
in the ECA study. Rather than concluding that the HMO was restricting access to
treatment, the authors cautioned that the inclusionary criteria may be different for this HMO relative to those observed in the ECA study and concluded that further studies need to be conducted to understand the course of treatment and illness of persons in HMOs.

**Summary of Applied Administrative Data Studies**

It is clear that administrative data are being used to inform researchers of the states of health and mental health care even though evidence exists suggesting that such analyses should proceed with great caution if at all. Though many researchers rely on these data, there are others who suggest that the use of administrative data may be premature and the accuracy of these large sets of data must be established before they are utilized. Reliability analyses provide data about how consistent these data are, but provide no information about the validity of data. Despite concerns around the use of administrative data, researchers are using these data to measure a variety of mental health concerns including outcomes (Gilbody et al., 2002; Melfi & Croghan, 1999), cost shifting (Leslie & Rosenheck, 1999), quality of care (Lohr, 1990; Weiner et al., 1990), the effects of managed care (Beinecke & Perlman, 1997; Johnson & McFarland, 1994; Weiner at al., 1990) and substance abuse utilization (Garnick, Horgan, Hendricks, & Comstock, 1996).

**Rationale and Research Questions**

Chart data, Medicaid claims data, and state information systems are all rich sources of data, but sources containing the most accurate information are difficult to establish without examining the agreement among all three. Kashner (1998) refers to chart data as the "gold standard" of data sets; however, there has been no definitive set of studies to suggest this. Chart data could be susceptible to recording requirements of the agency or accrediting body and not indicative of actual services provided. Similarly, claims data
may not reflect true service utilization because many organizations are aware of the restrictions reimbursing agencies place on services and it is possible that many services may never be billed or filed because of an anticipated rejection of the claim. More important, if a researcher were to use one data source alone, how different would service utilization appear when compared to another source? There has been no study to date that examines three sets of data simultaneously for agreement.

Triangulation, that is, an analysis of three systems of data to determine the most accurate system, will be applied to Medicaid claims data, Hawaii's behavioral health information system (MFASIS), and chart data from Hawai'i state's community mental health centers (CMHCs). These data files will be compared to explore the nature of relationships among these different sources of data. In Hawai'i, Medicaid claims data are maintained by the Hawai'i Department of Human Services. The Department of Health, Adult Mental Health Division (AMHD) maintains MFASIS, a computerized database that contains records of services to adults with serious mental illness provided by CMHCs. CMHCs have responsibility for tracking services and patient information in clinical charts, submitting bills to Medicaid, and entering data into the MFASIS system. It is hypothesized that chart data will include more services and the agreement between charts and MFASIS will be higher than agreement between charts and Medicaid, and MFASIS and Medicaid for service data. Chart data theoretically should not be influenced by billing restrictions and therefore should contain all services provided rather than only a subset of billable services. It is also hypothesized that agreement rates of psychiatric diagnosis will be greater using only the first three digits of the diagnostic code when compared to agreement rates using four and five digits across all three sources. Service
utilization patterns based on each set of data will be computed and compared to examine the different service portraits painted by different data sources.

Listed below are the questions this study proposes to answer, and what results are expected.

1. What is the agreement between chart data and MFASIS records?
   a. It is hypothesized that rates of agreement between chart data and MFASIS will be higher than other comparisons.

2. What is the rate of agreement between chart data and Medicaid claims data?
   a. It is hypothesized that the rate of agreement between chart data and Medicaid claims will be lower than other comparisons.

3. What is the rate of agreement between MFASIS records and Medicaid claims data?
   a. It is hypothesized that the rate of agreement between MFASIS and Medicaid will be better than that found for charts and Medicaid.

4. What is the rate of agreement for primary diagnosis?
   a. It is hypothesized that agreement rates found using only the first three digits of a 5 digit code will be better than rates using four and five digits.

5. What are the patterns of service utilization using each set of administrative data?
   a. It is hypothesized that services recorded in charts will significantly outnumber those recorded in MFASIS or Medicaid claims. It is also hypothesized that MFASIS will contain fewer recorded services than charts, but more than Medicaid claims.
CHAPTER THREE

Methods

Design

This study involved a quasi-experimental design utilizing archival data. Archival data were collected in a study by the University of Hawai‘i, in collaboration with AMHD and the State Medicaid Agency, funded by the Substance Abuse and Mental Health Services Administration (SAMHSA). This study, referred to as the “Preliminary Study,” included the collection of data from local databases and service data from state providers of mental health care (e.g., CMHCs). A review by the university’s Institutional Review Board allowed for the use of data collected from the preliminary study as long as identifying information (e.g., social security number and name) was eliminated once data were ready for analysis.

Preliminary Study

Five-hundred sixty-three persons with a serious mental illness using mental health services in the state of Hawai‘i participated in a study that examined satisfaction with mental health services. Participation included a one-hour structured interview to collect information regarding satisfaction with services and current health and mental health status. A serious mental illness was defined by DSM-IV codes 295-298 inclusive which represent the following diagnostic categories: Schizophrenia and other Psychotic Disorders, Bi-Polar Disorders, and Major Depressive Disorders. Persons enrolled in the study were receiving Medicaid and mental health services in a fee-for-service (FFS) or managed care (MC) environment. Participants were assessed by interviews on current mental health and health status, satisfaction, quality of life, and service utilization at
baseline and 6-month follow-up (see Wylie & Nathan, 1999). In addition, Medicaid claims data were collected for participants as was service data retained in the AMHD's management information system (MFASIS).

Chart reviews were also conducted for participants who received services at a CMHC. Three hundred sixty of the original 563 participants (64%) were identified as having received a service at a CMHC. Of those identified, 243 charts were reviewed fully, 10 were partially reviewed, one chart could not be located, one was unavailable due to client death, and 105 charts had no services recorded within a year previous to the study or had been closed by the CMHC due to inactivity. Participants with completed chart reviews were used as participants in this study.

Sources of Data

Data from three sources were used for these analyses. All data sources were collected and archived under the preliminary study. Data included: 1) service information extracted from charts; 2) service data from an automated system used by CMHCs for reporting purposes; and 3) Medicaid claims data from the State Medicaid Agency for mental health services provided by CMHCs.

Chart Review Data

Information collected through a review of CMHC charts were entered into a Microsoft Access database and linked to demographic information collected from the structured interview. Demographic information including name, study identification number, and social security number (SSN) were used to request data from other data sources.
A form was created to collect data from charts and included service location, type and date of service, diagnoses listed in client charts and client record number. The service categories include: provision of prescription, drug administration, medical check, blood draw, therapy/case management, group therapy, family therapy, evaluation/assessment, phone contact with client, treatment planning, attendance at Clubhouse, provision of a referral, paperwork, biopsychosocial rehabilitation and an “other” category in which the reviewers could place any additional services not captured by other categories (see Appendix A). Therapy and case management were coded as a single category because CMHCs in Hawai‘i consider supportive case management a preferred treatment approach for those with a serious mental illness. Often, the same service is called therapy when the person providing the service holds a Ph.D. or M.S.W.

Charts were reviewed for a period of one year prior to entrance in the preliminary study. The study began in November 1997 and baseline interviews were completed by November 1998. Therefore, chart data covered services provided from November 1996 through November 1998 with start and end dates for each participant depending on the date of entry into the study.

Charts were reviewed at the CMHC from which the person received services during the time period of interest. Charts were reviewed under supervision and in cooperation with the CMHC administrators in an area provided for the chart review. Each chart review took approximately one hour to complete. Data were then entered into a database secured in a locked office and password-protected to ensure confidentiality. A copy of that database was used for these analyses.
Diagnostic information was collected in addition to service information. The primary diagnosis reported in charts on the date closest to the baseline interview was recorded.

Medicaid Claims

As part of the preliminary study, Medicaid claims data were collected for participants whose clinical charts were reviewed and others involved in the larger study. The Medicaid claims database contained billing information about services that were billed to Medicaid by CMHCs. Information contained in this database included name, social security number, procedure code, provider code, transaction number and additional codes more relevant to the claim than these analyses. This study used submitted rather than adjudicated claims. For example, the Medicaid database also included denied claims, information related to appeals on denied claims and records of paid claims. In addition, only submitted claims that were coded with a psychiatric diagnosis were included.

Prior to submission to the State Medicaid Agency, claims data were completed by the AMHD clerks at CMHCs across the state. A database, separate from MFASIS, was used to generate claim forms compatible with those required by Medicaid. Services that were considered “reimbursable” were entered into the database at each site from charge and encounter slips completed by CMHC providers. Once data were entered into the database a claim form would be generated. The forms were mailed to Medicaid for payment.
Diagnostic information was attached to each claim. In order to conduct agreement analyses for diagnoses, the modal psychiatric diagnosis was identified for each participant and used for analyses and descriptive purposes.

*MFASIS*

The AMHD MFASIS was a computerized database that recorded all services at CMHCs regardless of billing status. Data-entry staff at CMHCs entered all data, regardless of billing potential, from a charge and encounter slip completed by the provider of the service at the CMHC. MFASIS was used for reporting purposes by the Research and Evaluation Unit of AMHD as well as for record-keeping purposes at CMHCs.

MFASIS service records for the 243 participants were provided by the Research and Evaluation unit of the AMHD for the full period between 1996 and 1998. The data used for each participant corresponded to the time period used for chart reviews.

Diagnostic information from MFASIS was requested separately. Diagnostic information was provided in a table with multiple diagnoses recorded at different dates. The diagnosis that was identified as “primary” and recorded at the date closest to but before entrance into the preliminary study was used for this study. This was consistent with instructions to chart reviewers who recorded the most recent diagnosis in the chart reviews. Primary diagnosis was merged into datasets using the participant’s client record number assigned by AMHD.
Procedures

Crosswalk Table

Each set of data contained different codes to represent services; therefore it was necessary to create a procedure to link these data with a coding system. The coding system, or crosswalk table, created common numerical codes for each service category under study. This allowed for the categorization of a group of service procedures in one database to be compared with a group of similar services in another.

Medicaid claims use CPT codes to identify services. These codes are typically five digits and specify the type of service (e.g., evaluation, medical evaluation, therapy) and the provider (e.g., physician, social worker) and can be modified to designate length of service. The coding approach allows for a more detailed analysis of service utilization; however, the other data sets in this study were less specific, so CPT codes were collapsed into categories more consistent with the other data systems. Modifiers were eliminated and only the core code was used for analyses.

MFASIS data, though less complex than CPT codes, allowed for some designation of the type of service provider (i.e., MD vs. non-MD) and reflected subtle differences in services provided. Charge tags, or billing tags, used by CMHCs show both CPT codes matched with MFASIS codes for many services. Matched codes located on charge tags were used to assist in the development of the crosswalk table.

The translational coding system was developed to maximize simplicity. The type of provider was not included in category definitions. The length of service was not considered either. This made it a challenge to create a coding system that allowed for
services to be matched and compared. Information located in a more complex coding system would be lost when forced into a simpler strategy.

First, categories from chart data were selected based on billing status since claims data would not contain non-billable services. A total of six categories were selected: 1) drug administration; 2) evaluation/assessment; 3) family therapy; 4) group therapy; 5) medication evaluation; and 6) therapy/case management. Then, a review of 48 MFASIS codes and definitions were completed. The services reflecting components of the chart data categories were grouped within each of the six categories. Finally, a review of psychiatric CPT codes from 1997 and 1998 was conducted. A crosswalk table that identified the chart review category, CPT codes, and equivalent MFASIS codes was created.

A panel of experts in the area of Hawai‘i’s mental health services reviewed the crosswalk table to validate the rationale used to create and categorize codes. Seven persons were selected. The panel included two people with Ph.D.s in psychology from the research and evaluation unit of the AMHD; one person with a M.A. in psychology who previously directed a Medicaid mental health clinic and was the AMHD information systems supervisor; a data analyst with expertise in MFASIS data; a data analyst with expertise in Medicaid claims; a registered nurse with experience in mental health systems in Hawai‘i; and a person with five years experience as a CMHC billing clerk. The panel reviewed the codes and crosswalk table and provided feedback that categorized some of the therapy/case management codes into medication evaluation codes. Appendix B shows the final crosswalk table used for analyses and is based on panel input and suggestions.
Eligibility

Individuals receiving Medicaid often “cycle” on and off Medicaid because of Medicaid-imposed eligibility restrictions (e.g., income, employment). Charts were reviewed without knowledge of Medicaid eligibility; therefore it was important to determine each person’s eligibility status during the study period to ensure that included service data examined from participants who were Medicaid eligible. If not, there would be a chance that a service may not have been submitted because of the knowledge that a participant was not Medicaid eligible and thus no claim would be present in Medicaid files. For each participant, an eligibility table was constructed from Medicaid eligibility files to determine which months the participant was Medicaid eligible. As a result, only those months of service in which participants were eligible were examined in this study and service comparisons only occurred across those months.

Data Cleaning and Preparation

Each dataset underwent several processes to create three sets of service data for analyses. Chart data included services that occurred between the date the participant was interviewed and one year prior to this interview. This year was considered the service period and was unique to each participant. In addition, service period was further defined to include only those months in which the participant was Medicaid eligible.

First, services in MFASIS and Medicaid claims were recoded to be consistent with the crosswalk table. Services not included in the crosswalk table were excluded. Then, using eligibility tables, those services that occurred in a month the participant was Medicaid eligible and whose date of the service fell within the one year service period
were kept for reliability analyses. Table 2 represents the data cleaning and preparation process.
Table 2: Data Cleaning and Preparation

<table>
<thead>
<tr>
<th>CHARTS</th>
<th>MFASIS</th>
<th>MEDICAID CLAIMS</th>
</tr>
</thead>
</table>
| Services extracted from Chart Reviews of all CMHCs in Hawaii. Reviewed Services for a 12 month period. Varying service periods were reviewed due to different dates of entry into the original study.  
- 15 original service categories  
- 13,846 services identified  
- n = 243 | Using chart record numbers and social security numbers from 243 participants in chart review, selected service data from MFASIS records between 1996 to 1998. Due to varying service periods of chart review, selected all services between these two years to ensure collection of all possible services for each participant.  
- 51 original service categories  
- 54,626 services identified  
- n = 219 | Using chart record numbers and social security number from 243 participants in chart review, selected service data from Medicaid claims files between 1996 and 1998. Due to varying service periods of chart review, selected all services between these two years to ensure collection of all possible services for each participant.  
- 1,015 original service categories (includes non-psychiatric codes)  
- 71,915 services identified  
- n = 221 |

Selected service categories that were included in crosswalk table. Six service categories selected. Outcome: 7,226 services identified  
n = 238 | Recode 11 MFASIS service codes into 6 service categories to match chart data categories using the crosswalk table. Select only those service categories for analysis. Outcome: 12,883 services identified  
n = 218 | Recode CPT codes into 6 service categories to match chart data categories using the crosswalk table. Select only those service categories for analysis. Excluded any services with negative charges and duplicate transactions (i.e. retained original submitted claims only). Outcome: 8,774 services identified  
n = 202 |

Selected only those service episodes that were provided during months of Medicaid eligibility. Outcome: 5,716 service episodes identified.  
n = 201 | Select only those service episodes that were provided in Medicaid-eligible months. Only those months that occurred within the one year service period were used. Outcome: 4,072 service episodes identified.  
n = 177 | Selected only those service episodes that were provided in Medicaid-eligible months. Only those months that occurred within the one year service period were used. Outcome: 3,986 service episodes identified  
n = 186 |

Retention rate of participants: 82.7% | Retention rate of participants: 80.8% | Retention rate of participants: 84.2% |
Agreement Analyses

To estimate reliability across data sets, percent agreement was calculated. That is, the number of agreements that a particular service was delivered at a specific time was divided by the number of possible agreements. First, each dataset was organized by service categories so that the rates of agreement could be compared within service categories. Each service encounter was given a unique code using the identification number of the participant, date of service and episode type.

Typically, the denominator used for percent agreement is comprised of the total number of observations, generally specified before the observations occur. Due to the nature of this data, the total number of observations could not be pre-determined. A substitute number, such as the number of days in a year, could not be used as the denominator because of the possibility that a service could be provided multiple times in a day. Therefore, the number of possible agreements was calculated as follows:

\[
\frac{(# \text{ of observations in data set } A) + (# \text{ of observations in data set } B)}{- # \text{ of agreements}}
\]

This approach eliminates the number already observed to be in agreement in order to eliminate duplication between the data sets.

Rates of agreement for those that did not receive a service were also calculated. This analysis used the entire sample (243) and calculated how many persons were identified by both data sources to have not received a particular service during the study period. The agreement index was comprised of the number of times two data sets concurred that a service was absent for each person divided by the total number of participants in the study. This agreement rate represents the percent of the original sample
that two data sets agree did not receive a service within each category. The rate is a good indicator of whether data sources can identify persons not using services.

In addition to using the agreement analysis presented above, an additional approach examined general agreement among data sources. The approach described earlier looks for exact matches, that is, a match occurring when a service episode in one set of data match an episode in another set of data by participant identification number, date of service and type of service. As mentioned previously, the accuracy of service dates vary (Robinson & Tataryn, 1997) so an alternative approach allowing some variability in the dates of service delivery was developed.

Rather than matching on specific service date, only the month in which the service occurred was used. Rather than looking at individual service dates across the year, individual months in which the participant used that particular service were examined in agreement analyses. A code was created that identified each month that a particular individual used a service. This allowed comparisons to be made among users in each month. Such an approach controls for some variability in exact dates of service and allows for an examination as to whether or not the databases capture the same individuals in the same months. The approach provides a snapshot of how many different people utilized the service in each month. In other words, the index looked at the “penetration rate” of services across users.

This approach borrowed from the idea of penetration rates. Penetration rates examine service use in a more practical and realistic approach by calculating how often one accesses a particular service in a given time relative to all potential service users. A penetration rate is the proportion of actual users to the rate of eligible users of a service.
within a specified period of time and is used as an indicator of access to health services (Stiles, Boothroyd, Snyder & Zong, 2002). The calculation developed for this research borrows the notion of examining the use a particular service within a specific period of time (i.e. a month) and compares the “penetration” of this service for each month by the user compared to other sources of data. Essentially, rather than using a specific date as is used in the exact match analysis, agreement of service use occurred if a participant used the service at least once in the month. In these analyses, the term penetration is used to differentiate these agreement analyses from precise matching and not in the traditional sense of the word.

This approach takes a more liberal view of agreement across data sources. It was expected that rates of agreement would be better using these analyses because less strict matching had to occur. Analyses were conducted across each data set using the above procedures. Results are presented by the six service categories discussed above.

*Service Utilization*

Service utilization was examined by comparing the average number of episodes over the study period in each service category. A comparison of means using analysis of variance (ANOVA) and Tukey post-hoc tests were used to identify significant differences between service categories.
CHAPTER FOUR

Results

The findings of this study are presented in four sections. The first section describes the results of data cleaning and preparation. The second section presents the agreement analyses for service data. Each dataset is compared to each of the other two, resulting in three agreement analyses per service category. In section three, agreement rates for diagnostic codes are discussed. The last section presents information regarding service utilization patterns among the three data sets.

Data Cleaning and Preparation

Of the 243 participants with chart review data, only 219 and 221 were in original MFASIS and Medicaid claims files, respectively. Fewer participants were left for analyses after data were cleaned. Retention rates of participants were very similar across data sets (as shown in Table 2). After the initial step of cleaning, there was a greater loss of service episodes for MFASIS and Medicaid data primarily due to the nature in which data were collected. More service data were eliminated for MFASIS and Medicaid claims because both covered a 2-year period rather than a one year period, like chart data.

Demographic characteristics of the data sets were evaluated with chi-square analyses and one-way analysis of variance (ANOVA) to examine if the loss of participants from data cleaning procedures affected the characteristics of participants in each data set. Table 3 and Table 4 present the demographic and diagnostic characteristics of the sample among the three data sets. When comparing original files to cleaned data files, no significant differences were found between gender in the chart data set, $\chi^2(1, \ N = 444) = .411, p = .521$, MFASIS data set, $\chi^2(1, \ N = 396) = .042, p = .838$, or the
Medicaid claims $\chi^2(1, \ N=407) = .055, p = .814$ (see Table 3). When comparing original files to cleaned data files, no significant differences were found between diagnostic category in charts, $\chi^2(10, \ N=444) = 2.876, p = .984$, MFASIS data, $\chi^2(7, \ N=371) = 1.909, p = .976$, or Medicaid claims data $\chi^2(6, \ N=337) = .001, p = 1.00$ (see Table 3).

Participants from the preliminary study were identified as having a diagnostic code between 295 - 298; however, other diagnostic categories were represented in files. Persons with severe mental illness often have co-morbid conditions such as substance use disorders that may have been recorded as the primary diagnosis and would explain this discrepancy.

The difference in age between original and final data was not significant in the chart data, $F(1, 443) = .446, p = .505$, MFASIS data, $F(1, 395) = .590, p = .443$, or the Medicaid Claims data, $F(1, 406) = .631, p = .427$ (see Table 4). Interestingly, across all three sets of data, two-thirds of persons were represented in the original sample, that is, of the original 243, 162 appeared at least once in all three data sources.

In addition, the demographic characteristics of the data sets were compared with each other to determine if there were significant differences between the groups. No significant differences were found in the age of participants between the three data sets $F(2, 682) = .097, p = .883$. No significant differences were found in gender between the three data sets $\chi^2(2, N=564) = .011, p = .994$. No significant differences were found in diagnostic group between the three data sets $\chi^2(22, N=554) = 15.677, p = .832$.

Because of the fact that persons can cycle in and out of eligibility for Medicaid as discussed earlier, it was important to examine the frequency of this occurring. In fact, this
happened rarely. No significant differences were found in number of eligible months between the three data sets \( F(2, 561) = 1.465, p = .232 \). Mean number of eligible months for chart data was 12.00 (SD = 2.23; range = 1 to 13), for MFASIS data was 12.08 (SD = 1.94; range = 1 to 13), and for Medicaid claims data was 12.33 (SD = 1.61; range = 3 to 13). The greatest permissible number of months of eligibility in this study was 13 (i.e. the month of the initial interview and the preceding 12). The mean number of eligible months was 12 or greater for services recorded across data sources, although some participants were Medicaid eligible for as few as a single month of the study.

Table 3: Gender and Diagnostic Category of Participants by Source

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Chart Original</th>
<th>Chart Final</th>
<th>MFASIS Original</th>
<th>MFASIS Final</th>
<th>Medicaid Original</th>
<th>Medicaid Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>128</td>
<td>112</td>
<td>119</td>
<td>98</td>
<td>121</td>
<td>104</td>
</tr>
<tr>
<td>Female</td>
<td>115</td>
<td>89</td>
<td>100</td>
<td>79</td>
<td>100</td>
<td>82</td>
</tr>
<tr>
<td>Total</td>
<td>243</td>
<td>201</td>
<td>219</td>
<td>177</td>
<td>221</td>
<td>186</td>
</tr>
<tr>
<td>Diagnostic Category and Disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>292 (Withdrawal from Substance Abuse)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>293 (Delirium)</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>294 (Cognitive d/o)</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>295 (Schizophrenia)</td>
<td>149</td>
<td>134</td>
<td>141</td>
<td>122</td>
<td>132</td>
<td>128</td>
</tr>
<tr>
<td>296 (Bi-polar d/o)</td>
<td>61</td>
<td>40</td>
<td>46</td>
<td>31</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>297 (Delusional d/o)</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>298 (Brief Psychotic d/o)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>300 (Anxiety d/o)</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>301 (Paranoid Personality)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>304 (Opiate Abuse)</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>305 (Alcohol Abuse)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>309 (Adjustment d/o)</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>311 (Depressive d/o NOS)</td>
<td>7</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>243</td>
<td>201</td>
<td>219</td>
<td>177</td>
<td>181</td>
<td>169</td>
</tr>
</tbody>
</table>
Table 4: Mean Age of Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Chart</th>
<th>MFASIS</th>
<th>Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Final</td>
<td>Original</td>
</tr>
<tr>
<td>Mean Age</td>
<td>43.05</td>
<td>43.69</td>
<td>42.98</td>
</tr>
<tr>
<td></td>
<td>SD=10.06</td>
<td>SD=10.21</td>
<td>SD=10.18</td>
</tr>
<tr>
<td></td>
<td>(n = 243)</td>
<td>(n = 201)</td>
<td>(n = 219)</td>
</tr>
</tbody>
</table>

Agreement Analyses

Drug Administration

Analysis of chart review records found that there were 690 episodes of drug administration (n = 45) while Medicaid claims showed 481 episodes (n = 44). MFASIS data suggested a higher number of episodes than the other two data sets (958; n = 43). A closer look at individual data found two outliers in which 104 and 204 drug administration services had been provided to these two participants according to MFASIS records. A check on dates found that each episode was provided on a different day suggesting that duplication of the episode had not occurred. An individual analysis of the files found no recorded drug administrations in charts or Medicaid claims for the participant with 104 episodes. Chart data and Medicaid data suggested that 17 and 14 administrations, respectively, had been provided to the participant with 204 service episodes in the time period of interest. In fact, 14 of the 17 episodes recorded in charts and 12 of the 14 recorded in Medicaid matched with MFASIS files.

Two analyses using exact matching were conducted – one with the outliers included and one without (see Table 5). The exact matches with all data found the highest percent of agreement occurred between chart data and MFASIS (28.3%); however, the agreement between charts and Medicaid claims was almost the same (27.9%) with slightly lower agreement between Medicaid claims data and MFASIS (21%). Analyses
with outliers excluded found an exact agreement rate of 27.0% between MFASIS (650 episodes, \( n = 41 \)) and Medicaid (467, \( n = 42 \)), a 6% increase. The agreement between chart data (673, \( n = 44 \)) and MFASIS increased almost 8% to 35.8%. Rates of agreement for persons in the original sample that did not receive any drug administration services across comparisons (non-users) were much higher (see Table 5), approximately 76% across comparisons.

The number of total penetration episodes for chart data was 358 (\( n = 45 \)) with similar numbers for MFASIS (347; \( n = 43 \)) and Medicaid (288; \( n = 44 \)). Agreement rates using penetration rates improved agreement with approximately half of all episodes matching for each comparison.

Table 5: Agreement Rates of Drug Administration

<table>
<thead>
<tr>
<th>Drug Administration</th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td># matched episodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– exact match</td>
<td>256 (( n = 30 ))</td>
<td>364 (( n = 32 ))</td>
<td>250 (( n = 27 ))</td>
</tr>
<tr>
<td>Percent Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– exact match</td>
<td>27.9%</td>
<td>28.3%</td>
<td>21.0%</td>
</tr>
<tr>
<td># matched episodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– exact match</td>
<td>244 (( n = 30 ))</td>
<td>350 (( n = 31 ))</td>
<td>238 (( n = 26 ))</td>
</tr>
<tr>
<td>Percent Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– exact match</td>
<td>27.2%</td>
<td>35.9%</td>
<td>27.0%</td>
</tr>
<tr>
<td># matched episodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Penetration</td>
<td>224 (( n = 31 ))</td>
<td>245 (( n = 30 ))</td>
<td>211 (( n = 28 ))</td>
</tr>
<tr>
<td>Percent Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Penetration</td>
<td>53.1%</td>
<td>53.3%</td>
<td>49.8%</td>
</tr>
<tr>
<td>Percent Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– non-users</td>
<td>76.1% (( n = 185 ))</td>
<td>77% (( n = 187 ))</td>
<td>75.7% (( n = 184 ))</td>
</tr>
</tbody>
</table>
Evaluation/Assessment

Like drug administration, there was a large discrepancy between the number of episodes for comparisons. Chart records identified only 52 episodes \((n=31)\) of evaluation/assessment whereas 369 episodes \((n=136)\) were found in Medicaid claims. MFASIS again reported a higher number of episodes than the other two data sets \((595; n=100)\). Again, individual level data were examined to find outliers. A closer analysis of data found two outliers in recoded MFASIS files - one person received 223 evaluation/assessments and another, 41. None of the events were found in chart or Medicaid records.

Two exact match analyses were completed. MFASIS and Medicaid agreement for exact matches including the outliers \((5.6\%)\) were higher than the two comparisons, though all agreement rates were extremely low. Without the outliers, agreement rates improved. Though the number of matches did not change, the MFASIS records without outliers \((331, n = 100)\) used to calculate agreement slightly increased the agreement rate. Using matches on penetration improved percentages across all three comparisons, particularly for the MFASIS and Medicaid comparison. Agreement rates of non-users were higher overall and suggest charts and MFASIS agreed more often on who did not receive an evaluation/assessment during the time period under study.
Table 6: Agreement Rates of Evaluation/Assessment

<table>
<thead>
<tr>
<th>Evaluation/Assessment</th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td># matched episodes – exact match</td>
<td>7 ( (n=7) )</td>
<td>6 ( (n=6) )</td>
<td>51 ( (n=33) )</td>
</tr>
<tr>
<td>Percent Agreement – exact match</td>
<td>1.7%</td>
<td>.94%</td>
<td>5.6%</td>
</tr>
<tr>
<td># matched episodes – exact match excluding outliers</td>
<td>n/a</td>
<td>6 ( (n=31) )</td>
<td>51 ( (n=26) )</td>
</tr>
<tr>
<td>Percent Agreement – exact match excluding outliers</td>
<td>n/a</td>
<td>1.6%</td>
<td>7.9%</td>
</tr>
<tr>
<td># matched episodes – Penetration</td>
<td>15 ( (n=14) )</td>
<td>11 ( (n=10) )</td>
<td>138 ( (n=65) )</td>
</tr>
<tr>
<td>Percent Agreement – Penetration</td>
<td>4.5%</td>
<td>4.0%</td>
<td>34.8%</td>
</tr>
<tr>
<td>Percent Agreement – non-users</td>
<td>42.0% ( (n=102) )</td>
<td>52.7% ( (n=128) )</td>
<td>34.6% ( (n=84) )</td>
</tr>
</tbody>
</table>

Family Therapy

There were very few records reflecting family therapy in all three sources. Charts reported the highest number of episodes \( (27, n=9) \). MFASIS \( (3, n=3) \) and Medicaid \( (2, n=2) \) reported even lower use. No outliers were identified. Exact matches found no agreement between charts and Medicaid or MFASIS. There was one exact agreement, and one penetration episode that matched MFASIS and Medicaid files. As would be expected, with so few persons identified as receiving a family therapy service in each data source, agreement rates for non-users were extremely high.
Table 7: Agreement Rates of Family Therapy

<table>
<thead>
<tr>
<th>Family Therapy</th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td># matched episodes – exact match</td>
<td>0</td>
<td>0</td>
<td>1 (n = 1)</td>
</tr>
<tr>
<td>Percent Agreement – exact match</td>
<td>0%</td>
<td>0%</td>
<td>25%</td>
</tr>
<tr>
<td># matched episodes – Penetration</td>
<td>0</td>
<td>0</td>
<td>1 (n = 1)</td>
</tr>
<tr>
<td>Percent Agreement – Penetration</td>
<td>0%</td>
<td>0%</td>
<td>25%</td>
</tr>
<tr>
<td>Percent Agreement – non-users</td>
<td>95.9% (n = 233)</td>
<td>95.1% (n = 231)</td>
<td>98.4% (n = 239)</td>
</tr>
</tbody>
</table>

**Group Therapy**

A number of duplicate Medicaid claims in the area of group therapy were deleted from the Medicaid database prior to analysis. These claims were associated with a particular CPT code, W9204. This code is used by the state to represent a clinic visit in a group setting that was conducted by personnel other than a M.D. or Ph.D. Duplicate claims were found with same charge, number of units, dates of service and other identifying codes used by claims clerks to identify individual claims including transaction number and unique sequence identification numbers. Thus, duplicate claims that matched on participant, date of service, service code, transaction number and unique sequence identification were eliminated before analyses.

Chart records had the highest number of recorded group therapy episodes (479, n = 38) (MFASIS, 371, n = 32; Medicaid, 131, n = 21). No outliers were identified. Rates of agreement were highest between charts and MFASIS (29.8%) with somewhat lower rates for the other comparisons. Penetration comparisons increased the rates of agreement for each comparison. Agreement rates of non-users were higher than for users.
Table 8: Agreement Rates of Group Therapy

<table>
<thead>
<tr>
<th>Group Therapy</th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td># matched episodes – exact match</td>
<td>84 (n = 15)</td>
<td>195 (n = 18)</td>
<td>82 (n = 15)</td>
</tr>
<tr>
<td>Percent Agreement – exact match</td>
<td>15.9%</td>
<td>29.8%</td>
<td>19.5%</td>
</tr>
<tr>
<td># matched episodes – Penetration</td>
<td>59 (n = 16)</td>
<td>96 (n = 21)</td>
<td>62 (n = 15)</td>
</tr>
<tr>
<td>Percent Agreement – Penetration</td>
<td>28.4%</td>
<td>39.3%</td>
<td>37.1%</td>
</tr>
<tr>
<td>Percent Agreement – non-users</td>
<td>82.7% (n = 201)</td>
<td>80.7% (n = 196)</td>
<td>84.4% (n = 205)</td>
</tr>
</tbody>
</table>

Medication Check

The variability in reporting of medication checks was high. Charts reported 2,161 episodes (n=179), MFASIS reported 1,041 (n=123) and Medicaid reported only 215 (n=49). An examination of data did not suggest duplicated records, miscoding or outliers. Despite wide variability, agreement between charts and MFASIS in the area of medication checks was one of the stronger rates compared to the other five service categories examined (31.9%, n=119). Penetration rate analysis examining service use in each month found higher rates of agreement for all comparisons and ranged from a low of 11.2% to a high of 41.5%. Consistent with the variability associated with reporting of medication checks, agreement rates for non-users were low. In contrast to the four services reviewed earlier, the non-agreement rate was lower than the agreement rates when comparing chart data and MFASIS data. This pattern is also present in the last service area examined, therapy/case management, and reflects a greater number of users identified by at least one data source.
### Table 9: Agreement Rates for Medication Checks

<table>
<thead>
<tr>
<th>Medication Check</th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td># matched episodes – exact match</td>
<td>143 ($n = 36$)</td>
<td>775 ($n = 113$)</td>
<td>159 ($n = 27$)</td>
</tr>
<tr>
<td>Percent Agreement – exact match</td>
<td>6.4%</td>
<td>31.9%</td>
<td>14.5%</td>
</tr>
<tr>
<td># matched episodes – Penetration</td>
<td>151 ($n = 40$)</td>
<td>576 ($n = 116$)</td>
<td>157 ($n = 29$)</td>
</tr>
<tr>
<td>Percent Agreement – Penetration</td>
<td>11.2%</td>
<td>41.5%</td>
<td>22.1%</td>
</tr>
<tr>
<td>Percent Agreement – non-users</td>
<td>23.5% ($n = 57$)</td>
<td>24.7% ($n = 60$)</td>
<td>42.8% ($n = 104$)</td>
</tr>
</tbody>
</table>

**Therapy/Case Management**

Several complications were identified with Medicaid claims when examining therapy and case management services. The first was the discovery that case management data, a group of services including case management assessment, referral and linkage, and support were billed to Medicaid in an aggregate manner. A “bundle” of case management episodes were sent to Medicaid. The bundle identified a “service to” and “service from” date that described the time period in which the case management occurred. The units of service, billed in 15 minute increments, were included as well. This presented a serious problem for analyses in that a specific date of service could not be established.

Management information systems support staff described that charge and encounter slips were submitted by CMHC personnel to billing clerks. The clerks kept files that identified the date of service for the case management service, then clerks would bundle the service into one claim using a database. Files that included the individual dates of service were partially recovered from only three of nine CMHCs in the state; however, it
is unknown whether these data represent all case management services provided in the study period. In fact, some files recovered were from years outside the study period. This eliminated the ability to make comparisons confidently for therapy/case management services across the three sets of data. Instead, the only comparison using exact matching that was made for this service category included chart data and MFASIS files. Chart records identified 2,307 episodes \((n = 187)\) while MFASIS files had 1,104 \((n = 133)\). Exact matching found 450 matches \((n = 121)\) with an agreement rate of 15.2\%. Comparison using charts and MFASIS found a penetration rate of 30.2\% \((421 \text{ matches}, n = 110)\).

Agreement rates using penetration analyses allowed for an examination of all three data sources. In spite of the fact that bundles were not consistently billed by calendar month but instead, by 30-day periods that could begin mid-month, penetration analyses were still completed using the month defined in the “service to” variable to estimate agreement. Charts and Medicaid claims agreed more often than the other comparisons despite the approximate service dates used for Medicaid claims. Rates for non-users were much lower than penetration agreement rates suggesting little agreement about who is not receiving therapy/case management services.
Table 10: Agreement Rates of Therapy/Case Management

<table>
<thead>
<tr>
<th>Medication Check</th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td># matched episodes - exact match</td>
<td>-</td>
<td>450 (n = 121)</td>
<td>-</td>
</tr>
<tr>
<td>Percent Agreement - exact match</td>
<td>-</td>
<td>15.2%</td>
<td>-</td>
</tr>
<tr>
<td># matched episodes - Penetration</td>
<td>781 (n = 157)</td>
<td>421 (n = 110)</td>
<td>467 (n = 112)</td>
</tr>
<tr>
<td>Percent Agreement - Penetration</td>
<td>43.3%</td>
<td>30.2%</td>
<td>27.7%</td>
</tr>
<tr>
<td>Percent Agreement - non-users</td>
<td>17.3% (n = 42)</td>
<td>18.1% (n = 44)</td>
<td>19.8% (n = 48)</td>
</tr>
</tbody>
</table>

Agreement of Diagnostic Codes

Rates of agreement were calculated for diagnostic codes. Diagnostic codes are typically five digits long with the last two digits providing more specification about the condition. Agreement analyses were conducted using three, four and five digits. In Table 11 the rates of agreement are presented and show increased agreement with fewer digits. Since participants were selected based on available chart data, not eligibility, diagnoses were gathered for all participants in each data set. Chart data was the only set that had diagnoses for every group member (N=243), thus, this total was used as the denominator in the agreement analyses. Higher percentages of agreement can be seen when using three digits compared to five digits.

Table 11: Agreement Rates of Diagnostic Codes

<table>
<thead>
<tr>
<th></th>
<th>Charts and Medicaid</th>
<th>Charts and MFASIS</th>
<th>MFASIS and Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 digits</td>
<td>56% (n = 136)</td>
<td>67.1% (n = 163)</td>
<td>63.4% (n = 154)</td>
</tr>
<tr>
<td>4 digits</td>
<td>39.9% (n = 97)</td>
<td>49% (n = 119)</td>
<td>56.8% (n = 138)</td>
</tr>
<tr>
<td>5 digits</td>
<td>33.7% (n = 82)</td>
<td>35% (n = 85)</td>
<td>48.6% (n = 118)</td>
</tr>
</tbody>
</table>
Service Utilization

The average number of services used per participant during the study was calculated to compare service utilization depicted by each set of data (see Figure 1). For drug administration, there were no significant differences in service utilization; however, the trend was toward significance, $F(2, 131) = 2.99; p = .054$, with MFASIS showing higher utilization. Comparisons did not approximate significance, $F(2, 129) = 2.01; p = 1.38$, when outliers were excluded.

Figure 1: Mean Number of Service Episodes During the Study by Service Category

<table>
<thead>
<tr>
<th>Service Category</th>
<th>Charts</th>
<th>MFASIS</th>
<th>Medicaid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug Administration</td>
<td>15.33</td>
<td>22.28</td>
<td>10.93</td>
</tr>
<tr>
<td>Evaluation/Assessment</td>
<td>1.68</td>
<td>5.95</td>
<td>2.71</td>
</tr>
<tr>
<td>Family Therapy</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Group Therapy</td>
<td>12.60</td>
<td>0.00</td>
<td>6.24</td>
</tr>
<tr>
<td>Medication Check</td>
<td>12.07</td>
<td>8.46</td>
<td>4.39</td>
</tr>
<tr>
<td>Therapy/Case Management</td>
<td>12.34</td>
<td>8.30</td>
<td>15.49</td>
</tr>
</tbody>
</table>

Comparisons among data sets regarding the average number of evaluation/assessment services per participant yielded no significant difference in service

---

1 Service episodes only include services during Medicaid-eligible months.
utilization, $F(2, 266) = 1.998; p = .138$ though the average number of episodes is greater when using MFASIS as a source ($M = 5.95$, $SD = 22.4$). This average was influenced by the outliers discussed earlier. When the outliers were excluded, means were much closer in number; however, a one-way ANOVA found that the average service utilization reported in MFASIS remained significantly higher than both charts and Medicaid claims, $F(2, 264) = 6.806; p = .001$. Post hoc comparisons found that MFASIS service utilization was significantly different than both chart data and Medicaid. Tukey post hoc tests indicate that mean number of evaluation/assessment episodes from MFASIS data was significantly higher than data from charts.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charts</td>
<td>1.67a</td>
<td>1.11</td>
</tr>
<tr>
<td>MFASIS</td>
<td>3.38b</td>
<td>2.88</td>
</tr>
<tr>
<td>Medicaid</td>
<td>2.71ab</td>
<td>2.02</td>
</tr>
<tr>
<td>Total</td>
<td>2.84</td>
<td>2.36</td>
</tr>
</tbody>
</table>

Table 12: ANOVA for Evaluation/Assessment

No significant difference was found in service utilization of family therapy $F(2, 13) = 1.01; p = .396$ or group therapy $F(2, 90) = 2.387; p = .098$. Analysis of variance for service utilization of medication checks found a significant difference when all three groups were compared $F(2, 350) = 6.479; p = .002$. Tukey post hoc tests found that service utilization of medication checks reported in charts was significantly higher than Medicaid claims.
Table 13: ANOVA for Medication Check

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charts</td>
<td>12.07\textsuperscript{a}</td>
<td>17.99</td>
</tr>
<tr>
<td>MFASIS</td>
<td>8.46\textsuperscript{a,b}</td>
<td>9.42</td>
</tr>
<tr>
<td>Medicaid</td>
<td>4.39\textsuperscript{b}</td>
<td>4.13</td>
</tr>
<tr>
<td>Total</td>
<td>9.74</td>
<td>14.32</td>
</tr>
</tbody>
</table>

\[ F(2, 350) = 6.479; p = .001 \]

\textsuperscript{a,b} a,b Means with the same letter in their superscript do not differ significantly from one another according to Tukey with a .05 limit on family-wise error rate.

Service utilization was further examined by looking at total numbers of services over a one-year period as recorded by each dataset excluding data from months during which participants were ineligible for Medicaid. Figure 2 demonstrates the variability in the number of service episodes by service category as reported by different data sources. The pattern in Figure 2 is different from Figure 1 because each service category represents a different number of participants which affects average service episodes.

Chart data consistently report a greater number of services when compared to Medicaid claims except for evaluation/assessment and therapy/case management services. MFASIS is highly variable and reports a greater number of evaluation/assessment and drug administration services but fewer therapy/case management services than charts and Medicaid claims.
Differences in absolute numbers of services provided across data sources may suggest a difference in how services are described in charts relative to how they are actually labeled for billing purposes. For example, chart reviews identified fewer evaluation/assessment services when compared to MFASIS and Medicaid claims. This pattern is very different for medication checks. Potentially, such services are billed under a different service or not at all. Overall, charts identified more services than the other sources which may indicate a system of service delivery in which many services are provided but not billed.

**Funding Source**

The datasets examined in this research include data from persons receiving mental health services under two different funding conditions. These data, collected under a
demonstration project to examine satisfaction with a fee-for-service approach compared to a managed care approach, include persons in both conditions. A greater percentage of persons in the sample were served under a fee-for-service (FFS) condition (63.8%, \( n = 155 \)) relative to the managed care (MC) condition (36.2%, \( n = 88 \)). As mentioned earlier, there may be a greater incentive to submit a claim in a fee-for-service environment compared to managed care because payment is dependent on the claim submission. Therefore it might be expected that higher levels of agreement among data sources would be found for services provided in a fee-for-service environment. This pattern is reflected in Table 14. When examining rates of agreement for each service category by funding condition, higher rates of agreement are seen for the FFS condition in nine instances compared to four instances for the managed care condition. However, the differences in rates of agreement attributable to differences in funding conditions are not consistent nor of a magnitude which explains the overall low rates of agreement observed in this study.
Table 14: Agreement Rates for Service Categories by Funding Source

<table>
<thead>
<tr>
<th>Service Type</th>
<th>FFS</th>
<th>MC</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Agreement</td>
<td>% Agreement</td>
<td>% Agreement</td>
</tr>
<tr>
<td>Drug Administration(^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charts and Medicaid</td>
<td>28.1%</td>
<td>19.8%</td>
<td>27.2%</td>
</tr>
<tr>
<td>Charts and MFASIS</td>
<td>35.9%</td>
<td>36.5%</td>
<td>35.9%</td>
</tr>
<tr>
<td>MFASIS and Medicaid</td>
<td>29.4%</td>
<td>15.0%</td>
<td>27.0%</td>
</tr>
<tr>
<td>Evaluation/Assessment(^a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charts and Medicaid</td>
<td>1.4%</td>
<td>3.5%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Charts and MFASIS</td>
<td>1.8%</td>
<td>0.1%</td>
<td>1.6%</td>
</tr>
<tr>
<td>MFASIS and Medicaid</td>
<td>8.1%</td>
<td>5.4%</td>
<td>7.9%</td>
</tr>
<tr>
<td>Family Therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charts and Medicaid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Charts and MFASIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MFASIS and Medicaid</td>
<td>25.0%</td>
<td>-</td>
<td>25.0%</td>
</tr>
<tr>
<td>Group Therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charts and Medicaid</td>
<td>15.8%</td>
<td>16.9%</td>
<td>15.9%</td>
</tr>
<tr>
<td>Charts and MFASIS</td>
<td>33.0%</td>
<td>5.2%</td>
<td>29.8%</td>
</tr>
<tr>
<td>MFASIS and Medicaid</td>
<td>20.9%</td>
<td>5.3%</td>
<td>19.5%</td>
</tr>
<tr>
<td>Medication Check</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charts and Medicaid</td>
<td>7.2%</td>
<td>1.3%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Charts and MFASIS</td>
<td>33.5%</td>
<td>22.0%</td>
<td>31.9%</td>
</tr>
<tr>
<td>MFASIS and Medicaid</td>
<td>15.9%</td>
<td>1.8%</td>
<td>14.5%</td>
</tr>
<tr>
<td>Therapy/Case Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charts and Medicaid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Charts and MFASIS</td>
<td>15.1%</td>
<td>15.4%</td>
<td>15.2%</td>
</tr>
<tr>
<td>MFASIS and Medicaid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^a\)excluded outliers
CHAPTER FIVE
Discussion

Four of five hypotheses, restated on the following page, were partially supported and one was fully supported. In three of the six service categories chart data agreed with MFASIS at higher rates than any other comparison. It was expected that charts and MFASIS would have higher rates in all service comparisons because both systems are based on documenting services without concerns about billing and potential for reimbursement. MFASIS and Medicaid claims had higher rates of agreement in three service categories suggesting that the basis for these hypotheses may not apply to these data sources. Charts and Medicaid should have had the lowest rates of agreement but in fact, charts and Medicaid had the lowest rates of agreement in only two of the service categories while never having the highest rate of agreement. These results partially supported the first three hypotheses.

Agreement rates for diagnostic codes were highest using three digits as predicted and seen in previous studies. In regards to service utilization estimates, charts had the highest number of service episodes in four of the six service categories, partially supporting the final hypothesis. Because charts typically contain all services provided as recorded by the provider, factors such as data entry error and reimbursement incentives should not have affected data quality for this medium. Therefore, charts should contain the highest number of service episodes in every service category. This was partially supported in that service utilization estimates were higher using chart data in four of the six service categories. A detailed summary of hypotheses and supporting evidence is presented below.
1. It was hypothesized that rates of agreement between chart data and MFASIS will be higher than other comparisons.
   a. Charts and MFASIS records had the highest rates of agreement in three of six service categories.

2. It was hypothesized that rates of agreement between chart data and Medicaid claims will be lower than other comparisons.
   a. Rates of agreement between chart data and Medicaid were lowest in two service categories and never were the highest rate of agreement among comparisons.

3. It was hypothesized that the rate of agreement between MFASIS and Medicaid will be better than that found for charts and Medicaid.
   a. MFASIS and Medicaid agreement rates were highest in three service categories.

4. It was hypothesized that agreement rates found using only the first three digits of a 5 digit code will be better than rates using four and five digits.
   a. Rates of agreement were higher when using three digits of the diagnostic code compared to four and five digits.

5. It was hypothesized that services recorded in charts will significantly outnumber those recorded in MFASIS or Medicaid claims. It was also hypothesized that MFASIS will contain fewer recorded services than charts, but more than Medicaid claims.
a. Average service utilization estimates using chart data were higher than other sources in three service categories. Charts had the greatest number of service episodes and MFASIS had the lowest.

Rates of agreement among collected chart data, MFASIS files, and Medicaid claims were very low in comparison to published studies. There are a number of possible reasons for these low rates of agreement. For instance, the low agreement rates seen in this research compared to other literature can be potentially explained by the methodology used in agreement calculations. The approach developed takes a conservative view of agreement. Traditional agreement analyses take the number of instances in which both data sources agree that there is an absence of an event and include this in calculations. For example, if two observers agreed that a family member engaged in exercise on the same two out of 10 days, it follows that they also agree that exercise did not occur on the other eight days. This approach which presumes agreement in the absence of an event tends to over-inflate agreement rates and would have done so in these analyses. On the other hand, the approach used in this study which defines agreement as the number of occurrences in each of two data sets minus the number of matches will consistently show a lower agreement rate than the earlier method. Capacity for demonstrating high rates of agreement is limited when not including data indicating agreement during the absence of an event. A further limitation in the approach occurs when there are fewer episodes of service in one data source and significantly more in another in that a ceiling effect is created by the data so that a 100% rate of agreement is impossible.
Not surprisingly, alternative calculations of agreement rates improved reliability estimates. The approach used in this study to examine if services may have been recorded within a temporally adjacent time frame, i.e. assessment of penetration rates, only identified whether an individual used the service in a particular month and did not address the intensity, or the number of times a service was utilized. The improvement in agreement suggested that there was some discrepancy between exact dates of service but there still were a number of participants and their services unaccounted for during comparisons. In other words, even when given a larger window around a date of service there was still a significant amount of disagreement between the databases. This lack of accuracy can only be improved upon by interventions and feedback to the primary front-end data users. Indeed, earlier research cited suggests that low rates of agreement can improve with staff training and regular monitoring of data quality (Fine et al., 2003; Worth & Mytinger, 1996).

Rates based solely on exact matching can be a challenge with these data because of the potential areas for error and such was likely seen in this study. Data entry may produce error when examining exact dates. Despite efforts to correct for this potential error by using a broader agreement indicator “penetration,” agreement rates remained low. These continuing low rates of agreement suggest that other factors are also likely to contribute to the observed low rates.

Chart data were collected using categories not specifically developed to match pre-existing coding systems. It is anticipated that a review of charts focusing exclusively on the presence or absence of a CPT defined event, or for that matter, on the presence of an MFASIS defined event would lead to higher rates of agreement for each related system.
However, the nature of this study which used a more global chart review process to allow comparisons across each system may have therein contributed to lower rates of agreement.

This study did not have the opportunity to take individual charts and compare services to MFASIS or Medicaid data in the way that would occur if charts were being audited to assure correspondence across systems. For example, in a Medicaid billing audit a reviewer would examine a clinical chart to look for evidence as documentation to support a claims submission. This approach may allow for a more focused search rather than one which reviews a time period in a chart and attempts to categorize and document service provision.

This matching process is also likely to be complicated by the fact that multiple services within the same service category can occur during one appointment. A visit to a psychiatrist may include some psychotherapy, a medical evaluation and medication prescription. Funding incentives and the potential for payment may impact whether three services are submitted for billing or just one. Therefore, that one visit may generate recording of one service or multiple services. Guidelines placed by the funding agency may limit how services can be billed so that if a health care organization is interested in tracking services, claims data may provide a limited picture.

It is also possible that information system codes may have been used for other purposes than assumed in this research. For example, the medication check procedure code in MFASIS is defined for the use of prescription administration and/or examination of medication side effects. MFASIS, a system built for record keeping and not for billing purposes, provided a system to log the frequency of prescribing medication. Given the
nature of mental illness, it is possible that medication administration may have been monitored or provided using the codes in this project by nurses who maintained non-billable records in MFASIS. Nurses may have also helped patients who dropped in without an appointment get a medication refill or prescription from the clinic psychiatrist. This may explain the high number of medication check procedure codes in MFASIS relative to Medicaid claims. Claims data, a system built for reimbursing services is not designed to capture services that may be provided but not assigned a cost.

The extremely low rates of agreement for evaluation/assessment services also indicate a lack of standard definitions across sources. Evaluation/assessment was defined as a service in which a formal psychological assessment was provided. Charts did not indicate that this service occurred often. MFASIS and Medicaid claims reported a much greater number of these assessments than charts. This may indicate that the procedure codes in Medicaid and MFASIS had a broader definition for evaluation/assessment than that used during the chart review component of the study. The Medicaid and MFASIS codes for evaluation/assessment may have included a number of other services not captured by the definition used for this study. Charts and MFASIS had extremely low levels of agreement as well, suggesting that error in the definition and categories used in the crosswalk table may have greatly impacted the ability to find matches.

Findings suggest that service categories that were more unique and required aspects of care that were not present in other services had better rates of agreement. Higher rates of agreement were seen for drug administration and group therapy, distinct services that contain specific qualifiers that are rarely duplicated in other mental health services. Medication checks, therapy/case management, and evaluation/assessment had much
lower rates of agreement. Though service definitions were based on previously accepted definitions used by the mental health system, there is potential overlap in how these services may be identified in charts. This potential may have affected coding of charts by researchers as well as billing clerks, thus introducing error in each data source.

An important lesson learned is that the development of rationally derived \textit{a priori} categories for classifying mental health services may not be optimal for establishing reliability between chart data and an administrative system. The nature of mental health services may not always be easily categorized as any one type of service. This forces clinicians, billing clerks and administrators to make decisions about how services should be categorized and billed. Such decisions affect how services are tracked and reported, ultimately affecting the patterns of service utilization seen with administrative data. When these data are then used to inform policy or resource allocation, mental health systems of care may be greatly affected by the decisions made.

For example, when looking further at the practical use of these data, service utilization patterns tell a slightly different story. Though the agreement across these data sets appears poor, most comparisons for average service utilization were not statistically significantly different from one another and the same general patterns of service utilization were observed across services (i.e. all data sources showed drug administration to occur more frequently than medication checks, etc.). However, if mean numbers were used to calculate costs of services the impact from the differences across sources would be dramatic. The numbers of services provided each month as characterized by the three data sets suggest that a direct examination of the intensity of service use, or the actual numbers of services provided each month, can vary dramatically from set to set.
Moreover, if administrators used one set exclusively over another without understanding the discrepancies in systems, over- and underestimation of service utilization could occur. This is particularly problematic when managing health care resources that are often scarce and limited.

The specifics of data management influenced the utility of working with these administrative data sets. Case management data submitted to Medicaid was provided in an efficient way for billing purposes, but once data were aggregated it became difficult to find exact or even approximate matches. This raises an important issue for data management systems. As organizations begin to automate services there may be competing resources and needs. In this case, case management data were entered into two separate systems – one for recording purposes and another for billing. The manner in which data were handled was different for each system. Data recording individual services were not maintained and could not be fully recovered whereas claims data were maintained. Interestingly, payments received from claims were not regularly entered or monitored, suggesting that this particular information system was more a tool to generate invoices than track billing and payment. Presumably, a billing system with expanded capacity and an emphasis on the full cycle of billing activities – service tracking, claims submission, adjustment when denied, tracking of payment – will include a greater number of data quality checks. It appears that attitude toward a data system greatly affects the data contained within.

Potvin and Champagne (1986) point out that the complexities of data management require experts in data systems, management, and statistical analysis in order to work with administrative data. Though some statistical measures, such as percent agreement,
appear straightforward, the process to calculate those statistics can be arduous and complicated. In the present study, data had to be extracted from a larger database using multiple participant identifiers. Multiple identifiers were needed to ensure that the correct participant’s data were collected. Data in each set had to be recoded so that sets could be compared. Recoding data using the crosswalk table introduced a challenge in that development of the table required expertise in understanding how charts are coded, how MFASIS definitions were used to record services and how CPT procedural codes might be equated.

Even once the stage of having a crosswalk strategy in place is reached, researchers will benefit from knowledge gained on the specifics of service delivery, data entry and data management. To work with claims data, an understanding of how services are submitted and what data are attached to those service episodes are necessary in order to identify duplications or errors in claims. Also, an understanding of the different types of claims (e.g., original, appealed) and evidence of payment or denial of the claim often are attached to a single episode. In the present study, obtaining this knowledge required multiple discussions with the data management team that housed and processed such data. Such discussions are imperative to working with these data – yet it is a time-consuming process and requires some basic knowledge of data systems and methodology. Organizations striving to use administrative data as a primary source for policy and decision-making in areas of outcome reviews, planning and service development would do well having personnel with a detailed understanding of how data work and the potential impact of data management on accuracy.
The inclusion of an expert panel to verify operational definitions of service categories was a unique aspect of this research that is not found in the literature. The benefit of this panel was that it provided a working and appropriate definition for a group of services that otherwise would not have been grouped together under claims data. Bringing a number of experts together to develop and fine tune definitions provided some validity to service categories used for analyses. In addition, this panel, which was comprised of clinicians, data entry and billing personnel, was a valuable resource in helping the researcher understand the nuances of data management systems. The panel was convened after data were collected and archived. As such, there was still some potential that data collected during chart reviews would not fall discretely into the administrative data categories. The differing rates of agreement between service categories suggest that it may have been valuable to convene the panel earlier in the research process. Certainly, a key recommendation evolving from this study is that health services researchers participate in extended discussions with some equivalent group of primary data system users prior to deciding on the chart collection procedures and the design of services research using administrative data.

HIPAA attempts to address concerns relevant to data quality by requiring health care organizations to use a shared standard for the exchange of information. HIPAA includes standardized service categories that allow organizations to develop internal service definitions and codes yet requires electronic submissions of service data to be consistent with the standard definitions of HIPAA. However, of concern for services research using administrative datasets are the regulations related to confidentiality introduced by HIPAA. No longer can identifying information be attached to electronic
claims information. Ultimately, each organization will be required to create unique identifiers for their own billing systems which will continue to keep billing systems separate from service records that contain personal information. In the current research, once data were recoded a variable had to be created that combined the person’s identification number, date of service and service code. Matching data across data sets required creating separate files of services with unduplicated data so that matches could be made for each service. Examination of penetration rates required a similar process.

Without a unique identifier to match individual data it would have been impossible to determine if a match were made. There is some research and methods that suggest when estimating population characteristics precise matching does not need to occur (Banks & Padiani, 2001) potentially limiting concern in this area. Despite these methods, linking service information with outcome information may become more difficult with the strict confidentiality procedures required by HIPAA and unless certain systems are set in place with research in mind, using automated systems may become even more difficult and eventually limit the use of administrative data in health services research.

The attempt to create a standard system of reporting services is promising. However, such a system will only work if fully implemented across all data systems used by providers. For example, the Medi-Cal Policy Institute (2001) study found that in some circumstances providers were not submitting claims at all because a lack of incentive. A similar phenomenon was seen with the MFASIS data where non-billable data were maintained in an antiquated fashion unrelated to CPT codes and without pressure to adopt potential HIPAA service standards. CMHC staff report that there was little incentive to track incoming payments from Medicaid because funding for service provision was
coming from multiple funding sources and there was not need to create profit.

Additionally there was no need to store certain types of service data once claims were submitted. HIPAA requires stringent rules about submission of patient information but not of the requirement to submit.

There are a few aspects of this study that limit the generalizability of findings. The sample is unique and not representative of all persons with a mental health issue. Services were restricted to those provided in a public mental health center where specific approaches to billing and recording may be quite different than other settings such as a private clinic, an in-patient hospital, or a day treatment center. Also, this study used data from a sample in one state. Because of state-specific processes, specific findings cannot be applied to other state mental health systems. Though this research may help the local state to better understand their own data systems and could contribute to improvement of data quality, it should not be assumed that all states collect and keep data in the same manner.

Despite these limitations, an important conclusion from this study is that researchers must always consider the context around how administrative data are gathered, recorded, maintained and used within the system. Incentives for accurate data entry, reimbursement structures and service definitions impact how data may appear when using it to examine service quality and utilization.

The results of this research validate the known limitations of administrative data. Findings also suggest that studies examining the reliability between extracted chart data and administrative data for mental health services be conducted with caution. Service categories should be informed by the existing systems that are being compared in order to
ensure shared operational definitions. In addition, a case by case analysis comparing chart records to administrative data and, vice versa, administrative data to chart records may help identify how services are being tracked in administrative data compared to chart data. Also, when using archived data it is invaluable and necessary to include persons that developed and maintained systems so that a greater understanding of the data can be applied to data analyses and interpretation of findings.

These conclusions emanate from the basic fact that this study found very low rates of agreement across three data sources used by one organization to track and report mental health services. In addition to contributing to the accumulating evidence of known limitations in the area of administrative data, this finding suggests that researchers and policy makers may not be able to expect data sets to approximate or converge around a “gold standard” of accuracy. It may be equally plausible to recognize that alternative approaches capture alternative versions of mental health service “reality.” This perspective subtly changes the focus of discussion to one which more fully attempts to understand the context and influences of the version of reality captured by a particular data source. The perspective also clarifies for researchers the compromises inherent in the use of administrative data.
APPENDIX A

Chart Review Record

<table>
<thead>
<tr>
<th>Record Number</th>
<th>Recorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent ID</td>
<td>Service Codes</td>
</tr>
<tr>
<td>Name</td>
<td>Rx = Prescription Given</td>
</tr>
<tr>
<td>Today's Date</td>
<td>Da = Drug Administration</td>
</tr>
<tr>
<td>Date of Baseline Interview</td>
<td>Med = Medical Check</td>
</tr>
<tr>
<td>D.O.B.</td>
<td>Bd = Blood Draw</td>
</tr>
<tr>
<td>MHC</td>
<td>Tx = Therapy/CaseManagement</td>
</tr>
<tr>
<td>Axis I</td>
<td>Gt = Group Therapy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Rx</th>
<th>Da</th>
<th>Med</th>
<th>Bd</th>
<th>Tx</th>
<th>Ft</th>
<th>Ev</th>
<th>Ph</th>
<th>TxP</th>
<th>CI</th>
<th>Rf</th>
<th>Pw</th>
<th>Bps</th>
<th>Oth</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Therapy/Case Management

Face-to-face session with a mental health professional (MD/PhD/other) that involves the handling of problems or issues relevant to the treatment of mental health concerns or illness; behavior modification and supportive therapies are included.

<table>
<thead>
<tr>
<th>Chart Record Category</th>
<th>CPT Codes (HI and Federal, 1997 and 1998)</th>
<th>MFASIS Codes</th>
</tr>
</thead>
</table>
| Therapy/Case Management | 99211 - 99215 Office or other outpatient visit 98  
90804 - 90815 Individual psychotherapy 98**  
90841 - 90845 Individual medical psychotherapy 97  
90845 Psychoanalysis 98  
90835 Psychiatric therapeutic procedure 97  
90836 Psychiatric therapeutic procedure 98  
90855 Interactive individual medical psychotherapy 97  
90841-45  
G0071 – G0074 Individual psychotherapy HI  
W9892 Monitoring/follow-up services HI  
W9205 Clinic Visit HI | 6: Individual Counseling/Therapy  
34: Crisis Stabilization |

**will include any modifiers e.g. -22, -52

---

# Group Therapy

Face to face session with a mental health professional (MD/PhD/other) that includes at least three consumers - typically evidenced in progress notes as a group session with a particular focus.

<table>
<thead>
<tr>
<th>Chart Record Category</th>
<th>CPT Codes (HI and Federal, 1997 and 1998)</th>
<th>MFASIS Codes</th>
</tr>
</thead>
</table>
| Group Therapy | 90857 Interactive group psychotherapy 98**  
90853 Group psychotherapy other than multiple family group 98  
W9204 Clinic Visit Group/Intake assessment HI | 7: Group Counseling/Therapy  
54: Group - Other |
**Family Therapy**

Face to face session with a mental health professional (MD/PhD/other) that includes the consumer and at least one family member.

<table>
<thead>
<tr>
<th>Chart Record Category</th>
<th>CPT Codes (HI and Federal, 1997 and 1998)</th>
<th>MFASIS Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Therapy</td>
<td>90847 Family psychotherapy conjoint w patient 98</td>
<td>8: Couple/Family/Counseling/Therapy</td>
</tr>
<tr>
<td></td>
<td>90849 Multiple family group psychotherapy 98</td>
<td></td>
</tr>
</tbody>
</table>

**Evaluation/Assessment**

A session that includes testing or psychological assessment conducted by a mental health professional (MD/PhD/other) for the purposes of treatment planning and/or evaluation.

<table>
<thead>
<tr>
<th>Chart Record Category</th>
<th>CPT Codes (HI and Federal, 1997 and 1998)</th>
<th>MFASIS Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation/Assessment</td>
<td>90801 Psychiatric diagnostic interview/evaluation 98</td>
<td>3: Evaluation (MD/PhD)</td>
</tr>
<tr>
<td></td>
<td>90820 = 90802 97</td>
<td>4: Treatment/Service Planning</td>
</tr>
<tr>
<td></td>
<td>90802 Interactive medical psychiatric diagnostic interview examination 98</td>
<td>50: Psychological Testing</td>
</tr>
<tr>
<td></td>
<td>99201 – 99204 Evaluation/Management</td>
<td>55: Intake</td>
</tr>
<tr>
<td></td>
<td>W9890 Case assessment HI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W9891 Case planning HI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90830 Psychological Testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>96100 Psychological testing w/report</td>
<td></td>
</tr>
<tr>
<td></td>
<td>96117 Neuropsychological testing w/report</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W9206 DHS Psychological Diagnostic/Evaluation HI</td>
<td></td>
</tr>
</tbody>
</table>
Administration of Drug

Evidence of administration of a psychotropic medication noted in progress notes or nursing notes.

<table>
<thead>
<tr>
<th>Chart Record Category</th>
<th>CPT Codes (HI and Federal, 1997 and 1998)</th>
<th>MFASIS Codes</th>
</tr>
</thead>
</table>
| Drug administration   | 90782 Therapeutic or diagnostic injection subcutaneous or intramuscular 98  
   90783 Intra-arterial therapeutic or diagnostic injection 98  
   90784 Intravenous therapeutic or diagnostic injection 98  
   90799 Unlisted therapeutic or diagnostic injection 98  
   J0110 Administration of injectable  
   J1631 Haldol intramuscular injection  
   J2680 Prolixin intramuscular injection | 29: Medication admin/ supervision |

Evaluation of Medication

Face-to-face session with a mental health professional (MD/PhD/other) that involves the handling of problems or issues relevant to the treatment of mental health concerns or illness; behavior modification and supportive therapies are included along with a review of current medications and/or side effects of prescribed psychotropic medication.

<table>
<thead>
<tr>
<th>Chart Record Category</th>
<th>CPT Codes (HI and Federal, 1997 and 1998)</th>
<th>MFASIS Codes</th>
</tr>
</thead>
</table>
| Medication Check      | 90801 - 90802 Psychiatric diagnostic interview/evaluation 98  
   90862 Pharmacologic management including prescription use and review of medication with minimal medical psychotherapy 98  
   90805** Individual psychotherapy with med management 98  
   90807** Individual psychotherapy with med management 98  
   90809** Individual psychotherapy with med management 98  
   90811** Individual psychotherapy with med management 98  
   90813** Individual psychotherapy with med management 98  
   90815** Individual psychotherapy with med management 98 | 28: Medication/prescription evaluation |
REFERENCES


