ROLE OF PATIENT SEVERITY IN PREDICTING LENGTH OF HOSPITAL STAY

by

Janet Durbin

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Community Health
University of Toronto

Canada

ABSTRACT

Role of Patient Severity in Predicting Length of Hospital Stay

Janet Durbin

Master of Science, 1997

Graduate Department of Community Health

University of Toronto

This study assessed the role of the Computerized Severity Index (CSI®) in predicting length of stay (LOS) for discharges from psychiatric inpatient units. Using the CSI, raters retrospectively reviewed medical charts, producing three continuous severity ratings for each patient- admission, maximum and discharge. Hierarchical and stepwise multiple regression analyses were conducted to predict LOS using CSI ratings and other patient data. Subgroups with psychotic disorders and major depressive disorders were modeled separately. In both groups severity measures predicted 8-10% variation in LOS controlling for discharge abstract variables, and final models predicted 14-19% variation in LOS. While severity improves capacity to predict LOS, prediction is still inadequate for funding purposes. Future work should concentrate on incorporating measures of personal and social functioning into severity ratings, and on assessing provider and site influences on LOS. In mental health, funding models based on program rather than individual performance may be more appropriate.

ACKNOWLEDGMENTS

I would like to warmly thank the following individuals:

My thesis supervisor Dr. George Pink for providing concrete feedback at every step, responding always with incredible speed, and having confidence in my skills.

Dr. Michael Murray for shepherding me through the statistical forest.

My daughters, Anna and Deborah, for their constant encouragement and pride in my effort, and for their incredible sweetness and wonderful hugs.

My husband, Edwin, for holding down the ship when I wasn't around, for believing I could and should do it and, above all, for caring.

And finally, Dr. Paula Goering, who was instrumental in every step of the project by providing practical support, scholarly input and moral support, and who has become the mentor that I spent many years looking for.

TABLE OF CONTENTS

CHAPTER 1:   **LITERATURE REVIEW**

CHAPTER 2:   **METHODOLOGY**

CHAPTER 3:   RESULTS

CHAPTER 4:   DISCUSSION

LIST OF TABLES

CHAPTER I

LITERATURE REVIEW

Introduction

During the last several decades the health care system has been undergoing reform in order to

meet opposing demands of containing health costs while serving greater numbers of people

with increasingly complex care needs. Fundamental changes to the funding, governance and

delivery of health services have occurred in response to escalating provincial and national

debts. There is a belief that, through restructuring and reform, high quality health services can

be provided more effectively, at lower cost. While health reform strategies vary, most give

increased emphasis to evidence-based decision making and the development of more cost-

effective and accountable health systems. Numerous technologies and methodologies have

emerged to inject greater accountability into service delivery and performance, and every

health sector is searching for reliable and valid measures of costs, outputs and outcomes that

can be monitored for different subgroups of service users (Canadian Institute for Health

Information, 1997).

In the institutional acute care sector, health care funding reform began in earnest in the early

1980s to control decades of dramatically rising hospital costs (Statistics, Canada, 1990). In

Canada many provinces moved from line-by-line to global budgeting to control cost increases,

but this approach had a limited capacity to respond to changes in hospital activity over time.

Global budgets that were based on an index year of hospital activity became less relevant over

time as hospitals altered treatment programs and approaches to serve changing patient

populations, applied new technologies and strived for increased efficiencies.

To quell rising costs in the United States, case mix funding emerged as a strategy for reimbursing inpatient services. This genre of funding links reimbursement to patient characteristics, setting a preset fee for each subgrouping of patients treated, regardless of the site of delivery or actual costs of care. The expected benefits are more equitable reimbursement, more incentives for productivity and opportunities to promote better practices. Case mix funding was implemented in 1983 in the United States under the Medicare Prospective Payment System (PPS) using Diagnostic Related Groups or DRGs® as the patient grouping system. For surgical and some medical cases, DRG based funding was perceived to provide appropriate levels of reimbursement. However, because DRGs did not perform well in psychiatry, psychiatric hospitals and psychiatric units in general hospitals were exempted from participating in the PPS until more appropriate case mix categories could be developed (Mitchell, Dickey, Liptzin et al, 1987).

Currently the mental health field is under considerable pressure to develop better funding models but efforts to define more accurate case mix groupings and predictors of resource use have been only minimally successful. Patient severity currently holds considerable interest as a potential grouping variable because it is believed to be an important determinant of hospitalization course. However, defining and measuring severity has been difficult. The purpose of this study is to examine the feasibility and value of using a promising measure of patient severity, the *Computerized Severity Index* for predicting resource use by psychiatric

patients during hospitalization. The study generates knowledge about whether this severity measure provides better information than existing administrative data for making decisions about hospital funding for psychiatric patients in Ontario.

The *Computerized Severity Index* (CSI®) (Stoskopf & Horn, 1991) produces a severity rating for each episode of inpatient care based on all the diagnoses in a patient's chart, and has been shown to produce reliable ratings (Thomas & Ashcraft, 1991). While Horn's evaluation of the Computerized Psychiatric Severity Index (Horn, Sharkey, Chambers et al, 1989), a precursor to CSI, demonstrated a strong association between severity and length of stay and case cost, published evaluations of the CSI have not included psychiatry (Averill, McGuire & Manning, 1992; Horn, Sharkey, Buckle et al, 1991; Thomas & Ashcraft, 1991). The goal of this study is to take advantage of the many years of work underlying development of the CSI rating system and to test its performance on a subset of psychiatric discharges from Ontario hospitals. Because the cost of collecting severity data is considerable, this study compares the improvement realized by using a severity rating to predict length of stay (LOS) over that achieved by a subset of patient variables that are currently collected for administrative purposes by Ontario hospitals. Because a reduced set of predictors is more practical for developing case mix groups, the study goes on to identify the subset of variables that maximizes prediction of inpatient LOS. The study is confined to a sample of discharges from three Ontario hospitals and two diagnostic groupings. If findings are positive, further testing of the CSI system on a larger and more heterogeneous sample would determine its appropriateness for use in system wide management and funding tools. International Severity

Information Systems, Inc. (ISIS), the American-based company that distributes the CSI software, provided a time limited site licence for use of the software, trained the severity raters and conducted rater reliability checks on a cost recovery basis.

Study Objectives

1. **To determine the feasibility of using the *Computerized Severity Index (CSI®)* to rate severity in psychiatric inpatients in Ontario.**

2. **To determine if the CSI severity measure predicts significantly more variation in patient length of stay than a subset of patient variables currently available in hospital discharge abstracts.**

3. **To specify a model that can predict length of stay using a parsimonious subset of study variables.**

Case Mix Funding

To make distribution of hospital funding more equitable, many jurisdictions are moving from global budgeting to a system that considers hospital case mix and volume. *Case mix systems* define groupings of treatment episodes which are clinically meaningful and are expected to consume similar amounts of resources. All patients in each grouping are funded at the same level regardless of days of care and actual resources used. Variability in treatment costs within a group is expected, but on average case reimbursement exceeds case costs. Groupings are designed to be exclusive and exhaustive so that there is only one appropriate group for any case. Hospital productivity and costs are reflected in the weighted case mix which is derived

from the distribution of episodes of care among the case mix groups and the amount of resources expected to be used by the average patient in each group (Lave, Jacobs & Markel, 1992; Botz, 1991). Case mix funding appeals to funders because risk for exceeding budgets is transferred to individual hospitals. It is perceived to be more equitable because levels of funding are independent of site of service delivery or historical funding patterns.

A major impediment to the introduction of funding based on case mix is the lack of suitable case mix systems, particularly in psychiatry. The DRG system (Fetter, Shin, Freeman et al., 1980), which underpins the American Medicare prospective payment system, assumes that the amount of resources required during a hospitalization episode is determined by a patient's primary diagnosis. In psychiatry where numerous studies have proven this assumption to be false (McCrone & Phelan, 1994; Horgan & Jencks, 1987), psychiatric hospitals and 50% of general hospital psychiatric units have been excluded from prospective payment. In a somewhat parallel situation in Ontario, psychiatry was excluded from the early years of acute care funding reform because of the poor performance of the case mix group (CMG®) system for grouping psychiatric inpatients.

Early Psychiatric Case Mix Systems

The early psychiatric DRGs and CMGs were based almost entirely on patient diagnosis and initially there were only 15 categories. In a study of more than 22,000 Medicare discharges for psychiatric problems, approximately 30% fell into only one DRG category "psychoses" and four categories accounted for 85% of discharges (Mitchell, Dickey, Liptzin, 1987). Studies

estimated that surgical DRGs explained up to 39% of variance in patient LOS compared with

psychiatric DRGs which explained about 2-15% of the variance (Frank & Lave, 1985; Horgan

& Jencks, 1987). Several factors may account for the poorer performance of the psychiatric

DRGs. First, given that guidelines for diagnosis and treatment are relatively imprecise in

psychiatry compared with surgery, it is not surprising that similarly diagnosed psychiatric

patients vary more in their consumption of hospital resources than surgical patients (Ashcraft,

Fries, Nerenz et al., 1989). Related to this, Huxley, Challis, Hughes et al (no date) note that

there is sometimes more than one effective approach to treating a particular condition or,

alternatively, no obvious solutions. As such, variation in provider judgement and practice

patterns can lead to different treatment programs for patients with similar conditions. Second,

Lyons and colleagues (1995) suggest that predictor paradigms for resource utilization are

inappropriate. It is not the psychiatric diagnosis but the symptoms and circumstances

associated with the diagnosis that affect the admission decision and the subsequent course of

inpatient psychiatric treatment. This explanation is supported by Sanderson and colleagues

(1995) who noted that length of stay has more to do with a patient's ability and support and

with the type of treatment than with diagnosis.

The poor performance of the initial psychiatric case mix groups prompted a search for

improved methods of classifying patients. It was widely believed that diagnosis was only one

of many factors defining treatment needs, and that, to be more effective, a classification system

should take into account factors such as severity of symptoms, current impairment in adaptive

functioning, social supports, disposition options and treatment objectives (Jencks & Goldman,

1987; Caton & Gralnick, 1987). As a result, the bulk of the patient classification research that followed assessed the performance of the initial DRG/CMG groups, further divided by patient, hospital, and system modifiers. The approach reflects the belief that medical diagnosis is relevant but not sufficient for categorizing patients and that more homogeneous groups can be realized by using additional classifying variables. For example, groups can be subdivided by patient characteristics such as age, repeat admissions, severity of illness; hospital descriptors such as presence of psychiatric unit, teaching status, number of beds, attending physician, treatment goals and philosophy, criteria for discharge; and system factors such as availability of community services and step down hospital services (such as day treatment), demand for psychiatric beds, insurance limits and distance to closest hospital.

Case Mix Funding in Ontario

In Canada, the Canadian Institute for Health Information (CIHI) is responsible for developing and modifying the CMG classification system although provinces can implement local variations. When the CMG system was implemented in the early 1980s, it was used primarily by hospitals for internal utilization management. However, with the advent of hospital funding reform in Ontario and in Alberta, and the decision to use case mix groups to make adjustments to hospital funding, CMGs came under closer scrutiny. In Ontario concerns were raised about the face validity and discriminatory power of the psychiatric CMGs. There were only 15 psychiatric CMGs; they were defined solely by diagnosis so that within group variation in clinical profiles and resource use was considerable; and group definitions were based on the ICD-9 diagnostic system rather than the DSM-IIIR system which is widely used in psychiatry.

In response to these criticisms, CIHI formed a Psychiatric Task Force to review and amend the CMGs. The Task Force altered the titles and definitions of the main psychiatric CMG categories to make them more compatible with the DSM-IIIR system. As well the main CMGs were split using variables intended to capture case complexity and severity. The revised 1994 system included 35 CMGs, defined by patient variables such as primary diagnostic group, Axis 3 diagnosis (physical conditions), ECT and age. These revised CMGs were used to make general hospital funding adjustments in 1995-96 and 1996-97. However funding to provincial and specialty psychiatric hospitals in Ontario is still based primarily on global budgets and concerns persist about whether the current psychiatric CMGs are accurate enough to form the basis for funding psychiatric services in the acute care sector. An evaluation of the 1994 psychiatric CMGs found that within group variation in length of stay was still considerable (Joint Policy and Planning Committee, 1995).

Currently in the province of Ontario, CMGs are being used to determine funding in acute care hospitals for inpatient and day surgery activity, including psychiatric care. If a more accurate classification system is not developed, hospital funding for psychiatric patients may be inappropriate. Good programs and practices will not be supported and hospitals may feel pressured to manipulate services and patient populations to remain financially viable.

Assessing Case Mix Systems

Assessment of a case mix system needs to consider what it costs to implement the system and how well it performs (Taube, Lee & Forthofer, 1984). *Costs* are determined by accessibility of

the information necessary to make the classification, training requirements for data collectors

and classifiers, and level of system auditing needed. *Performance* considers the reliability of

the classification system and the validity of the categories produced. If a classification system

is costly to implement, its feasibility is questionable, especially for system wide applications. If

a classification system does not perform well, payments to hospitals will be more or less than

their actual treatment costs.

Cost

Case mix systems can be based on data reported routinely in discharge abstracts for

administrative purposes; data typically recorded in a patient's medical chart; or primary data

collected by clinicians specifically to support case mix classification. Discharge abstract data

can be incomplete and inaccurate, and do not necessarily include the items most relevant to

predicting need for care and resource use (Hopkins & Carroll, 1994). These same problems can

apply to chart data. For example, in a chart review study conducted by Bezold and colleagues

(1996), 498 charts were disqualified (50% of eligible sample) because they lacked multiaxial

diagnosis data. Primary data collection involves use of specific assessment tools administered by

trained clinicians or independent raters. If properly planned, it should yield the most complete

information but also incurs the greatest cost.

Cost is a function of the number of data items required to make a classification, type of rater

(ie., independent rater or clinician), the need for rater training (initial and ongoing), and the

need for auditing. Auditing is necessary because providers may be tempted to assign diagnoses

or employ treatments which place patients in a higher paying category. For example, after prospective payment was introduced, Kiesler and Simpkins (1991) reported a substantial increase in admissions for affective disorders which were reimbursed at a higher rate than admissions for depressive neuroses. Periodic audits can mitigate the problem of manipulation but are not foolproof and can be expensive. The increased cost and burden to staff of direct data gathering can more easily be justified and accepted if gathered data serve multiple purposes - for example, care planning or outcomes measurement as well as case mix.

In Ontario a computerized algorithm classifies patients into CMGs using data captured in the electronic discharge abstract routinely reported to CIHI. Costs would be higher in a system which requires additional data gathering and classification by trained raters (Botz, Bestard, Demaray & Molloy, 1993; Charles & Schalm, 1992).

Reliability

Reliability refers to the consistency or repeatability of a measurement instrument. The three major types of reliability are stability, homogeneity and equivalence (Giovannetti, 1979). *Stability* refers to the consistency of the measures on repeated applications or test-retest reliability. *Homogeneity* refers to the extent to which an individual's responses to the items in an instrument are consistent (i.e., internal consistency). *Equivalence* is concerned with the consistency of results when different individuals assess the same individual with the same instrument at the same time (i.e., inter-rater reliability). Measures of equivalence represent the most important aspect of reliability in patient classification. The classification of a patient

should be the same, regardless of who is collecting data or making the classification. To the extent that subjective opinion can affect how a variable is recorded or how a classification is made, inter-rater reliability can be compromised. Alternatively, inter-rater reliability is enhanced through use of objective rating algorithms based on clearly defined data items, and through training and periodic monitoring of classifiers. Inter-rater reliability for categorical data can be assessed with a coefficient kappa (Cohen, 1960). If considerable resources are needed for training classifiers and conducting audits to maintain reliability, the costs of implementing a classification system increase.

Validity

Validating a classification system is a process of determining the confidence that can be placed on inferences made about patients in each category (Streiner & Norman, 1989). A case mix system is expected to group patients who are homogeneous in their clinical profile and use of hospital resources. Tests of validity need to address these expectations.

Content validity assesses whether the criteria used to classify patients seem to be necessary and sufficient to create the desired groups. A common method of determining content validity is to present the indicators and decision rules for making a classification to a panel of clinical experts. Following a systematic process the panel decides if the protocol appears to group patients who require the same level of care. *Face validity*, as with content validity, is concerned about the relevance of the classification criteria to the intended goal. Expert judgement also is used but the process is more informal. Face validity has value for patient

classification because clinicians will be more likely to accept that the methodology underlying the funding is congruent with their clinical goals ((Giovannetti, 1979; Thomas & Ashcraft, 1991; Streiner & Norman, 1989).

Criterion-related validity refers to the extent to which an instrument produces results which correspond to other measures of the phenomenon of interest. Because there are no gold standards in patient classification, *predictive validity* (rather than concurrent validity) is generally tested. Predictive validity is mainly determined by assessing the extent to which grouped patients are homogeneous in their resource use and different from those in other groups (Taube, Lee & Forthofer, 1984). While resource use is not the only indicator of effective care, it has become the main focus when assessing performance of case mix systems. The actual cost of treating a patient is the ideal measure of resource use but, until Canadian case cost data are available, length of stay is used as a proxy for resource utilization (Halpine and Ashworth, 1994; Kiesler, Simpkins & Moreau, 1990).

Predictive validity is generally assessed by applying a case mix system or set of independent predictors retrospectively to a set of hospital discharges. Performance is measured by the reduction in LOS variance (RIV) that is achieved by using the predictor variables or the average LOS per case mix group to predict actual LOS. The goal is to have as large a reduction in LOS variance as possible with as few groups or variables as possible. RIV can be calculated by regressing actual LOS on mean length of stay per group. The resulting coefficient of determination or squared multiple regression (R^2) statistic indicates the amount of

variance in LOS accounted for by the case mix system or set of independent predictors

(Halpine & Ashworth, 1994). Homogeneity of groups, another aspect of predictive validity,

can also be measured by the coefficient of variation (C.V.), in which the standard deviation in

LOS is divided by the mean LOS per group to bring the measurement of LOS variation to a

common scale. It is desirable to have a small C.V. with as few groups as possible. When a

C.V. is greater than one, there is substantial variation in LOS within the group. An analysis of

variance F test assesses whether case mix groups are statistically distinguishable. If group

means are not significantly different, it may be that some groups are superfluous. Alternatively

within group variation may be so large that groups overlap substantially (Horn, Chambers,

Sharkey et al, 1989).

A serious limitation of tests of predictive validity in patient classification is that they are based

on current rather than ideal practice. Low predictive validity may indicate a poorly performing

case mix system or set of predictors, or it may result from trying to model practice that is

suboptimal (Giovannetti, 1979; Hopkins & Carroll, 1994). Another limitation relates to

statistical stability of predictive models and generalizability of findings. Choca, Peterson,

Shanley et al (1988) demonstrated a substantial "shrinkage" in predictive power or R^2 when a

customized model based on analysis of a specific set of discharges was applied to a different

patient sample. This occurs because the regression equation, in striving to explain the

maximum amount of variance in LOS, captures both true differences among subjects and

chance variance which is sample-specific. The amount of shrinkage decreases as the number of

observations per independent variable increases and as R^2 increases (Streiner, 1994).

To assess generalizability, researchers often employ split-sample techniques, using one portion of their data to develop a model (derivation sample) and the remaining cases for validation (validation sample) (Thomas & Ashcraft, 1991). This strategy is less practical when the sample is small. Many researchers report the "adjusted R^2" statistic as the performance measure because it attempts to measure the true variation in the dependent variable, exclusive of error variance (Streiner, 1994).

Tests of predictive validity should not be conducted on patients whose LOS is altered by participation in a research protocol or specialized program. Tests of predictive validity need to be repeated periodically to assess the continued relevance of a case mix system. Botz and colleagues (1993) recommend ongoing recalibration of classification systems and cost weights in order to respond to changes in standards of care and current practice.

Unanticipated Consequences of Case Mix Funding

The introduction of case mix funding has had a considerable impact on clinical practice. Mechanic (1996) noted that institution of the Medicare prospective payment system resulted in reduced use of resources, and organizational and behavioural change in the general health sector. In a study of almost 900,000 psychiatric discharges from V.A. hospitals before and after implementation of case based funding, Rosenheck and Massari (1991) found a decline in average length of stay and increase in number of episodes of care. Yet, if the funding methodology is not properly developed and monitored, individual patients may not be receiving appropriate care.

Case mix reimbursement exerts its influence by transferring risk for not staying within budget from the funder to the hospital. If patients in case mix groups are not sufficiently homogeneous or if the associated level of reimbursement is not appropriate, desired practice will not be rewarded and gaming may result. Underpayment can adversely affect patient care by pressuring hospitals to discharge patients earlier (ie., "dumping"), select higher functioning patients in order to reduce expenses (ie., "skimming"), decrease provision of services or cost shift by transferring patients to sectors with different funding sources (e.g., long term care). Overpaid hospitals have few incentives to improve efficiency (Jencks, Horgan & Taube, 1987; Taube, Thompson, Burns et al., 1985).

It is important to avoid structuring case mix groups so that small changes in treatments or diagnoses can shift a patient into a higher paying category. For example, the CIHI CMGs are constructed such that the compensation provided to hospitals for patients in a short stay CMG (LOS less than 6 days) is significantly less than for patients with the same diagnosis in hospital for seven days or more. It could be argued that this provides an incentive to keep short stay patients in hospital for at least six days.

It also is important to use criteria for defining case mix groups that are independent of care. Diagnoses that develop during hospitalization and procedures performed while in hospital, whether scheduled or not, are indices that are used in the DRG and CMG case mix systems to define more expensive categories of care. Yet, use of these indices can be problematic because the effect of inadequate treatment cannot be disentangled from the legitimate need to deploy

more resources in response to a deterioration in a patient's condition or poor response to appropriate treatment. A case mix system that uses in-hospital indicators can inadvertently reward poor treatment practices. A further concern is that treatment variables are more vulnerable to manipulation and, if included in a case mix system, can encourage the practice of "charting for dollars".

Reimbursement inequities are a particular concern when systematic differences exist among hospitals in the costs of treating similar patients. This problem is relevant to psychiatry which is characterized by a highly differentiated service system where patients with a complex illness are often referred to higher levels of care (Jencks, Horgan, Goldman & Taube, 1987; Wellock, 1995). Specialized psychiatric facilities are expected to serve a complex and often refractory patient population defined by eligibility criteria, referral routes and treatment programs. In contrast, acute care facilities respond to a broader range of need, and emphasize stabilization and community linkage. Numerous studies have shown that hospitals with specialized psychiatric services would be inadequately funded under prospective payment (Freiman, Mitchell & Rosenbach, 1987; Goldman & Sharfstein, 1987; Dada, White, Stokes et al, 1992; Wellock, 1995). Services for involuntary psychiatric admissions constitute another sector within mental health care with systematically higher treatment costs (Holley, Kulczycki & Arboleda-Florez, 1994).

Initiatives to Improve Psychiatric Classification Systems

Existing Data Items

Initial studies to improve psychiatric DRGs incorporated or examined the potential role of items routinely collected in hospital discharge data sets. Modest associations with LOS were found among variables such as sex, age, prior hospitalizations, involuntary commitment, primary diagnosis, presence of secondary diagnoses or complications, type of treatment (see Table 1A). When these variables were incorporated into case mix systems, they increased only modestly the predictive power of the initial psychiatric DRG (see Table 1B) and performance was still poor compared with surgical groupings (Taube, Lee & Forthofer, 1984; English, Sharfstein & Scherl, 1986).

TABLE 1A

Summary of Findings on Association Between Predictor Variables and LOS

Predictor variable	Result (Reference)
gender	male = > ↓ LOS (Tucker, 1996; Hadley, Culhane & McGurrin, 1992) no association (Bezold, 1996)
age	older = > ↑ LOS (Ashcraft, 1989; Fortney, 1996; Jayaram, 1996;) younger/older = > ↑ LOS (Stoskopf, 1991; Taube, 1984)
psychiatric diagnosis	schizophrenia continuum, affective disorders ↑ LOS (Kato, 1995; Hadley, 1992; Chang, 1991)
psychiatric comorbidity	↑ LOS (Fortney, 1996)
medical comorbidity	↑ LOS (Fortney, 1996; Lyons, 1995; Kiesler, 1990)
substance abuse	↓ LOS (Lyons, 1995; Jayaram, 1996)
previous admissions	↓ LOS (Fortney, 1996) ↑ LOS (Bezold, 1996; Creed, 1997)
ECT	↑ LOS (Bezold, 1996; Stoskopf, 1991; Creed, 1997)
involuntary	↑ LOS (Holley, 1994)

TABLE 1B

Performance of alternative systems/variable groups to classify psychiatric patients

Study	System	Main Predictor Variables*	Variance Reduction	
			Alternative (%)	DRG (%)
Mitchell, 1987	Disease staging (>21 grps)	psychiatric ICD_9 codes (principal & secondary Dx)	5.8-12.0	4.4-8.2
Taube, 1984	Alternative DRGs (22 grps)	Dx, type of trmt, age, marital status, legal status, discharge status, prior m.h.care, referral status	11.8	3.2-7.6
English, 1986		age, sex, discharge against medical advice, medical complication	7.28	5.6
Mitchell, 1987	Clinically related groups (16 grps)	psychiatric ICD_9 codes, age	5.4-12.3	4.4-8.2

*All variables are available in hospital discharge data sets.

Severity of Illness

Severity of illness measures the relative state of health of patients with similar conditions. In

1985 Gordon and colleagues combined DSM-III ratings of severity of psychosocial stressors

(Axis IV) and level of functioning (Axis V) with DRG groups and found prediction of LOS to

be significantly improved over DRGs alone. While Axis IV and V ratings are not consistently

and reliably recorded by treating physicians (Goldman, Taube & Jencks, 1987; Bezold,

MacDowell & Kunkel, 1996), Gordon's research gave support to the belief that incorporating a

measure of illness severity into a patient classification system would improve predictive

validity. The work of Marie Ashcraft and colleagues (1989) supported further this view. She convened an expert advisory panel to identify potential severity markers, but only retained items which were relatively reliable and "ungameable". The resultant data set of patient behavioural and functional characteristics was assembled using data obtained from the discharge abstract, the patient's medical record and provider interviews. While Ashcraft's classification system explained only 11.5% of LOS variation in patients with mental disorders, a measure of severity of symptomatology on admission (based on the GAF) was the strongest single predictor of LOS.

Susan Horn and colleagues developed a psychiatric severity of illness index - the PSOII (Horn, Chambers, Sharkey et al., 1989) which evaluated patients along seven patient dimensions - peak signs and symptoms; complications while in hospital; comorbidities; dependency on hospital staff; availability of social support; rate of response to therapy; and resolution of acute symptoms on discharge. The index was determined by specially trained raters through in-depth reviews of patient records. While PSOII-adjusted DRGs explained 40-54% of LOS variation, the PSOII was criticized for being costly to assemble, dependent on rater judgement, and based on indicators which could be measures of treatment ineffectiveness rather than patient severity. Horn has achieved improvements in predicting LOS but at the expense of reduced reliability and validity, and increased cost.

Since its initial development the PSOII has gone through a series of modifications. A second version, *the Computerized Psychiatric Severity Index (CPSI)* was based on a patient's mental

status, treatment history, physical condition and social situation (Stoskopf & Horn, 1992). The

rating, which ranged from one to four, was diagnosis-specific and based on objective

measures. In a study of 306 discharged patients with schizophrenia and affective disorders, the

CPSI alone explained 14% of variation in LOS. Patients in the least severe category stayed an

average of 15 days compared with 39 days for the most severe group. When the CPSI was

combined with other items in the patient's medical record (including type of insurance), the

resultant model accounted for 33% of variation in hospital LOS.

The CPSI has since been incorporated into a broader rating system which assesses severity in

all hospital discharges. The *Computerized Severity Index* (CSI) uses diagnosis-specific

indicators to assess severity of every ICD-9 diagnosis in a patient's chart. Development of the

indicators and the algorithms for calculating the severity index were guided by expert clinical

panels in each medical specialty area. Criteria are objectively defined, and patients are rated

using information supplied by a medical records abstractor after reviewing a patient's chart.

Data entry and generation of the severity rating is fully computerized. The rating is based

primarily on patient data from the following domains - mental status, psychiatric history, past

response to treatment and medical complications, with the specific indicators and criteria for

assessing severity varying per diagnosis. Treatment data are not used. The tool produces both a

discrete and a continuous rating. Social factors also are considered to influence resource use

but have been excluded from the CSI rating until more reliable measures can be found (Gurny,

1994). Published evaluations have assessed performance of the CSI system in rating surgical

and medical patients but have been confined to the United States and not included psychiatry

(Averill, McGuire & Manning, 1992; Horn, Sharkey, Buckle et al, 1991; Hopkins & Carroll, 1994).

Lyons et al (1995) developed and tested a Severity of Psychiatric Illness (SPI) index that can be completed retrospectively by chart review or prospectively by a clinician. This 12 item multidimensional instrument assesses four domains - reason for admission; complications to the psychiatric disorder; complications to treatment; and severity and persistence of illness. The SPI has been used in studies concerned with quality management, decisions to admit and prediction of length of stay, and has been shown to be a reliable and valid tool (Lyons, Colletta, Devens & Finkel, 1995; Lyons, O'Mahoney, Doheny et al, 1995; Lyons, OMahoney, Miller et al, 1997).

The Health of the Nation Outcome Scales (HoNOS) is a prospective tool developed in the United Kingdom to measure both case mix and outcome. Clinician raters assess each patient on 12 items using all the information available to them. Scale domains include behaviour, impairment, symptoms and social functioning. The tool was designed to meet criteria of brevity, adequate coverage of clinical and social functions, reliability, relationship to more established scales and sensitivity to improvement. Currently the HoNOS is being used in several pilot projects to match patient characteristics to level of care. A recent study found that HoNOS scores provided a better prediction of in-patient resource use in psychiatry than Healthcare Resource Groups, a DRG-like case mix system developed in the United Kingdom (Wing, Curtis & Beevor, 1996; McCrone & Phelan, 1994).

These various projects and tools demonstrate that multidimensional severity rating systems hold promise for improving performance of psychiatric classification systems.

Social-Environmental Factors

While social and environmental factors affect decisions about admission, treatment and discharge for a variety of health conditions, their influence is particularly pronounced in mental health (Richman, Boutilier & Harris, 1984; Cagle & Banks, 1986; McCrone & Phelan, 1994). Patients with psychiatric problems may be admitted because their living environment can no longer tolerate them or because they can no longer tolerate their living environment (Goldman, Taube & Jencks, 1987). Psychosocial crises are a primary cause for presentation to hospital emergency departments and crisis programs, and may lead to an admission. Inadequate housing, living alone, marital disruption and unemployment are examples of indicators of social disequilibrium that have been found to increase risk for admission (Streiner, Goodman & Woodward, 1975; Kelly & Jones, 1995) and readmission (Goering, Wasylenki, Lancee & Freeman, 1984), and a relationship between non-hospital supports and length of stay has been postulated (Mitchell, Dickey, Liptzin et al, 1987; Boyer, Olfson, Kellermann et al, 1995).

Despite many examples where social-environmental factors influence hospitalization experiences, these domains are difficult to define and measure, and few studies have evaluated their influence in patient classification systems. For example, the concept of social support has various meanings, including the number of people a person has contact with in a fixed time period, a person's perception of available supports, extent of reciprocity in a person's

relationships and number of conflictual relationships. In discharge datasets or medical records, patient social support is not consistently defined or reported. Primary data collection to gather this information is costly, requires training and monitoring to increase data reliability, and might be resisted by clinical staff.

The few studies that have incorporated social factors into patient classification have used simple measures or proxies.The role of social support in explaining LOS has been assessed using marital status, next of kin, living alone, living with a significant other, alone on admission, and change of home address following hospitalization (Leff, Swartz, Ghler et al, 1985; Horgan & Jencks, 1987; Chang, Brenner & Bryant, 1991; Fortney, Booth & Smith, 1996; Cyr & Haley, 1983; Creed, Tomenson, Anthony et al, 1997). While Fortney et al (1996), Chang et al (1991) and Cyr & Haley (1983) found that having social support modestly reduced hospital stay, Horgan & Jencks (1987) reported mixed results. The severity rating produced by the CPSI (Stoskopf & Horn, 1991) included a psychosocial domain based on chart information about a patient's living situation (i.e., alone, in chaotic disturbed family situation, homeless) and admission status (i.e., voluntary, involuntary). While the CPSI performed well, the contribution of the psychosocial domain to the severity rating were not reported and psychosocial indicators were excluded from the next version of the CPSI due to measurement problems (Gurny, 1994).

Social area studies have developed an approach for measuring social factors which may be better suited to patient classification systems. These ecological studies use geographic areas

such as census tracts as the unit of analysis and rely on existing indicators (usually census data). The underlying assumption is that measures of social conditions such as unemployment rate and per capita income can influence individual need for health care (Cagle & Banks, 1986). Social area studies have identified a number of socio-economic and demographic variables that predict inpatient psychiatric admission. Miller, Dear & Streiner (1986) found that admission rates were higher in downtown core areas where more people were single, renters and less educated. Richman and colleagues (1984) found that proportions of not married, low income and less educated people were positively associated with use of hospital services. Indicators of isolation (i.e., single person household, unmarried, old people living alone) and poverty (unemployment, lacking a car, unskilled worker) correlated with psychiatric admission rates in a study by Jarman, Hirsh, White et al (1992), with proportion unmarried having the strongest association. While aggregate level relationships may not pertain to individuals (the "ecological fallacy"), the findings of social area studies parallel those of other studies and offer an approach to social factors measurement which merits further examination.

CHAPTER 2

METHODOLOGY

Study Objectives

This retrospective chart review study rates severity of a sample of discharged psychiatric inpatients using the CSI system. Staff from the medical records departments of three Ontario hospitals participated in data collection and were trained to use the CSI severity rating system.

Objectives of the study were:

1. *To determine the feasibility of using the Computerized Severity Index (CSI®) to rate severity in psychiatric inpatients in Ontario.*

2. *To determine if the CSI severity measure predicts significantly more variation in patient length of stay than a subset of patient variables currently available in hospital discharge abstracts.*

3. *To specify a model that can predict length of stay using a parsimonious subset of study variables.*

Study Sites

The study was conducted at three hospital sites in the Greater Toronto area. The Clarke Institute of Psychiatry is a specialty facility which operates 103 beds and discharges some 950 patients each year. Sunnybrook Health Science Centre is a teaching hospital which operates 40 beds for psychiatric patients and discharges over 400 patients annually. Whitby Mental Health Centre is a provincial psychiatric facility that operates a 36 bed acute care unit with

approximately 300 discharges per year.

Sampling Frame

The sample was randomly drawn from adult psychiatric patients discharged during the 1994-95 fiscal year. The study used a stratified sampling approach, with each stratum represented by one hospital. The sampling frame was defined by the eligibility criteria set out in Table 2. Non-repeating patients were selected to maintain independence of observations within the sample. Excluded from the study were patients in programs where LOS is fixed or influenced by a research protocol or other special initiative. Also excluded were atypical cases as defined by CIHI, that is, outliers[1], sign-outs, transfers to and from other facilities, and deaths (CIHI, 1995). Cases where length of stay is interrupted are excluded because a full course of treatment is not completed. Outlier cases are excluded because treatment goals of providers tend to shift from an acute to a long term care approach for these patients. Atypical cases constituted 17% of all psychiatric discharges from acute care facilities in Ontario during October 1992 to September 1993 (CIHI, 1994). Similar exclusion criteria have been used in other studies (Ashcraft, Fires, Nerenz et al, 1989; Fortney, Booth & Smith, 1996; Thomas & Ashcraft, 1991).

Only discharges with a most responsible diagnosis of either major depressive disorder or

---

[1]   Outlier cases exceed the trim LOS for the assigned CMG. The trim point is derived from the distribution of cases within the CMG using the formula: TRIM = 3rd quartile + 2*(interquartile range). CMG trim LOS values for this study were based on the 1995 LOS distributions for Canadian acute care facilities (CIHI, 1995).

schizophrenia and other psychotic disorders were included in the study. These two diagnostic groups contain the most responsible diagnosis for approximately 40% of typical discharges from acute care facilities in Canada (CIHI, 1994) and account for a considerable portion of inpatient treatment expenditures (Fortney, Booth & Smith, 1996; Goree, 1994). While study eligibility criteria narrow the sample and limit generalizability of findings, use of a more homogeneous, smaller sample is common in a pilot study (Ashcraft, Fries, Nerenz et al., 1989; Horn, Chambers & Sharkey et al, 1989). If findings are positive, subsequent studies can be conducted on more heterogeneous samples.

Table 2

Sample Inclusion and Exclusion Criteria

Inclusion criteria:

* admission during FY94-95

* most responsible diagnosis of major depressive disorder or schizophrenia and other psychoses[2]

* age >= 18 yrs

* unique individuals

Exclusion criteria:

* LOS exceeding the following trim points

depressive mood disorder - with ECT:	106 days
depressive mood disorder - no ECT:	64 days
schizophrenia - with ECT	133 days
schizophrenia - no ECT	70 days

* signouts, transfers, deaths

* other discharges not home, ie., court assessments

* discharges from any program where LOS is pre-determined

Sample Size

To generate the study sample, chart numbers were randomly drawn from all eligible FY94/95 hospital discharges using criteria set out in Table 2. With approximately 13 independent variables in the study and separate predictive models being developed for each diagnostic group

(1) schizophrenia or other psychoses (ICD-9 295.0 - 295.9, 297.0-297.3, 297.8-297.9, 298.3-298.9);
(2) major depressive disorder (ICD-9 296.2, 296.3, 296.5, 296.9, 311)

(see Data Analysis Approach), a minimum sample of 130 cases per diagnostic subgroup was needed to provide sufficient power for the planned regression analyses (Norman & Streiner, 1994). Over sampling at each site ensured that, even with some losses of observations, an adequate sample size would be achieved.

Study Variables

A description of the study variables and their operational definitions follows and is summarized in Table 3. Where possible, definitions for study variables are consistent with the categories used in the discharge data set submitted by hospitals to CIHI.

Demographic variables

Age and **gender** of each case are recorded. Age is used in its natural form as a continuous variable while gender is converted to a dichotomous variable where 0 = male and 1 = female.

Diagnosis based variables

The study sample is restricted to patients who have a most responsible diagnosis of either schizophrenia, paranoid states and non-organic psychoses; or major depressive disorders (see Table 2 for ICD_9 codes).

Based on the ICD_9 codes in a patient's chart, four other measures are calculated for each patient. These include a count of the **number of medical diagnoses**; a count of the **number of psychiatric diagnoses;** and two dichotomous variables indicating the absence (=0) or presence

(=1) of a **secondary personality or substance abuse disorder.** Diagnosis-based variables have been used with some success in other studies to identify more severe patients (Choca, Peterson, Shanley et al, 1988; Fortney, Booth & Smith, 1996; Ashcraft, Fries, Nerenz et al, 1989).

Social-Environmental Factors

Although measurement of social environment is not a central concern of this study, performance of the severity ratings will be compared to a base model which includes three patient social-environmental variables. **Marital status, education level and source of income** are items reported routinely by Ontario acute care psychiatric units and speciality hospitals to CIHI [3]. To reduce the number of dummy variables representing these categorical data in the regression equation, the response categories were collapsed. Marital status was converted into a dichotomous measure corresponding to whether the patient was or was not in a marital/common law relationship. Financial support was reduced to two categories indicating whether or not the patient's primary income source was public assistance. Education was collapsed into two categories - "up to completed high school education" and "some post secondary training or more".

Severity

The Computerized Severity Index (CSI) measures the severity of the episode for which a patient is hospitalized, and yields clinically credible and reliable ratings (Hopkins & Carroll, 1994). The

---

[3]   While Whitby Mental Health Centre does not report to CIHI, these data are captured in the hospital clinical database.

CSI rating is a continuous measure with no pre-set maximum. Higher scores indicate increasing levels of severity. The rater begins by entering each ICD_9 diagnosis in a patient's chart into the CSI system. Based on these diagnoses, the system generates from 15 to 150 indicators that raters score from 1 to 4 (low to high severity) using chart information. The system combines the indicator scores using a non-linear algorithm to produce a patient's overall severity rating. Weighting is not applied to give more emphasis to certain indicators.

Each chart review yields three ratings. The admission severity is based on information available during the patient's first 24 hours in hospital, maximum severity is based on the most severe signs and symptoms observed during the hospital stay; discharge severity is based on final status at discharge. Maximum severity is intended to capture any deterioration in a patient's condition during hospitalization but a rise in severity also can reflect poor quality or inefficient care. Maximum severity has been found in several studies to be superior to admission severity in predicting LOS and costs of care (Stoskopf and Horn, 1992; Thomas & Ashcraft, 1991; Hopkins & Carroll, 1994).

In a study that has used the CSI system to rate non psychiatric patient groups, the severity rating was completed in 20 to 30 minutes by an experienced rater (Thomas & Ashcraft, 1991). The present study also assesses the feasibility (ie., rating and training time) of using the CSI system to rate patient severity.

One analyst from each site and the principal investigator (PI) attended a three day training

program offered by the CSI vendor, ISIS, Inc. to learn how to apply the CSI rating methodology.

Each analyst was responsible for rating the discharge episodes selected for the study from her

site. The PI served as the primary link between the raters and ISIS staff who worked with raters

to identify and solve problems as they arose.

Treatment

Inclusion of treatment variables in case mix systems is controversial due to potential incentives

for the care provider to alter treatment in order to obtain higher payments. To avoid this

possibility, developers of the CSI produced a severity rating system that is based solely on patient

variables and is independent of treatments provided. Yet, health services researchers have shown

that use of **ECT** has a profound influence on LOS (Stoskopf & Horn, 1992; Bezold, MacDowell

& Kunkel, 1996), and ECT is included in the most recent version of the psychiatric CMGs

(CIHI, 1994). ECT is included in this study to assess whether its strong relationship to LOS

remains, after patient severity and other patient variables are entered into the predictive model.

ECT is recorded as a dichotomous variable (0=did not use ECT; 1=used ECT).

Hospital Site

Numerous studies have found that **facility** is a powerful predictor of LOS in psychiatric patients

(Horgan & Jencks, 1987; Holley, Kulczcki & Arboleda-Florez, 1994; Dada, White, Stokes et al,

1992; Kiesler, Simpkins & Morton, 1990; Wellock, 1995; Taube, Thompson, Burns et al, 1985).

One explanation is that hospitals have unique "institutional practice styles". Another is that

hospital sectors differ systematically in their treatment mandates. For example, the in-depth

patient assessments and more aggressive therapeutic activities, including medications monitoring, that tertiary care facilities offer often require more time than the stabilization and linkage priorities of most acute care facilities. These different roles reflect the structure of the mental health system where more complex patients are often referred to higher levels of care. This study cannot test assumptions about sector differences because, with only one hospital participating from each sector, the range of practice within a sector is not represented. Therefore facility variables are entered first into the study model to control for any variability in LOS that results from site differences. Because there are three facilities in the study, site is represented by two dummy variables for Whitby and the Clarke respectively.

Length of Stay (LOS)

The dependent variable in the analysis, **LOS**, is the study indicator of resource use. The actual cost of treating a patient is a more direct measure of resource use but in Canada hospitals do not record case-specific costs. LOS has been shown in many studies to be a suitable proxy as it has a close correspondence to actual hospital costs for psychiatric patients (Horn, Chambers, Sharkey & Horn, 1989; Stoskopf and Horn, 1992; Mitchell, Dickey, Liptzin et al, 1987). Fisher & Altaffer (1992) caution that LOS is only an appropriate performance measure for relatively short stay patient populations where episodes of care are completed and variation in LOS is not unduly influenced by long stay outliers. Because this study excludes discharges with LOS outside the CIHI trim point, it was appropriate to use length of stay as the dependent variable and as a proxy for resource utilization (Halpine and Ashworth, 1994).

Table 3

Study Variables

Independent:
Site: Clarke, Whitby   (Sunnybrook is the excluded category)
Demographic: age, gender
Dx based: # medical, # psychiatric, personality Dx, substance abuse Dx
Social-environmental: marital, income, education
Severity: admission, maximum, discharge
Treatment: ECT
Dependent:
Resource use: LOS

Data Analysis Approach

The variables outlined in the preceding section form the study data set. All data analyses were conducted using SAS software package for statistical analyses. A description of each study objective and strategy for evaluation follows.

Objective 1:   To determine the feasibility of using the Computerized Severity Index (CSI) to rate inpatient severity among psychiatric patients in Ontario hospitals.

A primary determinant of feasibility is the cost of assembling the required data set, as reflected both in the time required to rate cases and in the time required to train and monitor raters (to

maximize reliability). Medical record abstractors were asked to record the time required to rate cases in a log as well as any problems encountered. All raters participated in a three day training. Reliability was assessed by comparing ratings produced by study raters and an independent CSI trainer using the same set of patient charts. Inter-rater agreement per indicator was calculated. The intention was that raters who did not meet a criterion level of performance would be provided with further instruction and a re-evaluation.

Objective 2:.   To determine if the CSI severity measure predicts significantly more variation in patient length of stay than a subset of patient variables currently available in hospital discharge abstracts.

Hierarchical multiple regression analyses were performed to measure the extent to which the severity ratings make a unique contribution, above discharge abstract variables, in predicting LOS. Prior to regression modelling, a number of preliminary steps were conducted.

Because other studies have shown that the strength of the relationship between patient variables and LOS varies by diagnostic grouping (Stoskopf & Horn, 1992; Ashcraft, Fries & Nerenz, 1989), a separate predictive model was built for each diagnostic group. This approach is consistent with the substantial differences in disease processes and practice requirements that exist across diagnostic groups (Brooke, Hudak & Finstuen, 1994) and with the design of the current CIHI CMGs in which diagnostic groupings define macro level categories.

To confirm that the two diagnostic groupings represent different populations, a series of **chi-**

square analyses and one way analyses of variance were conducted, comparing patients in the two diagnostic subgroups on socio-demographic, illness and utilization variables. Per diagnostic grouping a **Pearson correlation coefficient (r)** was calculated for all pairs of independent variables to examine inter-relationships and to identify where collinearity exists between variable dyads.

To begin to understand the relationship between each predictor variable and LOS, a **one way ANOVA** was performed for each independent categorical variable and a simple linear regression was run for each continuous independent variable. These bivariate analyses were run separately for observations in each diagnostic grouping. To examine the relative contribution of the independent variables to predicting length of stay, **hierarchical linear regression analyses** were conducted with the dependent variable LOS.

Linear regression develops an equation for the line that best fits all the data points with the least amount of error and assumes a linear relationship between predictor and dependent variables. An estimate of the unstandardized regression coefficient (b) is produced for each independent variable and measures the strength of its relationship to the dependent variable. Significance of b is tested by a t test. The squared multiple correlation or R^2 gives the percentage of variance in the dependent variable accounted for by its linear relationship with the predictor variables. In assessing the nature of the relationship it is important to consider the size of R^2, the size of the b coefficients and the standard error of the regression (a measure of how far the average dependent variable departs from its forecasted value). Assessing evidence of patterns in residual plots

indicates the appropriateness of using a linear model (Norman & Streiner, 1986; Achen, 1982).

Hierarchical linear stepwise regression introduces variables, either singly or in clusters, in an order assigned in advance by the researcher. The difference in R^2 between regression equations indicates how much variance is uniquely attributable to the additional variable or variable cluster, net of the other variables in the model (Brooke, Hudak & Finstuen, 1994). Significance of the difference can be assessed statistically by an F test (expressing the ratio of the additional variance explained by the new variable or set of variables to the residual error variance) or a t test (assessing the significance of the *b* coefficient if only one variable is added). It also is important to judge the clinical significance of the improvement by looking at the increase achieved in R^2 and the size of the *b* weight (Norman & Streiner, 1994).

Hierarchical methods were used to determine whether the CSI severity ratings improve capacity to predict LOS over using data already available in discharge abstracts or clinical information systems. Variables were entered in four steps (see Table 4) and full rank methods were used to solve the equations. Step 1 only included site variables (i.e., Clarke and Whitby) in order to identify and control for variance in LOS caused by site differences. Step 2 added patient socio-demographic and diagnosis-based variables (referred to hereafter as abstract variables) to assess the full contribution of this cluster to predicting LOS. Step 3 was repeated twice to assess individually the contribution of admission and maximum severity after controlling for abstract and site variables. ECT was added in Step 4 to assess whether use of this treatment variable significantly prolonged LOS beyond what could be predicted using available patient data.

Table 4
Independent variables in each step of the hierarchical stepwise regression analysis

Step	Independent variables
1	Clarke, Whitby
2	age, gender, number of medical diagnoses, number of psychiatric diagnoses, presence of personality disorder, presence of drug disorder, marital status, income source, education
3A	step 2 variables + admission severity
3B	step 2 variables + maximum severity
4	best model (3A, 3B) + ECT

Objective 3:   *To specify a model that can predict length of stay using a parsimonious subset of*

                  study variables.

One purpose of seeking a more parsimonious subset of independent variables is to increase the

accuracy of the prediction equation. This is achieved by increasing the ratio of cases to variables

which increases the precision of the coefficient estimates. A second purpose relates to

functionality. A reduced set of predictors is easier to understand and use in defining case mix

categories (Streiner, 1994; Fox, 1991).

For each diagnostic subgroup, the model with the highest R^2 from the hierarchical regression

analyses was rerun and several strategies used to reduce the number of variables in the model. In

forward stepwise selection the procedure first selects the variable that best predicts LOS, then

adds the variable which, in combination with the first one, now best estimates LOS. This process

continues until adding another variable does not improve the predictive power of the equation above a preset criterion. *Backward selection* begins with all of the variables in the equation, taking them out one at a time until removing another one would produce an unacceptably large drop in predictive ability. Stepwise approaches do not produce unique solutions because they depend on the order that variables are entered. After the first step, the approach is constrained to dyads which include the first variable entered, and after the second step, to triplets involving the first two variables, and so on. Starting with a different first variable in a forward stepwise process could produce a different solution. However, if forward and backward methods produce similar results, it is more likely that there is a single best solution rather than multiple contenders. Stepwise solutions cannot be used for explanation because variables that have a strong explanatory role may not enter the equation if they are highly correlated with independent variables already in the model (Streiner, 1994; Freund & Little, 1991).

Because the solution produced by stepwise methods are not necessarily unique, an **all subsets analysis** also was conducted. This approach produces, for each subset of independent variables of a given size, the R^2 that different combinations of the independent variables produce. Several statistics can aid in selection of the subset model that accounts for the most variance without introducing unnecessary error. Mallows C(P) statistic is a measure of the total squared error for a subset model with "p" independent variables. C(P) decreases when unnecessary independent variables are removed but increases when relevant variables are excluded. A good model has C(P) close to or below p+1. If C(P) is larger than p+1, there is evidence of an incompletely specified model (i.e., missing variables). If C(P) is much less than p+1, the model is

overspecified, that is, contains too many variables. The difference between R^2 and adjusted R^2 also can aid model selection. Adjusted R^2 approaches R^2 as the ratio of cases to predictors increases and as R^2 increases. A smaller difference indicates less reliance on chance variation in the sample and greater generalizability of the model ((Freund and Little, 1991; Fox, 1991; Streiner, 1994).

CHAPTER 3

RESULTS

Reliability of CSI rating System

Raters at all three sites participated in inter-rater reliability checks. Each rater sent photocopies of

three charts (first, second and fifth charts rated) along with her severity ratings to the CSI trainer

who rated the charts and compared results. Some rating discrepancies occurred due to differences

in units of measurement for diagnostic tests between the study hospitals and CSI. When these

indicators were omitted, agreement levels exceeded 95% for all raters. After a strategy for

resolving measurement unit differences was developed, inter-rater reliability was assessed on

another chart at each site, and agreement levels exceeded 95%. Due to the short study period,

ongoing inter-rater reliability checks were not conducted and rater drift was not assessed.

Sample Description

Representativeness

While each site was expected to rate 160 discharged patients, only the Clarke Institute met this

goal. Due to organizational changes in their medical records department, the Sunnnybrook site

only rated 120 patients. Whitby reviewed 184 cases but the drawn sample included 90 ineligible

cases - 69 discharges from non-acute units and 21 cases that were not discharged home. As all

site samples included some ineligible observations (see Table 2), the final sample sizes are 157

(Clarke), 88 (Whitby) and 110 (Sunnybrook). The total study sample numbers 355 observations,

including 188 in the depressive disorders subgroup and 167 with psychotic disorders. This

sample size is smaller than expected but sufficient for the planned analysis.

To assess representativeness, the study sample was compared to all eligible hospital discharges at each site on proportion of females, mean age and mean length of stay. Comparisons were made within diagnostic grouping per site, and sample values that differed from the population value by more than 10% (of the population value) were flagged (see Table 5).

Using this 10% criteria, the Sunnybrook sample does not differ from the eligible hospital discharge population on any of the three variables. The Clarke sample is similar on age and ALOS but has a lower proportion of females in the depressive disorders subgroup. Differences in the Whitby sample are more substantial, with the sample ALOS exceeding the population ALOS by 37% in the psychotic disorders subgroup, and 16% in the depressive disorders subgroup. The Whitby sample also has a higher proportion of females in both diagnostic groups which might account for the longer sample ALOS. Females have been linked with longer hospital stays in other studies (Tucker & Brems, 1993).

No practices could be identified that would have resulted in an over sampling of females or of longer stay patients. Inadequate representation of short stay male patients would be a concern in a study whose objective is to link patient characteristics with LOS. Fortunately Whitby patients constituted only 25% of the study sample, and the other sites provide adequate representation of the short stay male subgroup.

Table 5
Comparison of sampling frame[4] and study sample per site

Patient char.	Clarke FY 94/95				Whitby FY 93/95				Sunnybrook FY94/95			
	Total discharges (n=421)		Sample (n=157)		Total discharges[5] (n=183)		Sample (n=88)		Total discharges (n=267)		Sample (n=110)	
	Sch	Dep	Sch n=76	Dep n=81	Sch	Dep	Sch n=37	Dep n=51	Sch	Dep	Sch n=54	Dep n=5
Female (%)	33%	63%	30%	53%	42%	48%	49%	55%	47%	64%	52%	64%
Av. age	35	44.5	38	48.2	39.1	38.1	39.8	40.5	42.7	49.9	41.3	49.7
ALOS	20.3	27	20.9	26.8	19.7	21.0	26.9	24.3	20.8	22.3	21.6	21.1

Characteristics of Diagnostic Subgroups

Tables 6A and 6B describe and compare each diagnostic subgroup in the sample on all study variables using chi square analyses and analyses of variance. Characteristics of patients indicate a very disabled population and are consistent with descriptions reported in other studies of populations in treatment for schizophrenia (Boyer, Olfson, Kellermann et al, 1995) and for depression (Fortney, Booth & Smith, 1996; Stoskopf & Horn, 1991). One exception is rates of secondary substance abuse disorders which are lower than expected in both subgroups.

As would be expected, subgroup profiles are significantly different. Regarding comorbidity, patients with depressive disorders are significantly more likely to have at least one additional diagnosis (medical or psychiatric). About 50% of the depressed subgroup have a medical

---

4   all hospital discharges in specified period that meet study eligibility criteria.

5   from Short Term Assessment and Treatment unit

diagnosis or more than one psychiatric diagnosis, compared with about one quarter of those with a psychotic disorder. The frequency of secondary personality disorders also is higher in the depressed group than the psychotic group (33% vs 8%). Regarding social characteristics, the depressed subgroup is more likely to be older (average age of 47 vs 40 years), female (57% vs 41%) and married (32% vs 11%), and is less likely to receive public assistance (41% vs 69%). Consistent with clinical practice, patients with depressive disorders are more likely to receive ECT while in hospital.

In both subgroups admission and maximum severity ratings range from 0 to about 100, with score distributions slightly skewed to the right. Discharge severity scores range from 0 to 65 and, as would be expected, are highly skewed to the right. While mean discharge scores are 10.1 (sd=10.3) and 8.8 (sd=9.3) for psychotic and depressive disorder subgroups respectively, 5% of scores exceed 26 in both groups. Patients with psychotic disorders have significantly higher severity ratings than depressed patients on admission and reach a higher maximum severity during their hospitalization, but at discharge do not differ in level of severity.

Despite these many differences, the subgroups do not differ on average length of stay.

Table 6A indicates the difficulty of consistently obtaining social-environmental data from discharge abstracts or medical charts. About 21% of cases lacked required information on income source or educational level. When cases with complete and missing data were compared on all other study variables using chi-square analysis and ANOVAs, differences were only

significant with respect to site. Some 45% of Sunnybrook cases were missing information on at least one of these indicators whereas the Clarke and Whitby were missing 9% and 11% respectively. It appears that Sunnybrook is less consistent than other sites about collecting this data although the problem could not be further isolated to a specific program or unit (i.e., by patient age or diagnostic profile).

In SAS, multiple regression procedures deal with missing values through casewise deletion. If education and income were retained in model building, the sample would be reduced by about 21% overall and 45% for Sunnybrook, thus substantially reducing the power of the analysis. Other options for managing missing data include: (1) imputation; (2) excluding variables with missing data from model building. Imputation was rejected because it can introduce new sources of error or bias (Norman & Streiner, 1994). Instead, income and education were excluded from subsequent regression analyses. However, in a post-hoc analysis on the reduced data set of complete cases, the impact of adding income and education to the final model was assessed. As there is no evidence of bias in the reduced sample, this analysis is appropriate but weakened by a smaller sample size.

Table 6A
Demographic and Clinical Characteristics of Study Sample by Diagnostic Group

Variable	Psychotic n=167+	Depressive n=188+	χ^2 (df)	Pr (Grps same)
Dx- based variables				
# med. Dx (%)				
0	74.9	54.3		
1	15.6	19.7	24.4(3)	0.001
2	8.4	13.8		
3+	1.2	12.2		
# psy. Dx (%)				
1	70.6	49.5	24.4 (2)	0.001
2	24.0	28.7		
3+	5.4	21.7		
Pers. dis (%)				
yes	8.4	33.0	31.8 (1)	0.001
no	91.6	67.0		
Drug disorder (%)				
yes	9.0	13.3	1.6 (1)	0.20
no	91.0	86.7		
Treatment				
ECT (%)				
yes	1.8	13.8	17.1 (1)	0.001
no	98.2	86.2		
Socio-demographic				
Gender (%)				
female	41.3	56.9	8.6 (1)	0.003
male	58.7	43.1		
Marital status (%)				
married	10.8	32.1	23.3 (1)	0.001
not mar.	89.2	67.9		
Income Source (%)				
public	68.5	41.1	23.6 (1)	0.001
not public	31.5	58.9		
(missing=41)	(n=146)	(n=168)		
Education (%)				
<= secondary	57.6	46.4	3.9 (1)	0.05
post secondary	42.3	53.6		
(missing=45)	(N=144)	(N=168)		
Age (x, sd, range)	39.5 (10.8, 21-71)	46.6 (17.8, 20-100)	F=21.0 df=1,353	0.0001

+ Sample sizes per variable except where missing values are indicated

Table 6B
Severity of Study Sample by Diagnostic Group

Variable	Psychotic n=167[+]	Depressive n=188[+]	F value (df1, df2)	Pr (Grps same)
Severity				
CSI_dis (mean, sd, range)	10.1[*] (10.3, 0-65)	8.8[+] (9.3, 0-51)	1.6 (1, 353)	0.21
CSI_adm (mean, sd, range)	37.4 (18.0, 0-95)	29.1 (18.0, 0-96)	13.6 (1, 353)	0.0001
CSI_max (mean, sd, range)	41.6 (19.9, 0-103)	35.0 (20.5, 0-100)	9.5 (1, 353)	0.002
Utilization				
LOS	22.5 (16.5, 1-70)	24.4 (16.5, 1-81)	1.3 (1, 353)	0.54

+ Sample sizes per variable except where missing values are indicated
[*]   skew>   1.5

Variable Correlations

Within each diagnostic subgroup pairwise correlations were calculated (see Tables 7A and 7B) in

order to better illuminate relationships among pairs of study variables and as an initial step in

assessing whether collinearity might be a concern. Highly correlated dyads need to be identified

because collinearity among predictor variables can result in coefficient estimates that have large

standard errors and are unstable (Fox, 1991).

Of note (and expected) in this data is the very high correlation between the admission and

maximum severity ratings (r>0.90, p<.0001). Published evaluations of the CSI have assessed the

association between each of these two ratings and LOS separately due to their high inter-

correlation (Stoskopf & Horn, 1992; Thomas & Ashcraft, 1991). In order to benefit from the

information that both ratings hold, the increase in severity between admission and maximum

ratings was computed and interpreted as an indicator of patient deterioration during hospitalization. As would be expected, the distribution of this difference score was highly skewed to the right, in part because there is no increase in severity in over 50% of cases. Difference scores ranged from 0 to 34 (mean=4.3, sd=6.3, skew=2.1) and from 0 to 49 (mean=5.9, sd=9.1, skew=2.2) in the psychotic and depressive subgroups respectively. Mean increases in severity did not differ significantly between groups. In both subgroups, the difference score is significantly correlated with maximum but not admission severity. In model building, the contributions of the admission, maximum and difference severity ratings are assessed separately. Then the predictive role of the difference score is assessed, controlling for admission severity.

Among the diagnosis-based variables, presence of substance abuse (DRUG) and personality disorders (PERS) are highly correlated with number of psychiatric disorders (PSY) (r>.0.45, p<.0001) in both diagnostic subgroups. In keeping with Fox's suggestion (1991) of model respecification as a strategy for coping with collinearity, PSY is excluded from the regression analyses as DRUG and PERS can adequately represent the influence of psychiatric comorbidity. This solution is conservative as the linear relationship between independent variables must be very strong before collinearity seriously degrades the precision of estimation (Fox, 1991).

The significant negative correlations between admission and maximum severity measures and the Clarke (dummy site variable) are unexpected, given the notion that specialty facilities usually treat a more difficult case mix (Goldman & Sharfstein, 1987). However, because the Clarke

sample differs from the others on several characteristics associated with severity (i.e., the Clarke has a lower rate of medical comorbidity and fewer females), the relationship between the Clarke and admission severity was re-assessed in a regression analysis, controlling for these possible confounders. Although the strength of the association decreased in the psychotic disorders subgroup, admission severity remained significantly lower at the Clarke than the other two sites in both subgroups. Further investigation of this unexpected finding could include reassessing the inter-rater reliability of the tool and examining clinical priorities and practices of the Clarke compared with the other sites.

The strong correlation between ECT and site in the depressive disorders subgroup reflects substantial differences in rates of use of this treatment which are 27%, 7% and 0% at the Clarke, Sunnybrook and Whitby respectively. At Whitby, there was no central ECT service available during the study period and individual psychiatrists were responsible for ECT administration. At the other two sites, the widely varying rates of use of ECT reflect different institutional and physician practice patterns. In the psychotic disorders subgroup, only three patients received ECT so rates of use are not a concern.

There are a number of significant pairwise correlations between study variables ranging between 0.20 -0.30. While Fox (1991) demonstrates that inter-regressor correlations must approach 0.9 before there is a substantial increase in the standard error of the b estimate, precision of coefficient estimates for these variables were monitored during model building.

Correlations among patient variables for subgroup with psychotic disorders (n=167)

Var (n)	med	psy	pers	drg	ECT	sex	mar	inc	edu	age	adm	max	dis	diff	Whit	Cl
med (167)	---	0.18	0.03	0.07	0.06	0.20*	0.23*	-0.18	-0.02	0.23*	0.19	0.29**	0.04	0.37***	0.19	-0.37***
psy (167)		---	0.45***	0.50***	-0.08	0.08	-0.04	-0.02	0.04	-0.02	0.11	0.10	0.05	0.00	0.10	-0.1?
pers (167)			---	-0.02	-0.04	0.05	0.03	0.06	-0.06	-0.19	0.07	0.02	0.02	-0.09	0.05	0.03
drug (167)				---	-0.04	0.03	-0.04	-0.05	0.02	0.01	0.08	0.07	0.02	-0.01	0.13	-0.12
ECT (167)					----	0.07	-0.05	0.08	-0.13	0.15	0.04	0.02	-0.05	-0.06	-0.07	0.06
sex (167)							0.18	-0.16	0.04	0.13	0.25*	0.25*	-0.09	0.07	0.08	-0.21*
mar (167)								-0.26*	0.14	0.02	-0.04	0.01	-0.07	0.15	0.28**	-0.16
inc (146)									-0.23*	-0.01	-0.09	-0.10	0.07	-0.07	-0.02	0.11
edu (144)										-0.13	-0.06	-0.03	-0.08	0.06	-0.13	-0.03
age (167)											0.04	0.06	-0.03	0.08	0.02	-0.12
adm (167)												0.95***	0.34***	0.13	0.10	-0.25*
max (167)													0.40***	0.43***	0.09	-0.30***
dis (167)														0.28**	-0.06	-0.16
diff (167)															-0.01	-0.25**

Variable code book:

mar (married):	inc (on public assistance)	sex:	ECT:	per (personality disorder):	drg (substance abuse disorder):	edu (education):
0=not	0=no	0=male	0=no	0=no	0=no	0= <=secondary
1=married	1=yes	1=fem	1=yes	1=yes	1=yes	1= > secondary

severity scores:
adm = admission
max = maximum
dis = discharge
diff = maximum - admission

med = # medical Dx

*** = p<= .0001
** = p<= .001
* = p<=.01

Table 7B
Correlations among patient variables for subgroup with depressive disorders (n=188)

Var (n)	med	psy	pers	drg	ECT	sex	mar	inc	edu	age	adm	max	dis	diff	Whit	C
med (188)	---	-0.06	-0.13	0.02	0.08	0.13	0.12	0.02	-0.03	0.33 ***	0.21 *	0.34 ***	0.30 ***	0.35 ***	-0.02	-(
psy (188)		---	0.69 ***	0.49 ***	-0.14	-0.29 ***	-0.06	0.10	-0.12	-0.13	-0.06	-0.04	-0.04	0.02	0.11	-(
pers (188)			---	0.09	-0.12	-0.24 *	-0.14	0.08	-0.09	-0.19 *	-0.18	-0.21 *	-0.15	-0.12	0.11	0
drug (188)				---	-0.11	-0.23 *	-0.10	0.05	-0.05	0.02	0.01	0.06	0.07	0.11	0.04	0.
ECT (188)					----	0.16	0.22 *	0.00	0.01	0.29 ***	0.02	0.08	0.04	0.14	-0.24 **	0. *
sex (188)							0.09	-0.03	0.00	0.10	0.15	0.18	0.03	0.09	-0.02	-(
mar (188)								-0.30 ***	0.12	0.28 ***	0.08	0.07	0.09	0.01	-0.01	-(
inc (168)									-0.08	-0.25 **	0.16	0.14	0.11	0.00	-0.07	0.
edu (166)										0.01	0.01	0.05	0.11	0.10	-0.11	-(
age (188)											0.06	0.11	0.15	0.14	-0.22 *	0.
adm (188)												0.90 ***	0.55 ***	0.05	0.09	-(*
max (188)													0.59 ***	0.49 ***	0.00	-(*
dis (188)														0.26 **	-0.18	-(*
diff (188)															-0.17	-(
Whit (188)																

See Table 7A for variable codebook and significance levels.

Predicting Length of Stay - Bivariate Analysis

Tables 8A and 8B present analyses of bivariate associations between predictor variables and LOS using analysis of variance and univariate regressions. Only associations with p values of <= 0.01 are reported as significant.

Among those with psychotic disorders, medical comorbidity increases LOS, and females stay significantly longer than males (26.5 days vs 19.7 days). Admission, maximum and difference severity are all positively associated with LOS. Psychiatric comorbidity is not associated with LOS, nor are social-demographic indicators or use of ECT.

In the depressive disorders subgroup, a different set of predictors are associated with LOS. Older patients stay longer, as do those who get ECT (36.3 days with ECT vs 22.5 days without ECT). Among severity measures, only the difference score significantly predicts LOS, with greater in-hospital deterioration associated with prolonged LOS. Surprisingly, admission severity is not linked to LOS. Comorbidity of any kind does not prolong LOS and, with the exception of age, socio-demographic variables do not predict LOS.

Dx- based variables	ALOS, sd	F value (df1, df2)	P
Med. comor (n) 0　　(125) 1　　(26) 2　　(14) 3+　(2)	20.0, 16.0 28.8, 18.0 32.9, 12.1 25.0, 5.7	4.33 (3, 163)	0.006
Psy. comor (n) 1　　(118) 2　　(40) 3+　(9)	21.9, 17.0 23.9, 15.6 24.1, 14.8	0.28 (2, 164)	0.759
Pers. dis no　　(153) yes　(14)	22.3, 16.4 23.1, 14.8	0.14 (1, 165)	0.707
Drug disorder no　　(152) yes　(15)	22.4, 16.7 23.1, 14.8	0.89 (1, 165)	0.886
Treatment	**ALOS, sd**	**F value (df1, df2)**	**P**
ECT no　　(164) yes　(3)	22.5, 16.6 19.0, 9.0	0.13 (1, 165)	0.714
Socio-demographic	**ALOS, sd**	**F value (df1, df2)**	**P**
Gender male　　(98) fe male　(69)	19.7, 15.9 26.5, 16.6	7.10 (1, 165)	0.009
Marital status married (18) not mar (149)	27.9, 16.7 21.8, 16.4	2.23 (1, 165)	0.138
Income Source not public (46) public　(100) missing　(21)	26.3, 14.5 21.1, 17.7	3.14 (1, 144)	0.078
Education <= secondary　(83) post secondary　(61) missing　(23)	23.3, 17.7 22.6, 15.9	0.06 (1, 142)	0.800
Age	0.003　(R^2)	0.44 (1, 165)	0.509
Severity	**R^2**	**F(df1, df2)**	**P**
CSI_dis	0.009	1.51 (1, 165)	0.220
CSI_adm	0.081	14.61 (1, 165)	0.0002
CSI_max	0.112	20.80 (1, 165)	0.0001
CSI_diff	0.057	10.00 (1, 165)	0.002

Diagnosis-based variables	ALOS, sd	F value (df1, df2)	P
Med. comor (n) 0　　　(102) 1　　　(27) 2　　　(26) 3+　　(23)	23.8, 17.1 23.5, 14.8 23.9, 16.0 29.7, 16.6	0.88 (3, 184)	0.453
Psy. comor (n) 1　　　(93) 2　　　(54) 3+　　(41)	25.0, 16.8 24.4, 17.3 23.1, 14.7	0.19 (2, 185)	0.829
Pers. dis no　　(126) yes　　(62)	24.9, 17.0 23.5, 15.4	0.29 (1, 186)	0.590
Drug disorder no　　(163) yes　　(25)	24.6, 16.7 23.4, 14.9	0.11 (1, 186)	0.745
Treatment			
ECT no　　(162) yes　　(26)	22.5, 15.1 36.3, 19.7	17.06 (1, 186)	0.0001
Socio-demographic			
Gender 　female　(107) 　male　　(81)	24.2, 15.7 24.7, 17.5	0.05 (1, 186)	0.815
Marital status 　married　(60) 　not mar　(127)	26.7, 16.6 23.5, 16.4	1.51 (1, 186)	0.220
Income Source 　public　　　　(69) 　not public　　(99) missing　　　　(20)	26.9, 17.1 22.9, 16.1	2.40 (1, 166)	0.124
Education 　<= secondary　　(77) 　post secondary　(89) missing　　　　　(22)	25.1, 14.7 24.8, 17.5	0.02 (1, 164)	0.899
Age	0.083　　(R^2)	16.84 (1, 186)	0.0001
Severity	R^2	F (df1,df2)	p
CSI_dis	0.0002	0.04 (1, 186)	0.839
CSI_adm	0.001	0.11 (1, 186)	0.741
CSI_max	0.024	4.62 (1, 186)	0.033
CSI_diff	0.092	18.86 (1, 186)	0.0001

Predicting Length of Stay - Hierarchical Multiple Regression Analyses

Tables 9A and 9B present results of the hierarchical multiple regression analyses. Site variables, entered in the first step, accounting for approximately 2% of variation in LOS. The patient variables typically available in discharge abstracts are entered in the second step and account for an additional 10% of LOS variation in the depressive disorders group and 7% among those with psychotic disorders. The third step is repeated three times, to assess separately the effects of admission, maximum and difference severity scores on LOS, net of site and discharge abstract variables.

In the psychotic disorders subgroup, both admission and maximum severity ratings significantly improve prediction of LOS, with maximum severity (Model 3B) realizing the largest increase in R^2 over the Step 2 Model (almost 8% increase). The coefficient estimate for the maximum severity rating indicates that every increase of 10 points is associated with an additional 2.5 days in hospital (95% confidence interval 1.2-3.7 days[6]). The combination of admission and difference severity (Model 4) marginally improves prediction over using admission alone ($p<=0.018$). ECT is not added to the model as only three patients received this treatment. This analysis indicates that, for those with a psychotic disorder, a severity measure which incorporates both patient status on admission and changes during hospitalization is most strongly predictive of length of hospital stay.

---

[6]   A z value of 1.96 was used to calculate all confidence interval estimates reported in this chapter as the t distribution approximates the normal distribution for n > 120.

In the depressive disorders subgroup, admission severity does not increase capacity to predict LOS after controlling for other variables. Maximum and difference severity both significantly improve the model, but the difference rating is far superior, accounting for almost 10% of variation in LOS. In this subgroup, it is not a patient's severity on admission but rather an escalation in severity after admission that distinguishes patients on resource use. As the difference coefficient of 0.62 in Model 3C indicates, after controlling for the other variables, every 10 point rise in the patient severity rating after admission prolongs LOS an average of 6.2 days (95% confidence interval 3.6-8.8 days). Combining admission and difference severity (Model 4) improves prediction over admission alone but not over the difference score alone.

The last step introduces the treatment variable ECT which marginally improves upon the predictive power of Model 3C (incremental $R^2 =2\%$, p=.021). The strong bivariate relationship between ECT and LOS ($R^2 =0.08$) is substantially reduced after including severity and other patient variables in the model. This occurs, in part, because ECT is correlated with being older, a variable already in the model that is associated with longer LOS. Use of ECT adds an estimated 8.4 days to hospital tenure but the 95% confidence interval is wide (1.2 days to 15.6 days) due to the small sample of ECT users - only 26 or 14% of individuals with depression received ECT.

These models assume a linear relationship between predictor variables and LOS. Two sets of analyses were conducted to verify this assumption. The residual LOS (actual - predicted) was plotted against the predicted LOS for each model in the hierarchical analyses. There were no patterns present that would suggest presence of a non-linear relationship. In addition, bivariate

relationships were examined for independent variables where a non-linear relationship with LOS was a possibility. For example, age has been shown to have a curvilinear relationship to LOS in some studies, with younger and older age groups staying longer (Stoskopf & Horn, 1992; Taube, Thompson, Burns et al, 1985). Thomas and Ashcraft (1991) noted that costs sometime increase faster than severity and assessed nonlinear prediction models in a sample of hospital discharges that did not include psychiatric patients. In the present study severity ratings and age were plotted against LOS but non-linear trends were not suggested in the data.

In both subgroups adjusted R^2 estimates fall below R^2 estimates, suggesting that the models may be over-specified. In the next section several different methods of solving regression equations are applied to create a more accurate predictive model.

Table 9A
Hierarchical multiple regression analyses to predict patient length of stay

Psychotic disorders subgroup (n=167)

Independent variables/variable blocks	R^2	adj R^2	b, se(b) - for bracketed var	T: $b=0$	Prob > ITI
1 Site*	0.021	0.009	-------	------	-------
2 Site + abstract**	0.099	0.054	-------	-------	-------
3A Site + abstract + (adm)	0.152	0.104	0.226, 0.072	3.14	0.002
3B Site + abstract + (max)	0.173	0.125	0.248, 0.067	3.75	0.0003
3C Site + abstract + (diff)	0.135	0.085	0.552, 0.220	2.53	0.012
4 Site + abstract + adm + (diff)	0.182	0.130	0.508, 0.213	2.38	0.018

*   Site includes two dummy variables - Whitby, Clarke.
**   Discharge abstract variables include age, sex, marital status, medical comorbidity, presence of personality or drug disorder.

Table 9B
Hierarchical multiple regression analyses to predict patient length of stay

Depressive disorders subgroup (n=188)

Independent variables/variable blocks	R^2	Adj R^2	b, se(b) - for bracketed var	T: $b=0$	Prob > ITI
1 Site*	0.021	0.011	-------	-------	-------
2 Site + abstract**	0.125	0.086	-------	------	-------
3A Site + abstract + (adm)	0.127	0.083	0.046, 0.069	0.66	0.509
3B Site + abstract + (max)	0.159	0.116	0.167, 0.063	2.67	0.008
3C Site + abstract + (diff)	0.221	0.182	0.618, 0.132	4.47	0.0001
4 Site + abstract + diff + (adm)	0.225	0.181	0.064, 0.066	0.97	0.332
5 Site + abstract + diff + (ECT)	0.245	0.202	8.424, 3.612	2.33	0.021

*   Site includes two dummy variables - Whitby, Clarke.
**   Discharge abstract variables include age, sex, marital status, medical comorbidity, presence of personality or drug disorder.

Predicting Length of Stay - Maximizing Parsimony and Accuracy

Both forward and backward stepwise and all subsets procedures are used to identify the subset of

study variables which maximizes predictive capacity while improving accuracy and parsimony.

The procedures are applied to the predictive model in each hierarchical regression analysis

(Tables 9A and 9B) that meets the following criteria:

- Model R^2 is the highest of all models in the current step, and significantly higher than models in preceding steps (using significance level = 0.01).

- R^2 for the model in the following step is not significantly higher.

Site variables are excluded from model building in this phase because case mix systems are

designed to be site independent. While it is likely that prediction of LOS would improve by

including site, the goal of this analysis is to identify a subset of patient variables that have

potential for defining case mix categories.

The following criteria aided selection of the final model:

- maximize adjusted R^2;

- minimize mean square error (MSE) of predicted LOS

- model C(P) approaches p+1 where p = number of independent variables.

Sections in earlier chapters (i.e., on Assessing Case Mix Systems - Chapter 1 and Data Analysis

Approach - Chapter 2) explain why these criteria are used.

Psychotic Disorders Subgroup

Stepwise and all subsets methods were applied to Model 3B which includes discharge abstract data and maximum severity rating. Forward and backward solutions produced the same two variable model containing **maximum severity** and **medical comorbidity**. The all subsets analysis and regression diagnostics confirmed this selection. Adjusted R^2 reached 13%. Model parameters indicate that patients stay an extra 2.4 days for every increase of 10 points in maximum severity (95% CI=1.2 to 3.8 days). Those with a medical problem stay an average of 3.7 additional days for each medical diagnosis present (95% CI=0.9 to 8.3 days), controlling for maximum severity. Because medical symptomatology is represented in the severity rating, it is surprising that medical comorbidity still contributes to the model after controlling for severity. This finding raises questions about the structure of the CSI, a matter which is further addressed in the Chapter 4.

Depressive Disorders Subgroup

Stepwise and all subsets solutions were applied to Model 3C which includes discharge abstract data and difference severity. Forward and backward stepwise methods produced the same two variable solution (including **difference severity** and **age**) which achieved an adjusted R^2 of 15%. An all subsets analysis confirmed that this solution is superior to all other two, three and four variable models, with a high adjusted R^2, low MSE and C(P) value close to p+1. According to model parameters, an increase of 10 points in a patient's severity after admission adds an average of 4.8 days to LOS (95% CI=2.4 to 7.2 days). Older patients stay longer, with each decade of life adding 2.4 days to LOS (95% CI=1.0 to 3.8 days), controlling for severity increase.

Consistent with the interest of this study in learning more about the role of ECT in predicting

LOS when combined with patient severity and other variables, this process was repeated on

Model 3C + ECT. As with the earlier analyses, the solutions produced by the forward and

backward stepwise methods converge, with the final model containing three independent

variables that include **ECT** as well as **difference severity** and **age**. Adjusted R^2 nears 18%

compared with 15% achieved in the preceding model without ECT. The all subsets analysis and

regression diagnostics indicate that no other models perform better. Parameter estimates indicate

that having ECT prolongs LOS by an average of over 9 days (95% CI=2.9 to 15.9 days), holding

other variables constant.

Effect of Income and Education on LOS

As noted earlier, two variables with high rates of missing data were excluded from model

building. However, as a weaker, post-hoc test of whether income source and education

significantly improve predictive capacity of the model, the final models were rerun on a subset of

complete cases with these variables included. In the psychotic disorders subgroup, neither

variable significantly improves the model. In the depressive disorders subgroup, income

marginally improves the model, with a *b* estimate of 5.7 (se(*b*)=2.5, p=0.021, n=168) and partial

R^2 of 0.027. The finding that being on public assistance prolongs LOS an average of 6 days for

those with depressive disorders (after controlling for severity, age and ECT) needs to be further

examined with larger samples, using more complex measures of financial status.

Interactions

While the developed models assume that effects of independent variables are additive, interactional effects can be present. For example, the impact on LOS of an increase in severity may be greater in older than younger patients, and a higher maximum severity might prolong LOS more for patients who also have medical disorders. A final set of analyses evaluated the significance of interaction effects by assessing the contribution of each two way interaction term separately while controlling for main effects. Using this hierarchical approach, none of the interaction terms were significant at a 0.05 level in either subgroup.

Goodness of Fit of Model

The low R^2 values achieved in all the models indicate that a considerable amount of variation in LOS remains unpredicted. The standard error of the regression[7] (in SAS expressed in standardized form as the coefficient of variation - CV) is another measure of fit (Achen, 1982). The CV values for study models range from 0.60-0.70, indicating a fair amount of variability between actual and predicted LOS. This issue is further addressed in Chapter 4.

---

[7]   A measure of how far the average dependent variable departs from its forecasted value.

Table 10
Final regression models* to predict patient length of stay

Psychotic disorders subgroup (n=167)

Independent variables	Incremental R²	Cumulative R²	Coefficients for Final Model			
			b	SE(b)	T: b=0	Prob > ITI
intercept	------	-------	11.14	2.78		
maximum severity	0.112	0.112	0.24	0.06	3.81	0.0002
medical comorbidity	0.024	0.136	3.71	1.75	2.12	0.0359

R² = 0.136   Adjusted R² = 0.125   C(P) = 0.85

Depressive disorders subgroup (n=188) - no ECT

Independent variables	Incremental R²	Cumulative R²	Coefficients for Final Model			
			b	SE(b)	T:b=0	Prob > ITI
intercept	------		10.52	3.21		
severity increase	0.0920	0.09	0.48	0.12	3.92	0.0001
age	0.0610	0. 153	0.24	0.07	3.66	0.0003

R² = 0.156   Adjusted R² = 0.146   C(P) = 0.94

Depressive disorders subgroup (n=188) - with ECT

Independent variables	Incremental R²	Cumulative R²	Coefficients for Final Model			
			b	SE(b)	T: b=0	Prob > ITI
intercept	------		11.90	3.19	3.72	0.0003
severity increase	0.092	0.092	0.45	0.12	3.68	0.0003
age	0.061	0.155	0.19	0.07	2.79	0.0058
ECT	0.036	0.189	9.44	3.32	2.85	0.0049

R² = 0.190   Adjusted R² = 0.176   C(P) = 2.14

*using forward stepwise methods.

CHAPTER 4

DISCUSSION

This chapter discusses the three questions posed at the beginning of this report and addresses a number of other issues arising from the research.

Feasibility

Regarding transferability to a Canadian setting, the CSI rating tool was applied as expected by study raters. Levels of inter-rater agreement were high when scores generated by study raters were compared with those produced by the CSI trainer for the same set of charts. Rater drift was not assessed. The only major adjustment required to make the tool compatible across countries was a conversion of measurement units for reporting results of diagnostic tests from Imperial (U.S.) to Standard International units (Canadian).

Most of the information required by the tool is available in medical charts. One notable exception is the social skills indicator. It is based, in part, on the Global Assessment Scale which constitutes Axis V of the DSM clinical diagnostic system in psychiatry but is not consistently completed. If data for making a rating are missing from a chart, the CSI system assumes that the indicator is normal which may compromise rating comparability across settings. Further, it can introduce a bias in the rating calculation because social functioning data are less consistently recorded in a patient's chart than symptom and behavioural data. CSI vendors recognize that information gaps may exist in charts and promote the CSI as a vehicle for increasing awareness

of charting practices and effecting improvements.

Time to complete ratings varied according to case complexity (e.g., number of diagnoses, length of stay) and also by rater. Although raters reported an improvement in speed as their familiarity with the tool increased, rating times still ranged from 30 minutes to over two hours. The CSI trainer and study raters identified a number of strategies that could be employed to reduce the rating time:

- concurrent completion of the CIHI discharge abstract and the CSI severity rating;

- rater working conditions that are more amenable to the task - e.g., dedicated time and a quiet space.

- increase rating efficiency - for example, by concentrating on "big event" days and not entering medical diagnoses that do not impact on the patient's admission;

Without knowing more about the structure of the tool, it is difficult to know whether the strategies proposed to increase efficiency would compromise the validity of the ratings. However, the lengthy rating times might discourage some facilities from using the CSI, especially when other chart review tools such as the SPI have substantially fewer items and are reporting average review times of 15-30 minutes (Lyons, Colletta, Devens et al, 1995).

Performance

The primary goal of this study is to assess capacity of patient severity to predict LOS relative to a set of patient variables available in discharge abstracts. While the study did not include an evaluation of the psychometric properties of the CSI severity score, it is possible to comment on

its structure and to make comparisons with other similar measures.

Structure and Properties of the CSI

The number of indicators used to rate patient severity depends on the number and nature of diagnoses in a patient's chart. However there is a core set of about 35 indicators that need to be answered in order to rate most psychiatric diagnoses. Most of the indicators focus on symptoms and behaviours, history of illness and response to treatment, although three items address other areas - social skills, daily functioning and motivation. (As noted earlier, social skills information is inconsistently reported.) Any medical diagnoses in the patient's chart generate numerous additional physical symptom questions.

Published articles and material distributed by CSI provide little information about the psychometric properties of the severity scores. This is an inherent limitation of a proprietary product distributed by a private company which is primarily concerned with maintaining product control, creating a positive product profile and expanding sales. While the CSI is promoted as having high inter-rater reliability and strong face validity (due to participation of expert physician panels in defining indicators and scoring methods), information on other scale characteristics (e.g., internal consistency, component subscales, relationship to other instruments) is lacking. Severity tools, by definition, are multidimensional and it is important to understand how a scale is constructed and what domains it measures. Confidence in the effectiveness of a tool is enhanced if expected relationships to other measures (e.g., established assessment instruments and outcome scales) and across patient subgroups (defined by age, gender, chronicity, etc) are

demonstrated. A systematic and comprehensive evaluation of the psychometric properties of the CSI is sorely needed. For example, to better understand the meaning of maximum and difference severity scores, an assessment of their relationship to in-hospital clinical and treatment events is necessary.

Prediction of Length of Stay

Several different analyses were conducted to assess performance of the CSI in predicting LOS. The hierarchical regression analyses reported in Tables 9A and 9B evaluated whether each severity rating predicted variation in LOS above what could be accounted for by patient demographic and diagnosis-based variables (available in discharge abstracts). Stepwise and all subsets methods were used to produce a reduced model that maximized prediction and accuracy (see Table 10). These analyses showed that, within each diagnostic grouping, the severity ratings perform differently. It is important to explore possible clinical explanations for this result and to confirm it in other studies. While this finding is consistent with the structure of the psychiatric DRGs and CMGs which are first defined by diagnostic group and then subdivided using different variable combinations within each group, the case mix systems developed in several more recent initiatives are less dependent on diagnoses (Ellis, Wackwitz & Foster, 1991; Herman & Mowbray, 1991).

Among those with depressive disorders, admission severity is not significantly associated with LOS after controlling for abstract variables, and maximum severity only accounts for an additional 3% of variation in LOS. Difference severity, an indicator of patient escalation in

severity while in hospital, has the strongest association, accounting for 10% more variation in patient days in hospital. The final model includes difference severity and age, and accounts for 16% of variation in LOS (19% when ECT is included). In the psychotic disorders subgroup all the severity ratings significantly improve prediction of LOS after the effects of other variables are held constant, but maximum severity performs best, accounting for an additional 8% variation in LOS. The final reduced model for this subgroup includes maximum severity and medical comorbidity, and predicts 14% of variance in LOS.

Only one other study is available that assesses predictive performance of the CSI tool in a psychiatric inpatient population. Stoskopf & Horn (1991) evaluated performance of an earlier version of the CSI in a sample of patients with diagnoses of schizophrenia and depressive disorders (including bipolar). In their study, maximum severity accounted for 14% of variance in the affective disorders group and 10% in the schizophrenia group (when no other variables were in the model). These percentages dropped when other variables were added. While a direct comparison between Stoskopf's study and ours cannot be made (due to differences in sampling frame, sample composition and model variables), our findings are somewhat similar and suggest that R^2 values of 9-11% are reasonable estimates of the predictive power of the CSI severity scores when combined with other patient variables. The predictive validity of the CSI in a large sample of medical and surgical cases also is consistent with the findings of this study. As would be expected in a sample of non-psychiatric discharges, the tool performed better, accounting for approximately 20% of variation in case costs (Thomas & Ashcraft, 1991).

The findings of other studies that have used retrospectively gathered chart data to predict LOS provide another context for viewing the performance of the CSI. In a study of 200 acute care discharges with a mix of psychiatric diagnoses, Chang, Brenner & Bryant (1991) were able to account for 20% of variance in LOS. Social-environmental indicators (i.e., being employed and living with a significant other) were significant contributers to the model. Lyons, Colletta, Devens & Finkel (1995) assessed performance of the Severity of Psychiatric Illness index in a sample of 244 older adults with mixed diagnoses, including 68% with a major affective disorder or schizophrenia. The tool predicted 9% of LOS variation in the full sample and 23% in a reduced, more homogeneous sample. In addition to level of psychiatric symptoms, substance abuse and medical complications, the extent of premorbid dysfunction during the previous year was a significant predictor. Bezold and colleagues (1996) tried to predict LOS in a sample of 400 psychiatric discharges with diverse diagnoses using a diverse data set that included treatment variables, Axis I, IV and V ratings and payer type. In bivariate analyses, payer type, Axis V rating, use of ECT and prior hospitalization were significant predictors. Overall the model accounted for 12% of variance in LOS; within age subgroups R^2 increased to up to 34%. Ashcraft, Fries & Nerenz (1989) analyzed a large sample of discharges from VA acute care settings, including over 4000 discharges with major depression and over 17,000 with schizophrenia disorders. In these two groups, predicted variation in LOS was 9% and 13% for depressive disorders and schizophrenia respectively, with a measure of symptoms and function, and assistance with ADL being major predictors. The levels of variance accounted for by the final models in the present study (19% and 14% for depressive disorders and psychotic disorders respectively) are similar but not superior to findings of these other investigations.

Role of Socio-Environmental Factors

In all of these investigations measures of patient functioning were significant predictors, reinforcing earlier comments that the predictive strength of the CSI might improve if measures of non-clinical aspects of patient status were incorporated. However the availability of such data in medical charts is inconsistent. Bezold, MacDowell & Kunkel (1996) lost 50% of their initial sample due to missing multiaxial diagnostic data, and Ashcraft, Fries & Nerenz (1989) relied on a provider completed questionnaire as well as chart review to collect their data set. In this study, medical charts did not provide reliable data on patient social skills, and even basic information on income source and educational level was not consistently available. One purpose of instituting a tool such as the CSI is to improve charting practices. However clinician training is needed to achieve this benefit.

A number of recently developed severity tools have included items on patient functioning that are psychometrically sound (Barker, Barron, McFarland et al, 1994; Wing, Curtis, Beevor, 1996; Ellis, Wackerwitz & Foster, 1991; Lyons, Colletta, Devens et al, 1995). For example, in addition to measuring symptoms, behaviours and physical problems, the SPI assesses dysfunction in the preceding 12 months, availability of family support, extent of family dysfunction, client problems with living situation and employment, and motivation. These questions constitute six out of 12 items in the index. The SPI can be completed retrospectively by chart review or prospectively by clinicians familiar with the patient being rated. The HoNOS, also a 12 item tool, assesses symptoms, behaviours, impairment and social problems, with the last two domains measured by six items and constituting 50% of the final score. It is completed prospectively by providers

trained in its delivery. Developers of the HoNOS have prepared a comprehensive training program that includes periodic rater reviews to avoid rater drift and maintain reliability. Because inpatient clinicians may have limited awareness of patient functioning outside the hospital, both the HoNOS and SPI have flexible scoring algorithms to accommodate sites where social functioning data are not available.

Prospective Data Collection

The use of a prospective approach creates an opportunity to support other functions such as care planning and outcomes monitoring. This increases cost-effectiveness of data collection and is likely to improve compliance if providers begin to rely on collected data for clinical decision making. However, it also increases the burden of data collection for providers and may be resisted in programs which lack a tradition of routine data collection.

The Joint Policy and Planning Committee (of the Ministry of Health and Ontario Hospital Association) is currently involved in development of an instrument designed to meet multiple purposes. The Resident Assessment Instrument - Mental Health (RAI- MH) is one of a family of tools that are intended to be prospectively employed across diverse settings in order to enhance case mix measurement, care planning, quality improvement monitoring and outcome measurement. The planned evaluation methodology for the RAI-MH is comprehensive, and will include extensive testing of reliability and four aspects of validity (content, criterion, convergent and predictive) (Hirdes, 1997). Among other challenges faced by this ambitious project will be acceptance by providers as completion of the tool will be very time consuming.

Provider and Site Influences

Even assuming a potential for some improvement in the CSI, performance levels demonstrated in this and other chart review studies remains disconcertingly low and, at first glance, seem inadequate for forming accurate case mix groups and developing hospital funding strategies that are fair and promote desired practice. However, the methodology used in all of these studies has a fundamental limitation that raises questions about the appropriateness of using R^2 as a performance indicator. In every study current, not ideal, practice is modeled. While models have been unable to account for 70-80% of patient variation in LOS, much of this variation may be caused by provider and site divergences from "best" practices rather than from unmeasured patient factors. Lyons, O'Mahoney & Larson (1991) found that attending psychiatrist was a significant predictor of LOS in a sample of 1366 psychiatric discharges from a private hospital, even after controlling for case mix differences. They cited physician practice styles and familiarity with the patient, and unit thresholds for discharge as likely underlying causes. Fortney, Booth & Smith (1996) studied variation in LOS in VA hospital units treating patients with depression. Even though the hospitals operated under a common administrative model and reimbursement system, and the treatment of depression is fairly well prescribed, almost one third of centres had average lengths of stay outside the expected range. Fortney attributes these differences to variation in hospital and physician practice styles.

One of the reasons for introducing case mix systems is to increase conformity in physician practice. As such one would expect performance of a case mix system to improve over time as practice patterns become more consistent across providers and sites. It is difficult to know

whether the low R^2 achieved in this and other studies results from a lack of understanding of and inability to measure factors that influence LOS, or from the need for providers to move toward more evidence-based, systematic delivery of care. Further progress is likely required on both fronts - case mix systems still need improvement and practitioners need to change. It also is likely that psychiatry will never achieve the precision of predicting hospital course and resource use that has been attained in other fields of medical care. As Fries and colleagues (1993) note "Treatment patterns (in psychiatry) are less well defined (than in other sectors of acute care), with multiple clinically accepted care models for similar diagnoses." (p.33)

Role of Outcomes in Case Mix Development

Even if case mix groups that are more homogeneous in hospital LOS can be developed, the appropriateness of the hospitalization period associated with each group needs to be established. As Bezold, MacDowell & Kunkel (1996) comment "LOS alone is not a final indicator of effective and efficient care." (p.422). At a time when hospitals are under considerable pressure to reduce LOS and when the shortest average LOS achieved by a cohort of hospitals often is adopted as a performance benchmark, it is essential that outcomes are incorporated into validity testing of new case mix systems. The clinical credibility of each case mix category will rest on its association with a best practice guideline and with a demonstrated capacity to achieve an appropriate level of therapeutic benefit. These associations are necessary to achieve a key benefit of implementing case mix funding, that of improving care practices.

Role of Treatment Variables in Case Mix Development

In this study the variation in LOS predicted by use of ECT drops from 8% to 2% when severity and other patient variables are included in the model. Yet the estimated *b* coefficient remains high, indicating an average increase of 9.4 days in LOS when ECT is used. It is appealing to use ECT as a defining variable in case mix groups as it is readily available and improves prediction accuracy of case mix systems. Yet inclusion of this variable also increases vulnerability to gaming. Moreover the widely varying rates of use of ECT across sites in this study raise concerns about how consistently ECT is applied within patient subgroups. It may be more valuable to look at the role of ECT in defining best practices within case mix groups than to use ECT as a grouping variable.

Current Initiatives in Case Mix and Level of Care

There is currently little dispute that measures of patient severity are critical for defining case mix systems. In progressive jurisdictions throughout Canada, the United States and the United Kingdom, efforts are underway to develop more powerful mental health information systems for system planning, management and funding. Many of these information systems include, as a core element, a severity measure that is, or will be, incorporated into criteria for defining case mix groupings and associated levels of care.

The Colorado Client Assessment Record (CCAR) is a problem checklist and level of functioning rating instrument that is applied to all admissions to the public mental health system. An eight level client typology has been developed from the CCAR which is based on 13 factors and three

broad dimensions - self care, acting out and emotions. The typology is independent of diagnosis. In validity testing, the typology differentiated client groups with respect to location, intensity and cost of services used. In an inpatient study, the CCAR typology provided evidence of case mix differences between two hospitals that were not apparent using the DRG classification system. However, LOS for patients in similar CCAR subgroups varied between the two sites. The Colorado Department of Mental Health is currently using CCAR and other user data for performance contracting with mental health services. Work continues to develop a service typology for each CCAR profile that can be used for client centred monitoring and funding (Ellis, Wackwitz & Foster, 1991).

The CCAR project has influenced development of severity tools and mental health information systems in numerous other locales. Mental health authorities in six other states are using the CCAR in its current or a revised form. CCAR served as a starting point for development of the HoNOS, a measure of health and social functioning that is part of a revised minimum data set (MDS) for mental health being evaluated in the United Kingdom. A team of UK researchers is conducting a pilot project to match resources to care for people with severe mental illness, using a case mix typology based on the HoNOS and other MDS data (Huxley, 1997). In Ontario, an initiative is underway in Windsor/Essex County to conduct CCAR assessments of current users of county case management programs as well as individuals on program waiting lists. The assessments will be used to adjust levels of case management intensity to identified need and to monitor changes in client functioning over time (Carruthers, 1997). Northwestern Ontario has implemented a data linkage system which includes a client severity rating based on the CCAR

problem profile (Lakehead Psychiatric Hospital Research Department, 1996).

Initiatives similar to the CCAR project are in progress elsewhere in North America but success in mapping case mix groups into associated levels of service use remains illusive. The Texas Department of Mental Health and Mental Retardation has implemented a MDS that includes measures of symptoms, functioning and supports, and outcomes. The Multnomah Community Ability Scale is the component of the MDS that measures client functional severity (Barker, Barron, McFarland & Bigelow, 1994). Work is underway to use this core set of data to define case mix groupings with associated levels of care but progress has been slow (Texas Department of Mental Health, 1996). Tennessee, New Hampshire and several other states have implemented a grouping system for mental health service users called Clinically Related Groups (CRGs). The main criteria for classifying consumers are diagnosis and level of functional impairment which is based on ratings of performance in activities of daily living; interpersonal functioning; concentration, task performance and pace; and adaptation to change. However, in a study that assessed treatment costs associated with the CRGs, the four groups collapsed into only two distinct cost categories (Tennessee Department of Mental Health, 1995). In Ontario the Ministry of Health is considering how severity and outcome measures might be incorporated into a minimum data set that all mental health services (inpatient and community) will be required to maintain.

As these examples illustrate, the task of defining case mix groups that predict inpatient hospital resource use or LOS has expanded into the larger task of linking case mix groups with a resource

use package that encompasses care received across all service settings (e.g., inpatient, outpatient and community) over a specified period of time. This approach is consistent with the goals of mental health reform (and health care reform in general) to create a seamless system of care, improve continuity of care, and increase use of less restrictive care alternatives.

It is now widely accepted that a severity measure based on multiple areas of a person's social and personal functioning should be included in any case mix typology intended for planning, funding and monitoring systems of mental health care. Many jurisdictions have incorporated severity measures into their mental health MDSs. Several have developed case mix groups based on this data that identify users with distinct clinical profiles. However the task of linking case mix groups with distinct bundles of services has proved to be much more difficult.

Study Limitations

Generalizability/External validity

The study sample was restricted to two diagnostic groups and to CIHI typical cases so findings can only apply to discharges within these categories. Only three hospitals participated in this study, each one from a different sector (ie., teaching facility, tertiary care, provincial hospital). The sample sites may not represent experiences in these or other hospital sectors (ie., community hospitals).

Power/Accuracy

The small study sample reduces the accuracy of the coefficient estimates and the power of the

study to detect significant associations between predictor variables and LOS. For example, the marginal association between income and LOS (p=.021) might have reached a higher level of significance if tested with a larger sample. (In this case, missing data further compromised sample size).

Shrinkage/Over-fitting

Some shrinkage is expected when study regression models are applied to new samples. In an attempt to minimize shrinkage, the difference between R^2 and adjusted R^2 was one criteria used for final model selection. The common strategy of splitting a study sample into derivation and validation subsamples during model development to assess shrinkage was not possible due to the small sample size.

Interpretation of CSI

Information on the structural properties of the CSI was not available from the CSI developer, thus limiting ability to interpret findings. For example, what is the meaning of an increase in patient severity after admission? Why did medical comorbidity continue to be a significant predictor after maximum severity was entered into the model?

Influence of Site/Provider Variation

Due to the study design and small sample, it was not possible to estimate the contribution of site or provider in predicting LOS. Therefore, the extent to which unpredicted variation in LOS in the study is caused by provider influences or unmeasured patient variables cannot be estimated.

Conclusion

Summary of CSI Evaluation

This study has demonstrated the feasibility of applying an inpatient severity rating system developed in the United States to a sample of psychiatric discharges from Ontario hospitals. The tool was applied reliably by raters across three hospital sites, largely because it is based on objective, clearly defined indicators of patient status. While most of the data required to make the severity ratings were available in medical charts, data to rate social skills were inconsistently reported. Implementation of the tool could be a stimulus for training clinicians to chart more consistently and reliably. Because the tool is completed by medical records personnel, the increased demand placed on clinicians for data collection is minimal. Because the rating is based on multiple, clearly defined indicators, vulnerability to gaming also is minimized.

The CSI could be a useful support to utilization review activities as it can identify cases that merit further examination - for example, individuals who are admitted with low severity scores, individuals discharged with high severities and those who increase in severity while in hospital. The severity ratings substantially improve capacity to predict LOS over a subset of patient items available in hospital discharge abstracts, accounting for an additional 8-10% variation in LOS. Predictive models developed in this study differ across diagnostic groups, indicating that diagnosis in some form should be incorporated into case mix systems. Difference severity is the strongest predictor of LOS in the depressive disorders subgroup, suggesting that measures of patient change during hospitalization should be included in a case mix system.

Despite its strengths the CSI has significant structural and administrative weaknesses. In the indicators used to make the ratings, there is a strong emphasis on clinical status, with little attention paid to social functioning and environmental supports. To give more emphasis to these domains would require more clinician training and development of a better set of indicators.

The CSI rates inpatients and relies on charted ICD-9 codes to generate the indicators necessary to make the ratings. An ambulatory version of the CSI exists which is consistent with the move to developing tools that can be used across treatment settings. However the time required to rate inpatients is excessive (up to two hours) and could become a serious limitation if the tool is applied prospectively or outside of a hospital setting. Less reliance on ICD-9 codes also is needed as community programs do not consistently collect diagnostic data.

The performance of the CSI in predicting inpatient LOS does not appear to be superior to other available chart review data sets although direct comparisons are not possible. Performance may have been weakened by the lack of information about social factors.

Users need more information on the properties of the score to better understand the clinical meaning of the ratings. It is possible that, after a site licence contract is signed with the vendor, more information about the tool is forthcoming.

Severity Measurement and Case Mix Development - Future Directions

Many jurisdictions are now viewing funding of mental health care from a broader perspective,

with an episode of care being defined as use of an array of services over time to resolve and manage a problem. As such, the tools that support funding methodologies need to be applicable across treatment settings. There also is a move to prospective data collection which may increase compliance and can support multiple activities including clinical decision-making, outcomes monitoring and case mix classification. Many jurisdictions are in the process of defining a minimum mental health data set that all mental health services (i.e., inpatient, outpatient, community) in a defined area are expected to maintain.

It is widely accepted that a multidimensional measure of patient severity based on symptoms, behaviours, social and personal functioning needs to be incorporated into a data set used to define case mix categories. Research evidence indicates that measures of social functioning and environmental supports improve case mix definitions although considerable variation in patient resource use still exists. Estimates of the contribution of provider and facility practice patterns to LOS variation are needed to understand how much variation remains unpredicted. It is likely that further improvements in capacity of severity measures to predict resource use can be obtained but mental health will probably never achieve the precision of other health areas in prescribing care practices and predicting levels of patient resource use. For this reason, funding models in mental health may need to be more flexible than in other health areas. For example, strategies for dealing with cases that substantially exceed funded LOS may need to provide more opportunities for explanation and justification. While patient centred funding models have many benefits, it may be that program based funding needs a continued role in mental health. Performance contracting that rewards or penalizes programs based on aggregate levels of performance may be a more

realistic funding approach in mental health. Accurate case mix classification systems would still

be needed so that performance benchmarks can be established for different subgroups of users.

Whatever solutions are adopted , the mental health field needs to accelerate its efforts in this and

other areas to meet increased demands for accountability. In the province of Ontario all health

sectors are under considerable pressure to implement information systems, tools and

methodologies that monitor program delivery, costs, outcomes and quality. As evidenced by the

work of the Health Services Restructuring Commission in Ontario, this data is needed to inform

the major reorganization of facilities, services and governing structures currently underway in the

province. The Joint Policy and Planning Committee and the Canadian Institute for Health

Information are making progress in the acute care sector in developing patient categories and

funding methodologies that are responsive to both patient and facility characteristics. The long

term care sector is using a resident assessment tool that supports case mix categorization, care

planning and quality control. While there is considerable pressure on mental health to implement

similar monitoring systems, progress has been slow. For example, as the present study indicates,

in the area of funding there has been limited success in developing tools that can support more

equitable funding and reward better practices. However, efforts to achieve change are

accelerating. Numerous initiatives are currently underway in Ontario to improve data quality in

mental health, in both institutional and community environments (Ontario Ministry of Health,

1997; Hirdes, 1997; Mental Health Policy Research Group, 1997; Lakehead Psychiatric Hospital

Research Department, 1996). Perhaps even more importantly, there is a growing acceptance that

the health care environment is evolving; that mental health services must be positioned to

demonstrate their efficiency and effectiveness; and that without adequate information the field

will be disadvantaged in its ability to compete for health care dollars.

References

Ashcraft, M. L. F., Fries, B. E., Nerenz, D. R., & et al. (1989). A psychiatric patient classification system: An alternative to diagnosis-related groups. Medical Care, 27(5), 543-557.

Averill, R. F., McGuire, T. E., Manning, B. E., & et al. (1992). A study of the relationship between severity of illness and hospital cost in New Jersey hospitals. Health Services Research, 27(5), 587-604.

Barker, S., Barron, N., McFarland, B. H., et al., & Bigelow, D. A. (1994). A community ability scale for chronically mentally ill consumers: Part 1. Reliability and validity. Community Mental Health Journal, 30(4), 363-379.

Bezold, H. S., MacDowell, M., & Kunkel, R. (1996). Predicting psychiatric length of stay. Administration and Policy in Mental Health, 23(5), 407-423.

Botz, C. K. (1991). Principles for funding on a case mix basis: Construction of case weights (RIWs). Healthcare Management Forum, , 22-32.

Botz, C. K., Bestard, S., Demaray, M., & et al. (1993). Resource utilization groups (RUGs): Defining chronic care, rehabilitation and nursing home case mix in Canada. Healthcare Management Forum, 6(4), 5-11.

Boyer, C. A., Olfson, M., Kellerman, S. L., Hansell, S., Walkup, J., Rosenfield, S., & Mechanic, D. (1995). Studying inpatient treatment practices in schizophrenia: an integrated methodology. Psychiatric Quarterly, 66(4), 293-320.

Brooke, P. P., Hudak, J. D., & Finstuen, K. (1994). Product-line evaluation of graduate medical education program costs. Hospital & Health Services Administration, 39(2), 265-277.

Cagle, L. T., & Banks, S. M. (1986). The validity of assessing mental health needs with social indicators. Evaluation and Program Planning, 9, 127-142.

Canadian Institute for Health Information. (1994). Length of stay database by CMG.

Canadian Institute for Health Information. (1995). Resource intensity weights. Summary of methodology 1995/96.

Canadian Institute for Health Information. (1997). Mental Health and Addictions Services: Existing Standards and Related Initiatives for Management Information (draft)

Carruthers, V. (personal communication, July 11, 1997)

Caton, C. L. M., & Gralnick, A. (1987). A review of issues surrounding length of psychiatric hospitalization. Hospital and Community Psychiatry, 38(8), 858-863.

Chang, G., Brenner, L., & Bryant, K. (1991). Factors predicting inpatient length of stay in a

CMHC. Hospital and Community Psychiatry, 42(8), 853-855.

Charles, C., & Schalm, C. (1992). Alberta's Resident Classification System for Long-Term Care

Facilities. Part I: Conceptual and Methodological Development. Canadian Journal on Aging,

11(3), 219-232.

Choca, J. P., Peterson, C. A., Shanley, L. A., Richards, H., & Mangoubi, E. (1988). Problems in

using statistical models to predict psychiatric length of stay: An illustration. Hospital and

Community Psychiatry, 39(2), 195-197.

Chu, C. (1994). Resource intensity weighing and case mix grouping: Assumptions and

implications for health service performance evaluation. Health Management Forum, 7(1), 24-31.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological

Measurement, 20, 37-46.

Creed, F., Tomenson, B., Anthony, P. & Tramner, M. (1997). Predicting length of stay in

psychiatry. Psychological Medicine, 27, 961-966.

Cyr, J. J., & Haley, G. A. (1983). Use of demographic and clinical characteristics in predicting

length of psychiatric hospital stay: A final evaluation. Journal of Consulting and Clinical

Psychology, 51(4), 637-640.

Dada, M., White, W. D., Stokes, H. H., & Kurzeja, P. (1992). Prospective payment for

psychiatric services. Journal of Health Politics, Policy and Law, 17(3), 483-508.

Ellis, R. H., Wackwitz, J. H., & Foster, M. (1991). Uses of an empirically derived client typology

based on level of functioning: Twelve years of the CCAR. The Journal of Mental Health

Administration, 18(2), 88-101.

English, J., Sharfstein, S., Scherl, D.,et al. (1986). Diagnosis-related groups and general hospital

psychiatry: The APA study. The American Journal of Psychiatry, 143(2), 131-139.

Fetter, R.B., Shin, Y., Freeman, J.L., Averill, R.F. and Thompson, J.D. (1980). Casemix

definition by diagnosis-related groups. Medical Care 18 (Supp.No.2), 1-53

Fisher, W. H., & Altaffer, F. (1992). Inpatient length of stay measures: Statistical and conceptual

issues. Administration and Policy in Mental Health, 19(5), 311-320.

Fortney, J. C., Booth, B. M., & Smith, G. R. (1996). Variation among VA hospitals in length of

stay for treatment of depression. Psychiatric Services, 47(6), 608-613.

Fox, J. (1991). Regression Diagnostics. Sage Publications.

Frank, R. G., & Lave, J. R. (1985). A plan for prospective payment for inpatient psychiatric care. Hospital and Community Psychiatry, 36(7), 775-776.

Freiman, M. P., Mitchell, J. B., & Rosenbach, M. L. (1987). An analysis of DRG-based reimbursement for psychiatric admissions to general hospitals. American Journal of Psychiatry., 144(5), 603-609.

Freund, R. J., Littell, R. C. (1991). SAS System for Regression, Second Edition, Cary, NC: SAS Institute, Inc.

Fries, B. E., Durance, P. W., Nerenz, D. R., & Ashcraft, M. L. F. (1993). A comprehensive payment model for short- and long-stay psychiatric patients. Health Care Financing Review, 15(2), 31-50.

Giovannetti, P. (1979). Understanding Patient Classification Systems. Journal of Nursing Administration, 4-9

Goeree, R. (1994). Evaluation of programs for the treatment of Schizophrenia: A health economic perspective Evaluation of programs for the treatment of schizophrenia: A health economic perspective. Canada: McMaster University, Centre for the Evaluation of Medicines, St. Joseph's Hospital.

Goering, P., Wasylenki, D., Lancee, W., & Freeman, S. J. J. (1984). From hospital to community. Six-month and two-year outcomes for 505 patients. The Journal of Nervous and Mental Diseases, 172(11), 667-673.

Goldman, H. H., & Sharfstein, S. S. (1987). Are specialized psychiatric services worth the higher cost? (EDITORIAL). American Journal of Psychiatry, 144(5), 626-628.

Goldman, H. H., Taube, C. A., & Jencks, S. F. (1987). Chapter 2: The organization of the psychiatric inpatient services system. Medical Care (Supplement), 25(9), S6-S21.

Gordon, R. E., Vijay, J., Sloate, S. G., Burket, R., & Gordon, K. K. (1985). Aggravating stress and functional level as predictors of length of psychiatric hospitalization. Hospital and Community Psychiatry, 36(7), 773-774.

Gurny, P. (personal communication, March, 1994)

Halpine, S., & Ashworth, M. A. (1994). Measuring Case Mix and Severity of Illness in Canada: Case Mix Groups vs Refined Diagnosis Related Groups. Healthcare Management Forum, 6(4), 20-26.

Herman, S. & Mowbray, C. (1991). Client Typology Based on Functioning Level Assessments: Utility for Service Planning and Monitoring, Journal of Mental Health Administration, 18(2),

101-115.

Hirdes, J. P. (1997)   Integrated health information systems: Making mental health fit.

Holley, H. L., Kulczycki, G., & Arboleda-Florez, J. (1994). Case-mix funding and legislated

psychiatric care. International Journal of Law and Psychiatry. 17(4), 377-393.

Hopkins, D. S. P., & Carroll, R. J. (1994). Severity adjustment models for CPI. In S. Horne & D.

S. P. Hopkins (Eds.), Clinical Practice Improvement: A New Technology for Developing Cost-

Effective Quality Health Care. (pp. 91-102). New York: Faulkner & Gray Inc.

Horgan, C., & Jencks, S. F. (1987). Chapter 3: Research on psychiatric classification and

payment systems. Medical Care (Supplement), 25(9), S22-S36.

Horn, S. D., Chambers, A. F., Sharkey, P. D., & et al. (1989). Psychiatric severity of illness: A

case mix study. Medical Care, 27(1), 69-84.

Horn, S. D., Sharkey, P. D., Buckle, J. M., Backofen, J. E., Averill, R. F., & Horn, R. A. (1991).

The relationship between severity of illness and hospital length of stay and mortality. Medical

Care, 29(4), 305-317.

Huxley, P., Challis, D., Hughes, J. & Schneider, J. (no date). The usefulness of Healthcare Resource Groups and the Minimum Data Set for mental health in supporting the joint commissioning of comprehensive mental illness services. University of Manchester.

Huxley, P. (personal communication, March 24, 1997)

Jarman, B., Hirsch, S., White, P., & Driscoll, R. (1992). Predicting psychiatric admission rates. British Medical Journal, 304, 1146-1151.

Jencks, S. F., & Goldman, H. H. (1987). Chapter 5: Implications of research for psychiatric prospective payment. Medical Care (Supplement), 25(9), S42-S51.

Jencks, S. F., Horgan, C., Goldman, H. H., & Taube, C. A. (1987). Chapter 1: The problem. Medical Care (Supplement), 25(9), S1-S5.

Jencks, S. F., Horgan, C., & Taube, C. A. (1987). Chapter 4: Evidence on provider response to prospective payment. Medical Care (Supplement), 25(9), S37-S41.

Joint Policy and Planning Committee. (1995). Final Recommendations of the Activity Measurement Working Group Regarding the Evaluation of CIHI's Psychiatric CMGs, Toronto.

Kelly, A., & Jones, W. (1995). Small area variation in the utilization of mental health services: Implications for health planning and allocation of resources. Canadian Journal of Psychiatry, 40(9), 527-532.

Kiesler, C. A., Simpkins, C., & Morton, T. (1990). Predicting length of hospital stay for psychiatric inpatients. Hospital and Community Psychiatry, 41(2), 149-154.

Kiesler, C. A., & Simpkins, C. (1991). The De Facto national system of psychiatric inpatient care. Piecing together the national puzzle. American Psychologist, 46(6), 579-584.

Lakehead Psychiatric Hospital Research Department. (1996). Northwestern Ontario Data Linkage System: Regional Report #5.

Lave, J. R., Jacobs, P., & Markel, F. (1992). Transitional funding: Changing Ontario's global budgeting system. Health Care Financing Review, 13(3), 77-84.

Leff, S., Swartz, B., Ghler, M., & et al. (1985). An initial exploration of functional level and length of psychiatric hospitalization in four different hospitals: Relationships and implications for a prospective payment system. Final report submitted by Human Services Research Institute to the NIMH. Contract 84MO45181701D,

Lyons, J. S., O'Mahoney, M. T., & Larson, D. B. (1991). The attending psychiatrist as a predictor of length of stay. Hospital and Community Psychiatry, 42(10), 1064-1065.

Lyons, J. S., Colletta, J., Devens, M., & Finkel, S. I. (1995). Validity of the severity of psychiatric illness rating scale in a sample of inpatients on a psychogeriatric unit. International Psychogeriatrics, 7(3), 407-416.

Lyons, J. S., O'Mahoney, M. T., Doheny, K. M., Dworkin, L. N., & Miller, S. I. (1995). The prediction of short-stay psychiatric inpatients. Administration and Policy in Mental Health, 23(1), 17-25.

Lyons, J. S., O'Mahoney, M. T., Miller, S. I., Neme, J., Kabat, J., & Miller, F. (1997). Predicting readmission to the psychiatric hospital in a managed care environment: Implications for quality indicators. American Journal of Psychiatry, 154(3), 337-340.

McCrone, P., & Phelan, M. (1994). Diagnosis and length of psychiatric in-patient stay. Psychological Medicine, 24, 1025-1030.

Mechanic, D. (1996). Technologies for the delivery of mental health care. International Journal of Technology Assessment in Health Care, 12(4), 673-687.

Mental Health Outcomes Project Workgroup. (1996). The Texas mental health outcomes system

(adults).

Mental Health Policy Resource Group. (1997). Community Mental Health Evaluation Initiative:
Request for Proposals. Toronto.

Miller, G. H., Dear, M., & Streiner, D. L. (1986). A model for predicting utilization of
psychiatric facilities. Canadian Journal of Psychiatry, 31(5), 424-430.

Mitchell, J. B., Dickey, B., Liptzin, B., & et al. (1987). Bringing psychiatric patient into the
medicare prospective payment system: Alternatives to DRGs. American Journal of Psychiatry,
144(5), 610-615.

Norman, G. R., & Streiner, D. L. (1986). PDQ Statistics. B.C. Becker Inc.

Norman, G. R., & Streiner, D. L. (1994). Biostatistics: The Bare Essentials. Mosby.

Ontario Ministry of Health. (1997). Mental Health Minimum Data Set Progress Report.

Richman, A., Boutilier, C., & Harris, P. (1984). The relevance of socio-demographic and
resource factors in the use of acute psychiatric in-patient care in the Atlantic Provinces of
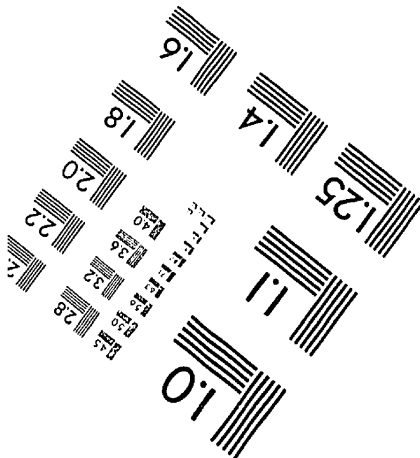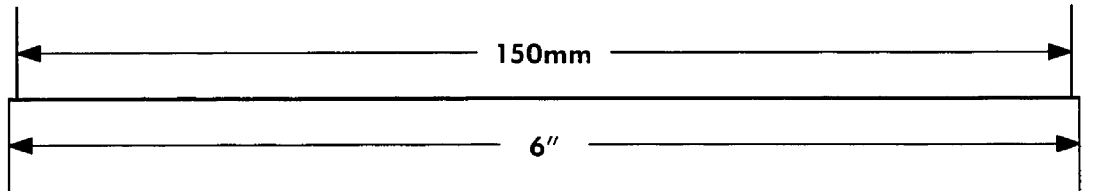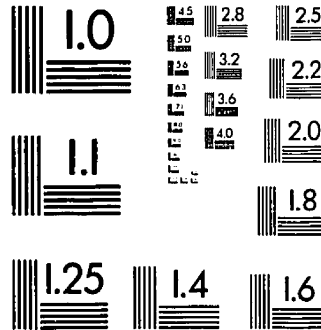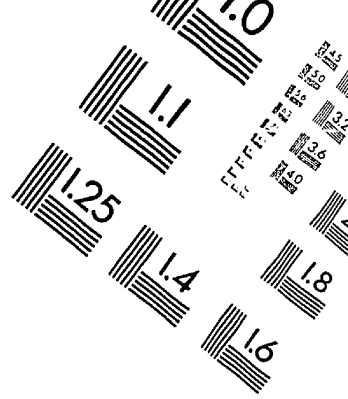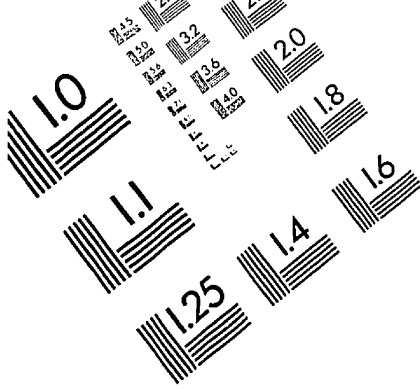Canada. Psychological Medicine, 14(1), 175-182.

Rosenheck, R., & Massari, L. (1991). Psychiatric inpatient care in the VA: Before, during and after DRG-based budgeting. American Journal of Psychiatry, 148(7), 888-891.

Sanderson, H. F., Anthony, P., & Mountney, L. M. (1995). Healthcare resource groups - version 2. Journal of Public Health Medicine, 17(3), 349-354.

Statistics Canada. (1990). Mental health statistics 1985-1986 revisited. Supplement no. 3, volume 2(1). Canada: Minister of Supply and Services (Catalogue 82-0036 (Quarterly).

Stoskopf, C., & Horn, S. (1991). The computerized psychiatric severity index as a predictor of inpatient length of stay for psychoses. Medical Care, 29, 179-195.

Stoskopf, C., & Horn, S. (1992). Predicting length of stay for patients with psychoses. Health Services Research, 26(6), 743-766.

Streiner, D. L., Goodman, J. T., & Woodward, C. A. (1975). Correlates of the hospitalization decision: A replicative study. Canadian Journal of Public Health, 66, 411-415.

Streiner, D. L., & Norman, G. R. (1989). Health Measurement Scales: A Practical Guide to their development and use. Oxford Medical Publications.

Streiner, D. L. (1994). Research methods in psychiatry. Regression in the service of the superego: The do's and don'ts of stepwise multiple regression. Canadian Journal of Psychiatry, 39(4), 191-196.

Taube, C., Sul Lee, E., & Forthofer, R. N. (1984). Diagnosis-related groups for Mental Disorders, Alcoholism and Drug Abuse: Evaluation and Alternatives. Hospital and Community Psychiatry, 35(5), 452-455.

Taube, C. A., Thompson, J. W., Burns, B. J., & et al. (1985). Prospective payment and psychiatric discharges from General Hospitals with and without psychiatric units. Hospital and Community Psychiatry, 36(7), 754-760.

Tennessee Department of Mental Health & Mental Retardation. (1995). Manual for the Tennessee clinically related group assessment.

Thomas, J. W., & Ashcraft, M. L. F. (1991). Measuring severity of illness: Six severity systems and their ability to explain cost variations. Inquiry, 28, 39-55.

Tucker, P., & Brems, C. (1993). Variables affecting length of psychiatric inpatient treatment. The Journal of Mental Health Administration, 20(1), 58-65.

Wellock, C. M. (1995). Is a diagnosis-based classification system appropriate for funding

psychiatric care in alberta? Canadian Journal of Psychiatry, 40(9), 507-513.

Wing, J. W., Curtis, R. H., & Beevor, A. S. (1996). HoNOS: Health of the nation outcome

scales. Report on research and development July 1993 - December 1995. London: Royal College

of Psychiatrists.

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

150mm

6″