Development of Quality Indicators for Inpatient Mental Healthcare: Strategy for Risk Adjustment

by

Christopher M. Perlman

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Health Studies and Gerontology

Waterloo, Ontario, Canada, 2009

Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Background and Purpose: Quality measurement is an essential, yet, complex component of mental health services that is often limited by a lack of clinically meaningful data across service providers. Understanding how services are organized, delivered, and effective is vital for ensuring and improving health care quality. In quality measurement of mental healthcare, structural indicators are common with fewer process and outcome indicators available. Using data from the RAI - Mental Health (RAI-MH), a comprehensive assessment system mandated for use in Ontario, this dissertation aims to define a set of mental health quality indicators (MHQIs), effectiveness quality indicators (EQIs), and risk adjustment strategy that can be used to evaluate and compare quality at the facility- and regional-levels.

Methodology: The MHQIs were developed using a retrospective analysis of two data sets: A pilot sample of 1,056 RAI-MH admission and discharge assessments collected from 7 inpatient mental health units in Ontario and a sample of 30,046 RAI-MH admission and discharge assessments collected from 70 Ontario hospitals as part of the Canadian Institute for Health Information Ontario Mental Health Reporting System. The MHQIs were chosen based on clinically meaningful domains identified by mental health and quality stakeholders, MHQI rates that were consistently above 5% or below 95% among hospitals, and appropriate variation in rates among hospitals in both sets of data. For each MHQI domain, regression modeling using generalized estimating equations was employed to choose risk adjustment variables and logistic or linear regression was used to perform risk adjustment to compare MHQI and EQI rates among hospitals and regions.

Results: A set of 27 MHQIs was defined measuring improvement and incidence/failure to improve in the following domains: depressive/psychosis/pain symptoms, cognitive/physical/social functioning, aggressive/ disruptive/violent behaviours, and control procedures. Also, 13 EQIs were defined to identify the magnitude of change in MHQI domains per 7 days between assessments. Regression models using generalized estimating equations identified between 1 and 8 risk adjustment covariates

for each MHQI. Risk adjustment using logistic and linear regression resulted in over 50% of hospitals and LHINs changing in rank based on MHQI and EQI scores.

Conclusion: This dissertation has developed an evidence-based set of MHQIs and EQIs based on a clinically rich set of data. Since the data is available provincially, the MHQIs and EQIs can be used for hospital based, regional, and public reports on quality of inpatient mental health services. The MHQIs/EQIs can be linked to care planning and funding using the RAI-MH to promote quality improvement and accountability for recipients, providers, managers, governors, and funders of mental health services. Opportunities are also available to extend the use of the MHQIs to community mental health, so that system level evaluations of quality can be developed.

ACKNOWLEDGEMENTS

I owe a great deal of thanks to my doctoral supervisor, Dr. John Hirdes. I immensely appreciate Dr. Hirdes' willingness to trust and challenge me through my professional development. Dr. Hirdes has provided me with so many diverse and rewarding opportunities to learn from others, share my knowledge, and build professional networks. I would also like to acknowledge the incredible influence that Dr. Hirdes has on improving the lives of others, including mine, through his tireless commitments to research, teaching, capacity building, mentorship, and so much more.

I thank the commitment of my thesis advisory committee for all of their support in the preparation and completion of this dissertation. I would like to thank Dr. Howard Barbaree for helping me to understand the implications for why evaluating and understanding mental health services is so important, Dr. Joel Dubin for his advice on the conceptual, methodological, and study design aspects of my thesis, and Dr. Ian McKillop for helping me to think critically about the importance and implications of measuring and evaluating quality. I also thank my internal/external examiner, Dr. Weizhen Dong, and my external examiner, Dr. Eric Latimer, for taking the interest and time to evaluate this dissertation and provide me with so many insightful suggestions and comments to enhance this research as it progresses beyond this dissertation.

Financial support for this dissertation was provided by the *ideas* for Health Project, the Ontario Ministry of Health and Long Term Care, Primary Health Care Transition Fund, and the Homewood Research Institute. I also acknowledge the support of the Canadian Institute for Health Information and interRAI for providing the data to carry out this research. Many thanks go to Dr Jeff Poss at the Univeristy of Waterloo and Dr. Rich Jones at the Hebrew Rehabilitation Centre, Boston, for their statistical and methodological advice around understanding the design and applications of quality indicators for health services.

I owe many thanks to many of the excellent staff at Homewood Health Centre in Guelph, Ontario for providing me with the opportunity and support to complete this dissertation and develop my career in mental health services and research. I offer very special thanks to Joe Power, Project

Manager, for promoting, supporting, and guiding my career development. I owe a great deal of gratitude to Dr. Edgardo Perez, President and Chief Executive Officer, for his willingness to take the time to share his expertise and provide me with the opportunity to learn and contribute at Homewood, and beyond.

I would like to thank my friends and family who have always supported me. I thank Dr. Janice Husted for all of her encouragement and very thoughtful advice about how to manage and enjoy my academic and personal journeys. Also, thanks to John Sax for being a great friend and helping me relax and have fun, no matter how stressful things seemed.

I owe many thanks to my parents, Mike and Mary Lee Perlman, and my brother, Steve Perlman, for providing guidance, support, and friendship through the completion of my doctorate, and life in general. My parents have always encouraged me to do the research to find the answers for what I seek out, and I think it is fitting that this advice has led me to a career in research.

Finally, I owe many thanks to Rebecca Thompson. As a mental health recreation therapist, Rebecca helped me by sharing her extensive knowledge about mental health services and the people who need them. Her endless passion and commitment to others is inspiring. As my wonderful wife, I thank Rebecca for her personal support, understanding, encouragement, and love as I navigated through this journey. I am so appreciative of her patience through this experience and so proud and fortunate to share my life with her.

TABLE OF CONTENTS

LIST of TABLES

LIST of FIGURES

1.0 INTRODUCTION

Over the past two decades there has been an increased emphasis on accountability for health services. Health care providers are expected to use evidence-based approaches for evaluating quality of care and effective resource allocation. These measurements can then be used to develop policies and decisions about future delivery and funding of services, provide information useful for national reporting or accreditation agencies, help providers improve practices and service delivery, and assist the public to choose appropriate and effective services (Hermann, Leff, Palmer et al., 2000).

Research on quality measurement is not as established for mental health care compared to other health sectors. The difficulty with and lack of research on quality in mental health services is related to the complexity of this health sector. In health sectors, such as surgical care, the primary outcomes (e.g., mortality) are often more concrete (Atherly, Fink, Campbell, et al., 2004). In mental health, a multitude of factors affect the way services are delivered and the outcomes they achieve. The factors might include the individual characteristics of patients, treatment providers, and variations in treatment effectiveness. The diversity of these factors requires special consideration for choosing what and how aspects of quality should be measured.

The purpose of this dissertation is to develop an approach for measuring quality of inpatient mental health services based on routinely collected clinical assessment data. With consideration for the unique characteristics of mental health services, a list of quality indicators (QIs) will be defined and a risk adjustment strategy will be developed for using these QIs to make fair comparisons between inpatient mental health hospitals and Local Health Integration Networks (LHINs) in Ontario. Before specifically describing these studies, an introduction to the burden of mental illness and the structure of mental health services in Ontario is first provided highlighting the complexity of mental health clients and treatment. Next, a review is provided of approaches for QI development and use including a description of risk adjustment approaches and applications for quality measurement followed by a

review of QI research and applications in mental health services, both internationally and in Canada. Finally, the specific purpose and research questions for this dissertation are presented.

1.1 The Burden of Mental Illness and Structure of Mental Health Services

Mental illness is a global issue affecting poverty stricken, developing, and industrialized countries. The World Health Organization (WHO) estimates the 12-month prevalence of mental illness to range between 12% and 26% in the Americas, 8% and 20% in Europe, 5% and 17% in the Middle East, and 4% and 9% in Asia (WHO World Mental Health Survey Consortium, 2004). Capturing a precise prevalence of mental illnesses, internationally, may be difficult for a variety of reasons including cultural differences in interpretation or perception of mental illness, stigma of reporting, and diagnostic practices. In Canada, it is estimated that at least 20% of the population will personally experience mental illness in their lifetime (Health Canada, 2002). About 12% of the Canadian population experience anxiety disorders in a given year, between 5% and 8% experience mood disorders, and about 1% experience schizophrenia (Health Canada, 2002). The prevalence of mental illness and substance use disorders ranges significantly between different regions in Canada. The prevalence of substance-use disorders, for example, is much higher in western and eastern provinces (12% to 14%) than Ontario and Quebec (8% to 9%) with more variation within larger cities such as Toronto and Montreal (Veldervizen, Urbanski, & Cairney, 2007).

Mental illnesses affect individuals of all age ranges. Hospitalizations due to mental illness in Canada accounted for 1% of all hospitalizations among those under the age of 15, 12% of 15 to 20 year olds, 10% of those age 25 to 44, 4% for those age 45 to 65, and 1% for those aged 65 (Health Canada, 2002). For all individuals, mental illness can reduce quality of life, promote poor physical health, and disrupt social and emotional functioning (Hoffman, Dukes, Wittchen, 2008). Mental illnesses, led by depression and substance abuse, account for 4 of the top 10 causes of disability, internationally (WHO, 2001). In Canada, suicide accounts for 24% of all deaths of persons age 15 to 24 and 16% of deaths for persons age 25 to 44. Compared to all medical conditions, persons with mental illness have the second

lowest number of quality adjusted life years following stroke (Jacobs, Dewa, Bland et al., 2007). Providing effective treatment of mental illness is not just important for quality of life, but for preserving life as well.

The direct and indirect economic costs of mental illness are tremendous. In the United States, for instance, the total estimated cost of anxiety disorders in 1990 was $42 billion (Hoffman, Dukes, & Wittchen, 2008). This included estimated costs based on medical care, psychiatric treatment, lost productivity and insurance, pharmaceutical, and mortality. In Canada, the economic burden of depression and distress in 1998 was estimated to be $14 billion (Stephens & Joubert, 2001). The largest portion of this cost was attributed to lost productivity (i.e., lost wages, employer losses, etc.). Costs related to treatment included $642 million for medications, $854 for physician visits, and costs not covered by public insurance for psychologist and social work at $278 million. Above all other costs, hospital-based care was the largest cost related to mental illness at $3.9 billion. These costs represent substantial burdens, not only to private or public payers, but to persons suffering from mental illness who may have limited funds to pay for medications and treatments.

The broad scope and impact of mental illness puts an even greater emphasis on the need to measure and ensure quality of care. Reducing the burden of mental illness will require that mental health treatments and services are delivered consistently, appropriately, and effectively. Quality measurement can be used to determine if and how these requirements are achieved and identify ways to improve mental health treatment. Offering and delivering services to attempt to alleviate the burden of mental illness on society and to the individual are important, and knowing that these services appropriately and effectively meet the needs of all stakeholders is essential.

The structure of mental health services in Ontario is complex involving a number of services offered in inpatient or outpatient settings. Outpatient services are generally the point of first contact for persons in need of care, usually originated through their primary care physician (Steel, McDonald, Silove, et al., 2006). Outpatient services can include intensive case management services affiliated with hospitals, stand-alone agencies, private practices, and primary care physicians (Goering, Wasylenki, &

Durbin, 2000). Aside from private practice therapists and primary care physicians, outpatient services involve long term case management, substance-use treatment, vocational rehabilitation, or crisis outreach and support. In Ontario, specialized outpatient services are available from over 60 teams that provide assertive community treatment (ACT). ACT is becoming a more common approach for community-based services as it includes specialized treatment teams who care for persons with complex, often chronic, mental health conditions (George, Durbin, & Koegl, 2008). When used effectively, ACT can reduce hospitalizations providing more efficient methods of delivering specialized care while helping persons maintain community tenure (Latimer, 1999).

Inpatient services are usually delivered in two hospital settings: specialized psychiatric hospitals (SPHs)[1] or general hospital psychiatric units (GHPUs; Goering et al., 2000). Specialized hospitals include services designed for specific mental health and addictions conditions (e.g., Eating disorders treatment, Concurrent programs for post traumatic stress and addiction, etc.). Services in SPHs are typically considered long-stay and are usually designed for specific mental health conditions (e.g., eating disorders, mood/anxiety, trauma, etc.). Services at SPHs may include fixed programs such as group therapies and activities where all patients participate. The majority of inpatient services are delivered in general hospital psychiatric units (GHPUs) that provide crisis stabilization and assessment to the acutely ill. Programming may be more individually oriented with access to specific services or programs (e.g. recreation therapy) based on referrals from the GHPU team. Stays on GHPUs are often shorter than SPHs and the mix of patient characteristics treated on one unit are often broader. A number of SPHs and GHPUs in Ontario also include beds designated for forensic psychiatric patients who are admitted due to court orders for psychiatric assessment. Forensic patients may also be receiving treatment as a result of a crime to which they were found not criminally responsible due to a mental illness and ordered to receive mental health treatment. While forensic patients are usually

[1] Formerly Provincial Psychiatric Hospitals.

admitted through the criminal justice system, most persons are referred to inpatient services from community based mental health or primary care services.

Most recipients of mental health services receive outpatient services, usually from primary care physicians and private practice clinicians. Among hospital based services, the majority are provided through GHPUs. The Canadian Institute for Health Information (CIHI) recently reported that 87% of mental health hospitalizations were in GHPUs but 51% of all hospital days were accounted by hospitalizations in SPHs (CIHI, 2008). Regardless of hospital type, schizophrenia, mood, and substance-use disorders accounted for the highest percentage of hospitalizations and organic (e.g. dementia), mood, and schizophrenia diagnoses account for the longest lengths of stay. In both types of hospitals, there has been a decreasing trend in both the number of separations and lengths of stay (LOS) since 2000/2001. In GHPUs the average LOS went from 36 days in 2000/2001 to 16 days in 2005/2006. A similar trend was found for SPHs where the LOS dropped from 160 days to 100 days in the same period.

The changing trends in service use toward CMH services are in line with reports that have emphasized a shift to community-based health care delivery (e.g., Kirby, 2004c; Commission on the Future of Health Care in Canada, 2002). These trends also raise questions around how the quality of mental health services is affected by changes in service use. For instance, how do fewer and shorter inpatient episodes affect short-term and long term outcomes experienced by service recipients? While system level analyses of such questions may be optimal they are difficult due to a lack of integrated information available between mental health sectors (New Freedom Commission on Mental Health, 2003). Sector specific evaluations may lend insight into the quality of mental health care by providing information about issues or concerns not relevant to other sectors. For inpatient mental health, admission to discharge evaluations of quality are important given the severity of conditions among mental health inpatients and the trend toward shorter lengths of stay. Evaluations of inpatient mental health services are needed for capturing information on the types of inpatient services received, their access and appropriateness, and their impact on those who receive them. Before discussing how

evaluations of the quality of inpatient mental health services can be measured it is important to consider the structure to which these services are governed and managed in Ontario.

1.2 Regional Management of Mental Health Services in Ontario

In 2006, the Ministry of Health and Long Term Care (MoHLTC) implemented a regionalized system to administer and manage health services, including mental health, by creating Local Health Integration Networks (LHINS). LHINS are not-for-profit corporations responsible for all planning and funding of existing or new health service providers. With this funding, LHINs establish service accountability agreements with health service providers that establish the nature, scope, and volume of services to be offered. Accountability agreements are also established between the MoHLTC and LHINs that outline service expectations, budgets, and expected service and system outcomes. Key processes implemented by LHINs are Integrated Health Service Plans (IHSP) used to develop plans for integrated services and determine health priorities with providers within each LHIN. The key areas of emphasis for IHSPs include (Bhasin & Williams, 2007):

- Renewing community engagement and partnerships concerning health care
- Improving the health status of Ontarians
- Ensuring equitable access to health care for all Ontarians
- Improving the quality of health outcomes
- Establishing a framework for a sustainable health system

The 14 LHINs were designed based on population size, not the geographic size, of the regions. Geographic distribution and specific characteristics associated with the population size and number of mental health treatment facilities in each LHIN can be found in Appendix A. Population sizes of LHINs range from 241,000 in the North West to 1,577,000 in the central region. The number of GHPUs in LHINs ranged from 5 to 25 while the number of SPHs among LHINs ranges from 0 to 3. The Canadian Institute for Health Information (CIHI) has reported a number of health service indicators by LHINs (see Table 2, Appendix A). Among all hospitalizations, the standardized

separation rate per 100K population ranged from under 300 to over 800 and the standardized rate of days stay per 100k ranged from about 3,600 to over 9,700. Separations refer to any instance where a patient leaves hospital due to death, discharge, transfer, or sign-out against medical advice. The 30-day readmission rate among patients with primary psychiatric diagnoses in acute hospitals ranged from about 6% to almost 12% and 1-year readmission rates among the same groups ranged from 18% to almost 28%.

The relationship between governance, accountability, and quality is a key focus of LHINs. A number of studies examining accountability have focused on how quality relates to funding of mental health services. A number of studies examining regional differences in health spending in the United States have found regions that spent more, per capita, did not perform better in several aspects of quality including appropriate use of procedures (Fisher, Wennber, Stukel, et al., 2003a), perception of care (Fowler, Gallagher, Anthony, et al., 2008), and survival following heart attack (Fisher, Wennberg, Stukel, et al. 2003b). Interestingly, physicians in some high spending regions with a greater number of hospital beds have actually reported more difficulty gaining access to these beds for their patients (Sirovich, Gottlieb, Welch, & Fisher, 2006). Based on these regional differences, calls have emerged for greater linkages between quality and funding of health services to correct flaws such as funding for more services, regardless of service quality (Wennber, Fisher & Skinner, 2002; Fischer, Goodman, Skinner, & Bronner, 2009). Quality measurement will be important to evaluate the impact of LHIN governance of funding coupled with LHINs' focus on working with community health service providers to develop plans to improve integration and quality.

Regionalization may have specific benefits for mental health services but there is also concern about fragmentation of services between regions. For instance, forensic mental health teams in Massachusetts reported being better able to manage re-entry to community among persons with mental illness recently after regionalized management of services was established (Hartwell, Fisher, & Deng, 2009). While specific clinical benefits can be identified within regions, the lack of uniform implementation of specialized practices and treatment (e.g., ACT) threatens the quality of mental

health services across regions (Latimer, 2005). Services that include public funding and interdisciplinary human resource strategies have been identified as one of ten critical success factors for LHINs (Secker, Goldenberg, Gibson, et al., 2006). However, not all LHINS have adopted mental health care as a priority. For instance, patients in seven of the fourteen LHINs that have identified mental health services as a priority were more than twice as likely to see a psychiatrist and over 20% more likely to see a social worker, occupational therapist, and dietician compared to LHINS where mental health services were not a priority (Martin & Hirdes, 2008). The challenge for implementing regionalized services will be to establish and implement consistent standards and practices of quality mental health services across regions, while responding to specific needs within regions.

1.3 Directions for Mental Health Services in Canada

In the last 10 years, several initiatives have reviewed mental health services at National and Provincial levels. In 2004, retired Senator Michael Kirby led a series of commissioned reviews through the Standing Senate Committee on Social Affairs, Science and Technology (also called the Kirby Commission) of perceptions of mental illness, mental health, and addictions (Kirby, 2004a; 2004b; 2004c). The aims of these reports were to document the incidence and economic burden of mental illness in Canada and to review mental health strategies and services used in other countries. These reports were undertaken to develop a National strategy for mental health and mental illness by focusing on prevention, promoting access to services, providing support to family and caregivers, and stimulating research.

A key component of the first report was to review practices and issues that were common across all Canadian provinces (Kirby, 2004a). The findings indentified:

- Needs for community based services for prevention and rehabilitation,
- Unequal distribution and quality of services, particularly in rural regions,
- Needs for improvement at the primary care level,
- A lack of human resources such as psychiatrists and psychologists,

- Needs for early intervention strategies,
- System fragmentation and a lack of integration.

Performance and accountability were common themes throughout the three reports. A lack of accountability was evident at the system level, primarily in terms of a clear distinction as to where the responsibility for mental health issues and services ultimately rest. A primary recommendation of the Kirby Commission was to establish a National Health Care Council to measure and report system performance by evaluating cost-effectiveness, efficiency, quality, and patient outcomes. The Kirby Commission also called for the clarification of the roles and responsibilities of different levels of government and an improvement in the sharing of accountability across these levels. A national data collection system that could be used consistently across provinces was also promoted to improve integration and accountability of mental health services (Kirby, 2004c). These recommendations emphasize the need for a common strategy to measure quality of mental health care.

In many provinces, including Ontario, actions had already been implemented to build a foundation to improve quality and accountability of mental health services. The Ontario Ministry of Health and Long Term Care (MoHLTC) has outlined a vision and principles for reform of mental health services to improve expectations and standards to which services are delivered (Ontario MoHLTC, 1999). The document, *Making It Happen*, emphasizes the importance of providing effective and early treatment to individuals with severe mental illness characterized by multiple needs and to encourage an active role for care recipients and families in treatment decisions, service planning, and evaluation. *Making It Happen* provides a framework for the roles of inpatient and outpatient services and includes a recommendation to develop operational goals and performance indicators to illustrate system/service responsibility and accountability.

In 2003, the MoHLTC released the *Mental Health Accountability Framework*, a guideline for monitoring the accountability, efficiency, and effectiveness of mental health services and supports (Ontario MoHLTC, 2003). A central component of this framework is to conceptually describe a set a of performance domains and performance indicators for use in hospital report cards, service improvement

initiatives, accreditation, operating plans, and other operational or quality review activities. Eight performance domains and 69 performance indicators were established. The majority of these indicators describe the process and structure of service delivery with the goal that clinical outcome based indicators be added in the future. To date, the most common outcome indicators used at a provincial level are based on perception of and satisfaction with care. No indicators measuring change in clinical, social, or role functioning, for example, have been implemented. Due to recent mandates of comprehensive assessment systems and reporting infrastructure, there exists an opportunity in Ontario for implementing QIs based on clinical outcomes that is internationally unique. Before discussing this opportunity, an introduction to the construction of quality indicators and the role of risk adjustment in measuring quality is provided. This is followed by a description of mental health QI development activities and research internationally and in Canada.

1.4 Definition of Quality and Quality Measurement

Defining quality of care is complicated by the multidimensional perspectives of health stakeholders. Recipients of health services may have different opinions from those who deliver health services, and perspectives may vary even within stakeholder groups. At a simplistic level, quality health care includes the maximization of benefit with the minimization of risk to the patient (Donabedian, 1980). This involves technical quality where treatment leads to desired goals (e.g., improvement, maintenance of functioning, etc.) without introducing excess opportunity for decline or harm and patient centred care where treatment is humane, appropriate, and includes patient choice (Brook, McGlynn, & Shekelle, 2000). To promote quality, practice standards and guidelines of care are provided by governing bodies, professional health care associations, and patient advocate groups to drive good health care quality through the delivery of effective, appropriate, and safe treatment. Adherence to these standards and their impact on service recipients can be evaluated using quality measurement (Fauman, 1989).

Quality measurement involves the utilization of available information to construct quality indicators (QIs) to evaluate health services. Quality indicators provide markers for different aspects of quality but are not definitive measures of quality (Bowen & Kreindler, 2008). Definitive quality measures are rare as they are difficult to quantify; encompassing definitions of good or bad quality using quality measures are typically not appropriate (Lifford, Mohammed, Spiegelhalter, & Thompson, 2004). Indicators of quality are more tangible as they provide flags of trends and potential quality problems. Conclusions about quality of care are often reached by measuring and comparing a number of QIs designed to reflect different aspects of quality of health services.

Most QIs are designed to reflect three dimensions of health care: structures, processes, and outcomes. Using these dimensions, good quality is said to occur when the needed and appropriate health care structures are in place so that suitable process of care can be delivered to achieve optimal outcomes (Donabedian, 1980; Donabedian, 1982). Structures typically refer to the characteristics of the health care delivery system (e.g., doctor to patient ratio) but can also include the governance and financial mechanisms for health care. Process indicators typically measure the type and intensity of services available or offered (e.g., medication prescribing based on best practice guidelines). Outcomes reflect the results of a person's interaction with health care and usually provide information on the impact of the structure and process of care on service recipients (McGrath & Tempier, 2003). Outcomes can include, but are not limited to, changes in symptoms and functioning throughout or following treatment, the prevalence of adverse events, and mortality (Hermann, Rollins, & Chan, 2007; Iezzoni, 2003). Using structure, process, and outcome QIs information on the organization, delivery, and impact of health services can be evaluated.

For over 30 years there has been debate about whether process or outcome QIs provide the best description of health care quality (McAuliffe, 1979). In most health sectors, process measures are the most commonly used QIs. Since process QIs are often based on administrative databases common across facilities (e.g., pharmacy or billing information) they tend to be more feasible to calculate. Process indicators can help clinicians target where quality improvement efforts should be targeted by

directly identifying inappropriate or inadequate interventions (Hermann & Palmer, 2002; Lilford, et al. 2004). Process measures are effective for measuring specific aspects of health care but are limited in their ability to account for less tangible aspects of care (e.g., clinical experience, therapeutic relationship, etc.) that could affect quality (Mant, 2001). Outcome QIs, however, may provide more meaningful information about quality because they are a reflection of the impact of the process of care (Hermann, 2005; Srebnik, Hendryx, Stevenson et al., 1997). Outcome QIs have clinical relevance for describing change regardless of treatment offered, their application from multiple perspectives (e.g., observational or self-rated outcomes) at multiple levels (e.g., unit, hospital, region, province, etc), and their efficiency in describing the overall impact of care (Hermann & Palmer, 2002). A mixture of process and outcome measures is optimal in order to have a balanced quality measurement system linking the availability and delivery of services to their impact on service recipients (McGrath & Termpier, 2003; Hermann, 2005). In combination with process indicators, the measurement of outcomes can promote evidence based practice, closing the gap between the science of what constitutes effective care and what care is actually delivered (Lehman, Goldman, Dixon, & Churchill, 2004; Hermann, et al. 2000).

1.5 Designing and Using Quality Indicators

There are several considerations for choosing and designing effective QIs, regardless of the type of quality being measured. At a high level, QIs should be *meaningful*, *feasible*, and *actionable* (Hermann & Palmer, 2002). A *meaningful* QI addresses a problem area, is clinically important, evidence based, and psychometrically sound. The interpretation of the QI has to be based on carefully defined criteria that specifically reflect a domain of quality (Hermann et al., 2000). For instance, the prevalence of certain psychiatric symptoms or diagnoses at admission may be clinically meaningful but may not be a reflection of poor quality. Instead, a high prevalence of physical restraint use may be a more meaningful QI because it indicates use of inappropriate treatment. A *feasible* QI is efficient and based on available yet meaningful data (Fries, Morris, Aliaga, & Jones, 2003). The utility of QIs is

limited when standard, reliable, valid, and accessible data needed to calculate QIs is not commonly available. Further, the calculation of the QIs based on commonly available data needs to be performed uniformly and consistently to ensure the accuracy of results over time. Finally, an *actionable* QI needs to be modifiable by revealing information that can be used by clinical staff, administrators, policy makers or researchers to take actions towards improving the quality of care. For instance, linking outcome QIs to a care planning process or tangible treatment options creates a feedback loop that can drive continuous quality improvement. As treatments and services are implemented the QI will evaluate the impact and effectiveness of those services which can drive further inquiry and improvement to the care processes.

QIs are typically constructed as prevalence or rate based measures. Prevalence measures are based on single point of assessment and consider a certain threshold of an event at a given point in time as an indicator of quality. For instance, the percent of inpatients with schizophrenia that are prescribed an appropriate dosage of antipsychotic medication following an initial mental health assessment is a prevalence based process QI (Hermann, Finnerty, Provost, et al, 2002). In most cases, prevalence QIs should not be based on the admission assessment to a facility as the facility will not have had time to intervene in the quality problem (Hermann, 2005). However, in some cases where QIs reflect processes such as physical restraint use, admission prevalence QIs may be considered. The threshold of an acceptable level for that QI will need to be determined through risk adjustment and benchmarking (Hermann & Provost, 2003).

Rate based QIs describe change in individual outcomes or treatment patterns based on separated periods of observation. Rate QIs can measure outcomes from the perspective of improvement (e.g., a reduction in symptoms), failure to improve (e.g., no change in symptoms that are expected to change), or incidence (e.g., increase of symptoms). Since failure to improve in or incidence of symptoms or functioning may both be considered an adverse event, they can be combined as a single QI (Hirdes et al., 2004; Jones et al., in press). To use rate based QIs the time between measurement points needs to be sensitive to the period of time in which the outcome is expected to occur. For

instance, should the QI reflect an outcome to occur by discharge, prior to discharge, or at some point after discharge? Without defining this time period, QIs may be misleading by over-reporting what is perceived as good or poor quality. The issue of time between assessments will be discussed further in later sections of this introduction.

Choosing the appropriate numerator and denominator is a critical step in defining prevalence and rate QIs. Typically, certain groups of individuals are excluded from the denominator if they are not at risk of the quality event. For instance, improvement in physical functioning can only occur in individuals with physical functioning deficits. The numerator is determined by specifically defining the quality event. The quality domain of interest and the measures available to assess this domain are important for defining the numerator to protect the content validity of the QI (Fries et al., 2003). For instance, if improvement in cognitive symptoms is the outcome of interest but the only measure available is the number of patients who completed a cognitive rehabilitation group the QI can only measure utilization of cognitive rehabilitation, not improvement in cognition. The definitions of numerators and denominators need to be liberal enough to detect a quality problem yet conservative enough not to overestimate that problem (Berg, Fries, Jones et al. 2001). Poorly defined numerators and denominators for QIs risk inaccurate interpretation, generalization, and conclusions about quality.

The purpose of measuring quality of care is an essential consideration when constructing QIs. Depending on the purpose, the expertise of the target audience for interpreting QIs will vary greatly thus driving the need for accurate yet interpretable QIs. Stakeholders of QI information include healthcare providers, consumers, researchers, quality improvement organizations, public reporting agencies, and health services funders (Hussey, Mattke, Morse, & Ridgely, 2007). Information from QIs aimed at informing quality improvement activities at the agency level or by benchmarking against other agencies are likely to be used by individuals familiar with practice patterns and staff to verify results. Accreditation agencies such as Accreditation Canada or the Joint Commission of Accreditation of Health Organizations (JCAHO) in the United States have expertise in evaluating health quality and are able to verify results by interviewing staff and patients. Consumers of reports on quality from the

general public may be less knowledgeable about health care and, particularly the technical aspects of the interpretation of quality. As such, QIs need to be scientifically and clinically sound to ensure they are true indicators of a quality problem but need to be presented in a way that can be easily understood by lay audiences.

Once QIs are developed or chosen, methods for using them to evaluate and compare quality need to be considered. Sources for variation in Outcome QIs, for instance, can include actual differences in quality of care as well as intrinsic patient characteristics (Mant, 2001). Intrinsic patient characteristics, or risks, are variables related to QI scores but are not influenced by or reflective of the quality of care that is delivered (Iezzoni, 2003). When these variables, or their prevalence, are unequally distributed among facilities being compared, QI comparisons might lead to misleading assumptions about facilities that are providing better or worse care (Mor, Berg, Angelelli, et al. 2003). Therefore, consideration for risk adjusting QIs is important for ensuring that differences in quality can be attributable to an actual quality problem and not the inherent characteristics of patients admitted to each facility. The next section will describe risk adjustment for quality measurement in more detail.

1.6 Risk Adjustment of Quality Indicators

1.6.1 Definition and Rationale for Risk Adjustment

Risk adjustment (RA) is a technique for controlling, often statistically, the unequal distribution of intrinsic individual characteristics on health care utilization, costs, and outcomes (Hermann et al., 2007; Zimmerman, 2003). In health economics and accounting, patient factors are adjusted to predict costs based on level of utilization of health services (Lorenz & Sederer, 2001). These practices can prevent under selection of high cost patients into health services by providing fair compensation for treating high cost users (Hendryx, Beigel, & Doucette, 2001). For quality comparisons, RA is useful for insuring comparisons between health providers are more equitable by not penalizing providers who treat higher risk patients (Iezzoni, 1997; Richardson, Tarnow-Mordi, & Lee, 1999). Given the focus of

this dissertation is on quality measurement in health care, financial or economic risk adjustment methods will not be reviewed.

There is debate as to whether RA should be used for measuring health care quality (Arling, Karon, Sainfort, et al., 1997; Fries, Morris, Aliaga, & Jones, 2003; Hirdes et al., 2004). The argument against risk adjustment centers on a concept of recognized risk: Facilities that accept "riskier" patients (i.e., those that have a greater likelihood of a quality problem) should put in place policies and procedures to appropriately treat these patients to prevent the quality problem. Under this assumption, RA would benefit facilities that accept riskier patients and could promote poorer quality of care (e.g., more incentive to admit high risk patients and less incentive for implementing high quality services for those patients).

Counter to the assumption of recognized risk is that risk adjustment, when applied appropriately, can remove incentives and promote better quality care. Risk adjustment can improve the clinical, administrative, and economic transparency of quality reporting (Lorenz & Sederer, 2001). Clinicians may be more willing to participate in reporting and benchmarking activities because of the assurance risk adjustment provides against penalizing treatment of high or low risk patients groups. Risk adjustment may also improve the interpretation of quality reports so that stakeholders can fairly evaluate and compare service providers, reinforcing to the public that funding for health care funding is applied to high quality services. By providing meaningful interpretations of QIs with RA, clinicians and decision makers can identify opportunities for quality improvement by sharing best practices between stronger and weaker performing providers (Lied, Kazandjian, & Hohman, 1999). Without RA, this opportunity may not exist or may be based on inaccurate conclusions about which providers perform better or worse. Also, certain adjustment procedures allow for the identification of interactions between risk adjusters and the quality of care provided (Zimmerman, 2003; Hendryx, 2005). For instance, if cognitive functioning is used to adjust a QI measuring improvement in aggressive behaviour and results in reduced rankings of certain facilities, investigations can be done to understand what contextual factors contribute to poorer performance in these facilities (e.g., staff specialization).

Accounting for the intrinsic patient characteristics allows for the identification of services characteristics that affect quality of care.

Risk adjustment is also important for providing fair comparisons of quality by controlling for selection bias of patients into treatment. In health research, randomized control trials (RCT) are viewed as the gold standard for measuring treatment impact on outcomes (Iezzoni, 2003). While conclusive and effective in artificial treatment environments where patients can be carefully selected and assigned to groups, RCTs are not realistic for measuring quality of care in natural treatment settings. The issue with studies using controlled experiments is that the link between natural clinical practice and outcome is lost (Dickey, Hermann, & Eisen, 1998). In many instances facilities cannot control the intrinsic characteristics that patients bring into treatment (Iezzoni, 2003; Richardson et al., 1999). In other instances, particularly for specialized facilities, only patients deemed appropriate or in need of specialized treatment are admitted. Although perfectly equivalent groups cannot be created and the rigor of experimental designs cannot be matched, quality measurement needs risk adjustment to account for such indirect or direct instances of selection bias (Iezzoni, 1997).

The choice to risk adjust a QI, the variables to adjust, and the method to perform the adjustment could all have dramatic effects on policies, funding, access, reputation, and perceptions of health care providers (Shahian & Normand, 2008). Considering the extrapolated conclusions that media and the lay public may form based on public reports of quality, the accuracy and appropriateness of risk adjustment is essential. Therefore, important considerations for choosing whether RA is needed include what variables to adjust, and how to adjustment should be done.

While RA should be considered for all QIs, there are instances where RA may not be appropriate (Hermann et al., 2007). When evaluating the quality of care for an individual patient, RA may not be needed unless individual results are to be compared to standards or benchmarks. Also, some process QIs that are fully under the control of the facility and have well defined denominator groups may not require RA (e.g., the availability of ECT services for individuals with severe depression). Finally, instances where RA leads to the same results as unadjusted QIs, the gain from risk

adjustment is small relative to the cost, or there is a lack of data on risk factors it may be more appropriate to report unadjusted QIs (Hendryx, 2004). In the latter case, it may also be inappropriate to report QIs at all. The lack of available information to perform RA should not negate the recognition that RA is needed.

1.6.2 Choosing Risk Adjustment Variables (RAVs)

Choosing appropriate risk adjustment variables (RAVs) is, perhaps, the most important step in RA. Hendryx and Teague (2001) suggest "the identification and testing of risk variables and models should be addressed with no less care than that invested in the development of performance indicators" (p.254). The method of RA chosen and interpretation of results will be obsolete or misleading if appropriate RAVs are not chosen (Weissman, Rosenheck, & Essock, 2002) (Iezzoni, 2003; Cuffel, 2004). Table 1 summarizes a list of guidelines that have been suggested by experts to inform the choice of useful risk adjusters for mental health QIs (Hendryx, 2004; Hendryx & Teague, 2001).

Table 1. Criteria for selecting risk adjustment variables *

Criteria	Description	Examples
1. Reliability and Validity	Based on quality data producing consistent and accurate ratings	Internal consistency, inter-rater reliability, content validity
2. Correlated to QI	Statistically related to the QI in a multivariate context	Cognitive impairment and aggressive behaviour
3. Outside of Facility Control	Factors not influenced by or related to facility actions	Severity of illness at time of admission
4. Variability Among Facilities	Providers differ in rates or prevalence of risk adjuster	Significantly different prevalence of dementia diagnoses
5. Theoretical or Clinical Relationship to Outcome	A priori relationship established through prior research or clinical experience	"Medical meaningfulness" (Iezzoni, 2003, p.33)
6. Not Susceptible to Manipulation or Gaming	No incentive for manipulation to improve ranking	"Up coding" diagnoses that result in more favorable performance
7. Influences Performance Interpretation	Inclusion of risk adjuster leads to different interpretations of performance than unadjusted results	Significantly changes performance rankings of some comparison groups
8. No Disadvantage to Vulnerable Groups	Reflects risk of outcome, not risk of provision of poor quality of care	Adjustment on race would mask racial differences in quality of care.

* adapted from Hendryx, 2004

The nature of the relationship between RAVs and the QI is important to consider for determining which RAVs are appropriate. The tendency may be to include all variables that are significantly related to the QI of interest to account for all possible variance of QI scores leaving on differences in quality of services provided; however, the purpose of RA is not to explain as much variance of the QI as possible, but to control for factors that influence that QI but are beyond a facility's control (Hirdes et al., 2004). While a statistical and clinical relationship between a RAV and a QI is important, the degree to which a potential RAV represents a quality problem, itself, will make such a variable problematic for evaluating quality (Hendryx & Teague, 2001; Lorenz & Sederer, 2001). Therefore, a first step in choosing RAVs is to determine which variables can be considered adjustment variables and those considered contextual variables (Lin, Degendorger, Durbin et al., 2001). While both adjustment and contextual variables are related to a QI, adjustment variables are not related to the treatment or services provided while contextual variables can include treatments or services. Contextual variables should not be included in RA because they, themselves, may be the source of a quality problem. For instance, analgesic medication use might be strongly related to improvement in pain but adjusting the use of this medication to measure pain could mask inappropriate usage of analgesics. Adjustment variables should be used for risk adjustment to compare quality while contextual and adjustment variables can be used to explain differences in quality.

The issue of gaming should also be considered when choosing variables to adjust. Gaming refers to a facility's ability to manipulate the reporting of data to alter the results of quality measurement (Hendryx, 2004). If a facility is adjusted for a RAV that they should be able to influence through interventions, there is more incentive for that facility to over-report that RAV less incentive to implement interventions for that RAV. This process is similar to *risk selection* in economic research where health insurers use risk adjustment to choose clients with risk profiles that predict less service use (Lorenz & Sederer, 2001). Careful consideration is needed to select variables that are not susceptible to gaming to prevent such events from skewing true estimates of quality of care.

1.6.3 Types of Risk Adjustment Variables

There are a wide variety of risk adjusters that have been identified for QIs in health care. Domains typically used for risk adjusting QIs include sociodemographic variables, prior service utilization, and diagnostic or clinical status (Iezzoni, 2003; Hermann et al., 2007). There is debate as to whether sociodemographic variables should be included as risk adjusters. Variables such as age, sex, ethnicity, and socioeconomic status tend to be the most commonly cited risk adjusters for mental health outcomes and QIs (Banks, Pandiani, & Bramley, 2001; Hermann et al., 2007). Some mental health symptoms change with age, can be expressed differently in men and women, and may vary in prevalence based on ethnicity and socioeconomic status. However, adjusting for sociodemographic variables could actually mask the identification of important quality problems (Iezzoni, 1997). For example, in a large sample study of older adults in the U.S., African American patients were found to have received worse quality of care than Caucasian patients in the U.S. (Schneider, Zaslavsky, & Epstein, 2002). While the provision of quality care may not be directly racially motivated, there may be differences in the quality of treatment facilities to which different racial groups are admitted. For instance, black residents admitted to Medicaid-reimbursed nursing homes that were primarily homogeneously black had a higher number of hospitalizations than nursing homes with a heterogeneous mix of black and white residents (Gruneir, Miller, Feng, Intrator, & Mor, 2008). These results suggest that race, per se, did not drive quality; but, instead, the types of facilities an individual is able to obtain service (either due to finances or geography) might influence the quality of care. Adjusting for race would actually mask the quality problem (i.e., inequality in access to quality treatment).

Several service utilization variables such as prior hospitalizations, prior outpatient visits, and length of stay in hospital have also been used as RAVs (Greenberg & Rosenheck, 2005). While these factors might affect outcomes of subsequent hospital stays, they may themselves be indicators of poor quality of care, particularly at a system level. For instance, adjusting for length of stay may actually create a disincentive for facilities to efficiently treat an individual and facilitate a return to a less

restrictive care setting, such as community tenure, in the fewest days possible. Instead, it may be appropriate to reward hospitals who are able to achieve such an outcome by creating time-dependentQIs. Ultimately, the choice of using utilization RAVs will depend on the purpose of quality measurement and the QI being measured. When comparing across facilities, utilization factors may be better studied as contextual variables.

Diagnostic and clinical status are considered stronger measures for risk adjustment than socio-demographic or utilization information. While sociodemographic variables may be related to a QI, they are often proxies for underlying clinical status that may be directly related to outcomes (Nicholl, 2007). Diagnostic and clinical factors include mental health or medical diagnoses, concurrent illness, measures for illness severity, and functional status. Clinical data drawn from multiple information sources (e.g., family, patient, referral source, etc.) and multiple perspectives (i.e., multidisciplinary treatment team) is recommended rather than diagnoses, alone, to protect against gaming and account for illness severity (Hendryx et al., 2001). Unfortunately, there has typically been a lack of good quality data across a health system for developing clinically based RAVs. Administrative data that include minimal information about patient demographics, diagnoses, pharmacy, and financial information are typically most commonly available for risk adjustment (Iezzoni, 2003). In health sectors such as long term care (Morris, Nonemaker, Murphy, & Hawes, 1997) and Home Care (Hirdes et al., 2004), rich sources of clinical data are available across facilities that can be used to develop clinically meaningful RAVs. Later in this introduction the use of similar data sources in inpatient psychiatry will be identified.

The choice of RAVs depends on data available, the relationship between the RAV and the QI, and the distribution of the RAV among groups being compared. Consideration is also needed for the relationship between the RAVs chosen for each QI. The use of too many or too few RAVs could lead to over-adjustment of a QI, essentially masking existing quality problems that do exist (Dalby, Hirdes, & Fries, 2005). Over adjustment could occur if the RAVs, themselves, are strongly correlated with measures of the same construct, are not related to selection bias among facilities, or have no clinically

meaningful relationship with the QI being compared (Day, Byar, & Green, 1980). Spurious RAVs will affect the precision to which QIs are compared among facilities when performing quality measurement (Schisterman, Cole, & Platt, 2009). Therefore, the choice of RAVs will depend on the existence of meaningful and independent relationships with the QI.

1.6.4 Techniques for Risk Adjustment

A number of risk adjustment techniques for measuring and comparing quality are available. The approaches range from simple techniques using uncomplicated univariate comparisons to complex multivariate models and stratifications. These techniques can be divided into indirect and direct methods of adjustment. The following are brief descriptions of risk adjustment techniques that have been described in a number of guides and introductions to risk adjustment for measuring quality (Hendryx et al., 2001; Hendryx, 2004; Iezzoni, 2003).

1.6.4.1 Indirect Techniques: Linear, Logistic, and Hierarchical Regression

Regression models are the most common methods of risk adjusting QIs. Linear regression such as Ordinary Least Squares is used for QIs with continuous scores (e.g., percentage change) while logistic regression is used for dichotomous (e.g., occurrence of improvement) QI scores (Shwartz & Ash, 2003). Using these approaches, QIs are first modeled at a population level (typically across all cases in data available) with RAVs as the independent variables to obtain parameter estimates for use at the individual level. Scores for each risk adjuster are then entered into the equation to determine the expected score for each person. The predicted scores can be averaged for each group being compared (e.g., hospital) to produce each group's predicted score (Daley, Iezzoni, & Shwartz, 2003). The ratio of observed to expected QI scores is then calculated and multiplied by the population mean QI score(e.g., mean across all hospitals) to produce the risk adjusted QI score (Berg et al., 2001). Essentially, the adjusted score represents the QI score expected if the hospital admitted patients with an average case mix (Morris, Murphy, Mor, et al., 2002).

Hierarchical regression models are also becoming more common techniques for risk adjustment. Extensions of linear and logistic regression, hierarchical regression model parameters that vary at different levels; for instance, at the individual, facility, and regional level (Raudenbush & Bryk, 2002). Hierarchical models address some of the disadvantages of traditional linear or logistic regression. For instance, traditional regression techniques cannot account for non-random relationships between observations within provider groups i.e., nested effects where observations are clustered within hospitals (Lambert, Doucette, & Bickman, 2001; Cuffel, 2004). Since individuals are not randomly distributed among facilities, hierarchical models can be used to account for clustering of within-group observations. Also, facilities with small sample sizes that would have otherwise been excluded or produced unstable QI rates using tradition regression can be included using hierarchical models to produce more reliable QI rates (Burgess, Christiansen, Michalak, & Morris, 2000a). Hierarchical regression can include facilities with small samples by controlling regression-to-the-mean. Regression-to-the-mean is the tendency, particularly in smaller groups, for more extreme (outlier) pre-test scores to have post-test scores closer to the grand mean, regardless of events (e.g., interventions) that occur between pre and post tests (Morton & Torgerson, 2009).

1.6.4.2 Direct Adjustment Methods: Stratification, Direct Weighting, and Propensity Scores

Stratification is the least complex method of directly adjusting QIs. In stratification, facility results are calculated within separate risk strata based on different levels or categories of the RAV (e.g., age groups; Berg et al., 2001). Each facility is assigned a QI score in each strata and results between facilities are compared within each strata. Multiple QI scores for each facility could be helpful for identifying specific types of individuals who are experiencing good or poor quality. For instance, are outcomes as good among individuals with mildly severe depressive symptoms compared to those with very severe symptoms? Similarly, comparisons can be made in quality for individuals with mild symptoms in one facility compared to individuals with mild symptoms in other facilities. This approach is useful if one or two risk adjusters are needed but may be cumbersome with more than two

risk adjusters (Hendryx et al., 2001a; Berg et al., 2001). Results of basic stratification are easy to interpret for almost any stakeholder, regardless of research or statistical experience.

A more complex method of stratification involves weighting strata QI scores. The QI score within a stratum is multiplied by the distribution of that stratum within the total sample and summed with all other weighted strata scores to produce a single QI score for each facility (Ash, Swartz, & Pekoz, 2003). For instance, if a stratum of patients with mild depression symptoms in one hospital had a QI score of 10 and patients with mild depression represented 25% of all individuals from all hospitals being compared, then the adjusted score from this strata would be 2.5 (10 x .25). If this same hospital had 4 other depression strata with adjusted scores of 5.2, 2.3, 1.8, and 1.2, respectively, the hospital's total risk adjusted QI score would be 13. This method accounts for population distributions of risk adjusters and may be more efficient for brief reporting since one score is produced rather than multiple scores across strata. This method requires careful specification of weights drawn from real populations to prevent errors in interpretation or else misclassification will result (Wilcosky & Chambless, 1985). For both methods of stratification, it may be difficult to determine objective intervals for strata leading to arbitrary "cut points" for risk adjusters based on continuous responses. Due to these shortcomings more advanced methods of direct adjustment such as propensity scores may be useful.

Propensity score adjustment is a newer technique being considered for QI research and reporting. Propensity scores represent the likelihood of being assigned to treatment (e.g., being admitted to a hospital) given the presence of selected covariates (Rosenbaum & Rubin, 1983). Propensity scores attempt to balance the distribution of baseline risk factors across comparison groups so that comparisons can be made within groups with similar profiles (Rosenbaum & Rubin, 1984). This technique is particularly useful for dealing with selection bias and has been applied in observational health services research where case-control assignment is not practical (Love, 2008; Vanderweele, 2006). For measuring quality, propensity adjustment accounts for the effect of RAVs on QIs by balancing the distribution of these risks instead of adjusting the effect of the RAV on the QI across individuals (Huang, Frangakis, Dominici, Diette, & Wu, 2005).

Propensity scores are calculated using regression modeling where all covariates related to an outcome (or QI) are included as independent variables predicting treatment group as dependent(Weitzen, Lapane, Toledano et al., 2004). Several methods can then be used to adjust QIs using propensity scores, the most common being stratification and matching (D'Agostino, 1998; Huang et al., 2005). In stratification, the estimated propensity scores are ordered from highest to lowest and stratified into quantiles (and most often quintiles) for each treatment group. For instance, if there were 10 facilities being compared, each facility would have 5 strata of propensity scores ranging from patients with low propensity (i.e., likelihood) of being admitted to that facility to patients with a high propensity or likelihood of being admitted to that facility, compared to all other facilities. The use of quintile strata are estimated to remove about 90% of the bias due to confounding by RAVs (Cochran, 1968), an estimation confirmed using quintile stratification on propensity scores (Leon & Hedeker, 2002). Patients actually admitted to a treatment group can then be compared to others that have similar propensity scores that were not admitted to the treatment group. Scores can also be combined across strata for each treatment group using the same process as stratification weighting. Propensity scores can also be used to match patients from different treatment groups with similar characteristics and compare their outcomes (Love, Cebul, Thomas, & Dawson, 2003). The danger with matching is that information about quality from cases that cannot be matched is lost (Austin & Lee, 2009). Therefore, while more technically accurate, matching may not be as useful for quality comparisons given the potential loss of information about quality among unmatched pairs, particularly from facilities with small sample sizes.

The ability to compare QI scores between groups where risk factors are balanced make propensity scores more accessible and transparent in reporting than regression adjustment (VanderWeele, 2006). A commonly cited disadvantage for applying propensity scores to risk adjusting QIs, however, is the difficulty of applying propensity scores when more than two groups are compared (VanderWeele, 2006; Hendryx, et al., 2001; Hendryx, 2004). Using multinomial regression, Imbens (2000) developed a method of applying propensity scores to compare more than 2 groups, a process

called the "multiple propensity score" (Wang, Donnan, Steinke, & McDonald, 2001). Huang and colleagues (2005) adopted this method and provide an in-depth illustration for using propensity scores to adjust satisfaction with asthma care by balancing the distribution of RAVs among 20 physician groups. For each patient, 20 propensity scores (1 for each physician group) were assigned to each patient using multinomial regression. Patients were then stratified into quintiles for each physician group and risk adjusted scores produced within each stratum. Propensity score odds ratios were then compared to risk adjusted odds ratios from hierarchical modeling. They found that there was a 75% difference between the two methods in the absolute rankings of physician groups and a 50% difference in quintile rankings (physician groups that moved into a different quintile rank). Similar methods have also been used for comparing quality of coronary artery bypass surgery (Shahian & Normand, 2008). In this study, propensity score stratification was used to ensure RAV balance between hospitals for comparison of mortality rates.

1.6.4.3 Factors affecting Risk Adjustment

A number of factors can be considered when determining which technique to use for risk adjustment of QIs. First, the number of RAVs will inform the simplicity of the risk adjustment technique to be used. Stratification or weighted stratification can be used for QIs that have only 1 or 2 RAVs as long as the RAVs can be easily divided into distinct strata. QIs that require adjustment of more than one RAV should be considered for indirect adjustment using regression or direct adjustment using propensity scores.

Second, the sample size among groups being compared is an important consideration for risk adjustment, particularly those concerned with direct adjustment using stratification. The risk of empty strata becomes a concern when the number of individuals within a comparison group is small, the strata are based on variables with little variability, or the number of strata for a risk adjuster is large (Wilcosky & Chambless, 1985). As such, direct adjustment techniques may not be useful for intra-provider comparisons such as program or clinician comparisons within a given facility. Sample size

can also be problematic for indirect adjustment techniques such as logistic and linear regression, particularly for facilities with small samples (Burgess, Christiansen, Michalak, & Morris, 2000). Due to regression to the mean, facilities with smaller samples may be more likely to have QI scores well above expected scores derived by regression simply due to chance variation (Ash, Shwartz, & Pekoz, 2003). Hierarchical models can account for regression-to-the-mean so that facilities with small samples do not have to be omitted (Burgess, Christiansen, Michalak, & Morris et al., 2000b; Huang et al., 2005).

Finally, the impact of selection bias and ascertainment at the facility level may influence risk adjustment results. Ascertainment bias refers to differences in the facility's ability to detect differences in patient characteristics, or quality problems, often due to differences in the experience of assessors (Berg et al., 2001; Morris et al., 2002). Variables usually prone to ascertainment bias are those that are subjective or difficult to directly observe (e.g., pain) and thus prone to differential effects of assessment. Selection bias, in the context of quality measurement, refers to differences in the characteristics, or case mix, of patients admitted to facilities being compared (Dalby et al., 2005). In psychiatry, for instance, one GHPU may treat a high percentage of patients with cognitive impairment and fewer with addictions while a second treats patients with more addictions and less cognitive impairment. If cognitive impairment and addictions are significantly related to the likelihood of an outcome, then those interested in evaluating such outcomes would need to control for the selection bias of patient characteristics between these two GPUs. In quality measurement of long term care in the United States, these two biases have been identified as potential sources of variability in facility QI scores over time (Morris et al., 2002).

Several approaches have been used to adjust for ascertainment and selection biases. The Facility Admission Profile (FAP) was developed for nursing home QIs in the U.S. to account for these biases using regression adjustment (Morris et al., 2002). The FAP reflects the proportion of individuals admitted to the facility with a condition that places them at high risk of triggering a QI condition. For instance, for a QI measuring change in depression symptoms the FAP would be the baseline prevalence of depressive symptoms. After entering the FAP to risk adjustment models with other RAVs, Morris

and colleagues found the FAP had a minimal impact on QI scores and were not recommended for use. A measure similar to the FAP called the Agency Intake Profile (AIP) was developed for Home Care Quality Indicators and compared to a case mix index (CMI) to adjust for selection and ascertainment biases (Dalby et al., 2005). The CMI score was based on a combination of clinical symptoms and represented a measure of clinical complexity. The AIP tended to minimize differences in QI rates between agencies and health regions and had a greater impact on change in QI rates than individual RAVs and the CMI. The authors recommended the use of the AIP in instances where a very conservative approach to risk adjustment is warranted, such as public report cards.

The use of facility level characteristics in an individual level model such as those used in studies of FAP may also be inappropriate. The issue is that logistic regression treats the provider effects as fixed meaning that individual observations are assumed to be independent (Cuffel, 2004). In cases where ascertainment or selection bias occurs, independence of observations within facilities cannot be assumed. Hierarchical models are able to account for facility RAVs as a source of random variation and can produce an estimate of the amount of variation in a QI that is attributable to the facility characteristics (Cuffel, 2004). However, while hierarchical models provide control for individual clustering within facilities and variation of RAVs between facilities they do not account for the proportion of high or low risk patients treated by a service provider (Ash, Shwartz, & Pekoz, 2003).

The choice of risk adjusters, the variety of techniques to perform risk adjustment, and the factors that may influence risk adjustment indicate the complexity of applying risk adjustment to quality measurement. As the next section will describe, mental health services present further challenges to quality measurement highlighting the need for risk adjustment.

1.7 Challenges to Quality Measurement for Mental Health Care

Non-acute health sectors such as mental health present specific challenges to quality measurement that are different from medical sectors such as acute care. In mental health, many problems or actions beyond clinical intervention may influence outcomes and the outcomes,

themselves, may not be as tangible as in acute medical settings (Fauman, 1989). The variety of approaches to treatment and unique characteristics of mental health conditions make quality measurement difficult. For instance, the symptoms of some mental illnesses such as psychosis or substance use disorders may include a lack of insight or motivation for treatment (Goldberg, Green-Paden, Lehman, Anthony, & Gold, 2001), thus influencing the likelihood of improvement regardless of treatments provided. The interaction between different treatment processes may also influence patient outcomes. For example, two patients with similar severities of depression may respond differently to the same antidepressant medication. For one patient, the response could be mediated by the presence of individual or group therapy while the other may experience improvement based solely on the medication (Antonuccio, Danton, & DeNelsky, 1995). The severity, chronicity, and diversity of many mental health conditions coupled with the variety of treatment options suggests that quality measurement will need to consider a diverse array of indicators to reflect the multifaceted nature of mental health care.

Other factors beyond illness characteristics, treatment options, and adherence can affect patient outcomes and may be directly, indirectly, or not related to treatment processes. Mental health admission criteria and diagnoses are rarely linked to specific treatments, and treatments are often less precise and less predictive of outcomes than in other health sectors (Lin, Degendorger, Durbin, Prendergast, & Goering, 2001). Instead, intangible qualities such as the therapeutic alliance between clinicians and patients, the insight and empathy of mental health professionals, environmental factors such as living conditions and income, personal factors such as family relationship, the chronicity of mental illness, the presence of concurrent physical illness, and individual choice for treatment all affect treatment outcomes (Health Canada, 1994).

Evaluators of quality also need to be sensitive to the challenges of treating a diverse array of complex conditions, often in one setting. Mental health services have to be equipped to treat various conditions that are often unrelated on a case by case basis. For example, among Canadian general and psychiatric hospitals in 2005/2006, 51% of hospital separations for organic disorders were for persons

aged 65 years or more while 52% of hospital separations for schizophrenia were for persons less than 45 years of age and 50% for mood disorders were for persons less than 24 years of age (CIHI, 2008). This example reflects the diversity of mental health at two levels. First, inpatient services treat patients at a wide range of ages and need to be able to accommodate the variety of peripheral conditions that might be associated with persons of different age (e.g., mobility. concurrent conditions, support needs, etc.). Second, the range of conditions, themselves, are quite different (e.g., dementia vs. mood disorders) and require specific treatments and expertise. This highlights the need for QIs that reflect the diversity of conditions treated and a mechanism for insuring indicators are compared in ways that do not penalize service providers for treating difficult and diverse mental health conditions.

The structural complexity of how mental health services are delivered and governed also leads to difficulty for measuring quality. Inpatient and outpatient services are typically managed and administered differently and often use unique forms of health information management (Goering et al., 2000). It is also difficult to track individuals as they move between inpatient and outpatient settings because of differences in information gathering and infrastructure. This can create difficulty for quality measurement and quality improvement, particularly for identifying contextual factors of treatment settings that influence individual outcomes.

1.8 Quality Indicator and Risk Adjustment Applications for Mental Health Services

1.8.1 Quality Indicator Initiatives and Research in Mental Health Services

Quality indicator development has become a key initiative of international government, accreditation, and research organizations. Some of the largest initiatives have emerged from Australia, United Kingdom, the United States (U.S.) and Canada. The Australian National Mental Health Working Group (NMHWG) developed a set of 13 QIs linked to datasets for inpatient, residential, and outpatient services (NMHWG, 2005). Benchmarks were established for an expanded set of 25 structure, process, resource utilization, and outcome QIs between inpatient services in Australia (Meehan, Stedman, Neuendorf, Francisco, & Neilson, 2007). The United Kingdom National Health

Service (UK-NHS) developed a framework for quality measurement of mental health services using structural and process indicators to measure clinical, patient, and capacity/capability aspects of care. Common indicators among these organizations include clinical negligence, psychiatric readmissions, prevalence of suicide, and transition of care between inpatient and community treatment.

In the U.S., the American Psychiatric Association (APA), Mental Health Statistics Improvement Program (MHSIP, 1996), the National Association of State Mental Health Program Directors (NASMHPD, 1998), and the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO, 1999) have all suggested quality domains or indicators for measurement at a national level. Overall, 56 national and state organizations have been documented in the U.S. as developing or using quality indicators (Hermann, 2005).

The American Psychiatric Association (APA) established a task force to develop QIs for mental health (APA Task Force on Quality Indicators, 1999). Quality domains were defined in a framework for developing QIs for different sectors of mental health including child, adolescent, and adult services. The domains address four dimensions of quality: Access to effective and appropriate care, quality of care, perception of care, and outcome of care. Specific QIs for each domain were not provided.

The Mental Health Statistics Improvement Program (MHSIP) was originally established to determine and design methods of collecting and interpreting data to inform mental healthcare (MHSIP, 1996). Through mental health report cards the MHSIP makes recommendations for the type of information to collect, the method of collecting that information, the process of interpreting and reporting the information, and the utilization of its use in decision making. MHSIP released its first mental health report card in 1996 which centered on the consumer's satisfaction and needs (MHSIP 1996). In the latest mental health report card the MHSIP identified 52 QIs and prioritized them based on information from 982 persons representing different mental health stakeholder groups, the majority being advocates, consumers, family members, and providers (Ganju, Smith, Adams, et al., 2005). The indicators include domains of structure (e.g., availability of services), process (e.g., participation in

treatment planning), outcomes (e.g., improvement in functioning), social support (e.g., prevalence of low social relationships), cultural sensitivity (e.g, perception of cultural needs considered in treatment planning), and safety (e.g., rate of medication errors). They are divided into universal indicators as well as population specific (e.g., children) and setting specific (e.g., inpatient) domains. Risk adjustment was not addressed for any of the QIs.

The National Association of State Mental Health Program Directors (NASMHPD) has also established a set of 46 QIs for mental health services. The NASMHPD QIs follow the framework of the first MHSIP report (MHSIP, 1996) with an added domain of "Structure/Plan Management" (NASMHPD, 1998). A total of 46 indicators were proposed, with 32 based on existing data sources in the U.S. The NASMHPD indicators span the entire mental health system (i.e., inpatient, outpatient, children, adults, etc.). While most indicators are process oriented, a number also focus on outcomes. These include consumer perception of care, improvement in school or employment, improved functioning, symptom relief, consumer injuries, elopement, and involvement in criminal justice system.

The JCAHO QIs are mandated for use among all inpatient mental health treatment facilities. JCAHO recently released the specification manual for *Hospital Based Inpatient Psychiatric Services (HBIPS) Core Measure Set Version 2.0* which provides the definitions and data standards for the 7 QIs to be reported (JCAHO, 2008). They were developed with consultation and consideration from the NASMHPD, the APA, and other national mental health agencies. The 7 QIs address the use of assessment, issues related to patient safety, and continuity/transitional care processes. Similar to most other QI initiatives listed, the JCAHO indicators are based on abstracted data from hospital medical records and are all process based with no specifications for risk adjustment.

An inventory for QIs for mental health has been developed by the Centre for Quality Assessment and Improvement in Mental Health (www.CQAIMH.org). Beginning with the identification of 86 process QIs, the inventory has now grown to include over 200 process QIs for mental health (Hermann et al., 2000; Hermann, 2005). The majority of QIs assessed appropriateness, access, and continuity of care. The most common diagnostic groups targeted by the QIs were

schizophrenia and depression while the most common treatment modalities targeted were medication and psychosocial modalities. The web-based inventory lists indictors based on diagnosis, special population (e.g., older adults), level of research evidence, and clinical setting (e.g., inpatient facility). Sets of QIs from this inventory have also been designated for specific groups such as patients with schizophrenia (Hermann et al., 2002).

There is an emerging literature that explores the use of many QIs for benchmarking mental health treatment. Using published reports of prevalence, statistical benchmarks have been identified for 56 mental health process indicators (Hermann, Mattke, Somekh, et al., 2006). As well, mental health stakeholders (e.g., accreditors, public or prior payer, clinician, advocate, etc.) identified 28 QIs for benchmarking quality in the U.S (Hermann, Palmer, et al. 2004). The 28 prevalence-based process indicators (Appendix B: Table 1) measure access, assessment, coordination, safety, treatment process (guidelines), continuity, and prevention among a mix of diagnostic groups, age groups and treatment settings. Specific to inpatient mental health for adults, 12 QIs have been identified for international benchmarking (Hermann et al., 2006). These indicators (Appendix B: Table 2) assess quality domains such as treatment, coordination, and continuity of services. One outcome indicator, mortality for persons with severe psychiatric disorders, is also included. No risk adjustment was applied to these QIs. This could be problematic considering benchmarked QIs in Australia were influenced by patient case mix and facility service characteristics (Meehan et al., 2007). Therefore, consideration is needed for risk adjusting QIs for patient case mix prior to establishing benchmarks.

A number of studies have also described the development or validation of specific QIs. The use of hospital readmission rates as a QI has received varying levels of support and criticism (Rosenheck, Fontana & Stolar, 1999; Humphreys & Weingardt, 2000; Craig, Fennig, Tanenber-Kurant, & Bromet, 2000; Lyons, O'Mahoney, Miller et al., 1997). The measurement of readmission has been problematic due to the variety of definitions in the timeframe of readmission; timeframes have ranged from 14 days to 1 year post discharge (Craig et al., 2000; Hendryx, Moore, Leeper, Reynolds, & Davis, 2001; Hendryx, Russo, Stegner, et al., 2003; Humphreys & Weingardt, 2000; Lyons et al., 1997). Consensus

on a timeframe definition for readmission is essential for determining the point at which readmission represents a quality problem compared to an event related to the development of new, or chronicity of old, symptoms or conditions. Timing of readmission is also important for considering the quality of community mental health supports and services. If the readmission was 30 days post discharge, it is not reasonable to attribute this to the quality of community mental health service. However, a readmission 6 months post discharge brings into question the quality of community services offered.

The use of readmission as a QI for inpatient mental health services compared to its use for CMH is somewhat unclear. Several studies have examined the relationship between clinical symptoms or outcomes and readmission with conflicting results. For instance, several studies have found no differences in clinical outcomes among patients who were later readmitted and those who were new admissions for mental health and substance use treatment (Lyons et al., 1997; Humphreys & Weingardt, 2000). These results suggest the clinical outcomes achieved during inpatient care were not related to later readmission, indicating that readmission may not be related to the quality of inpatient treatment. On the other hand, the presence of psychotic symptoms at discharge has been found to be related to future readmissions (Hodgson, Lewis, & Boardman, 2001). These results suggest that readmission may be useful as a QI for inpatient and outpatient mental health services, but that it is a proxy for quality rather than a direct indication of quality. Readmission is not necessarily a sole outcome of poor inpatient care, but may reflect inappropriate inpatient treatment, improper or lack of referral for follow-up treatment, a lack of continuity of services, the quality of resource allocation, and level of service need with the mental health care system (Hodgson et al., 2001). Responsibility for readmission, therefore, would rest on the quality of care from both inpatient and community service settings.

There has been very little inquiry into the development of outcome QIs for mental health. A number of outcome measures have been identified as key outcomes for acute inpatient mental health services including readmission, improvement in symptoms and functioning, satisfaction, and suicide/self-injury (Gerlamo, 2004). Numerous assessments tools have also been developed for

assessing mental health outcomes including the Behaviour & Symptom Identification Scale (BASIS; Eisen, Wilconx, Leff, et al., 1999) and the Health of the Nations Outcome scale (Goldney, Fisher, Walmsley, et al., 1996). Standardized approaches to the measurement of these outcomes or use of available measures have not occurred at a system level. Many of these scales are specific to certain types of clinical symptoms and functioning and lack data to support the use of risk adjustment for comparing outcomes. For instance, the BASIS assessment has measures of depression/anxiety, social and role functioning, psychosis, and addictive behaviours but no information is available on other demographic, diagnostic, or concurrent symptoms that could influence outcomes.

In the initiatives discussed thus far there has been limited use of risk adjustment for measuring QIs. This limitation, and the limited use of outcome QIs, may be due to the lack of available or rich data useful for outcomes and risk adjustment, particularly at a system level. While process QIs are quite useful for evaluating the quality of the services available, they provide no indication of the impact of these services on patients. For outcome QIs to be used effectively to compare mental health facilities, proper accounting of such QIs and consideration for risk adjustment is needed. In the next section a review of risk adjustment applications used for mental health QIs will highlight several outcome QIs that have been used to evaluate mental health quality.

1.8.2 Applications and Research on Risk Adjustment in Mental Health Services

Research on risk adjustment for measuring quality of mental health services is less common than research on risk adjustment for mental health service utilization and costs. For instance, a review of research examining risk adjustment used in mental health services, only 15 risk adjustment models were identified for outcomes compared to 72 for service utilization and costs (Hermann, 2007). While costs and utilization are important to system functioning, properly measuring and comparing processes and outcomes may be essential for accountability and improvement in system delivery.

Most research on risk adjustment of QIs for mental health services has emerged in the last 7 to 10 years with most studies focused on quality of outpatient services. Hendryx, Dyck, & Srebnik (1999)

used linear regression to risk adjust functional status, quality of life, and satisfaction outcomes among adults receiving services from 6 CMH agencies. Risk adjusters included age, sex, race, diagnosis, substance abuse, as well as baseline measures of the three outcomes. Among all models, the baseline level of each outcome accounted for the most variance in predicting time 2 outcome. Agency rankings differed significantly between unadjusted and adjusted rankings. Using the same data, Hendryx and Teague (2001a) found that agency rankings also differed based on the composition of the risk adjustment model. Models that included administrative data (demographics and diagnoses) explained little variance (~6%) compared to models that included clinical data (~30%). Both of these studies were problematic because the baseline information was collected from clients who had already received services, thus violating the assumption that RAVs be unrelated to provider intervention. Selection bias may have also been a factor since clients were solicited for participation. Clients who chose to participate may have different characteristics from those who did not participate. While these limitations question the utility of the specific risk adjustment models in predicting these outcomes in subsequent evaluations, they do provide evidence that the RAV and RA method can have an impact on quality rankings.

Banks, Pandiani, and Bramley (2001) tested three methods of risk adjusting the rate of change in criminal justice involvement following community mental health services. Rather than regression, they used weighted stratification, a basic pre-post test, and a mixed procedure combining stratification weighting and pre-post evaluation. The mixed approach involved stratification on age, gender, and focus of treatment for each agency and measurement of pre-post change scores within each stratum. The change scores were weighted by the distribution of that stratum among all agencies. Using this approach they found a substantial amount of variation of criminal involvement rates following treatment between community agencies.

As discussed previously, rehospitalization may be a relevant QI for mental health (Craig, et al. 2000) and substance abuse treatment (Humphreys & Weingardt, 2000). Early attempts to risk adjust rehospitalization relied on simple pre-post comparisons controlling the current rate of rehospitalization

with the rate of rehospitalization from the previous year (Banks, Pandiani, Schacht, & Gauvin, 1999). Such methods are problematic as prior rehospitalization may be a result of a consistent quality problem and ignore patient factors that have been found to predict rehospitalization including severity of illness, concurrent substance use, and functioning (Hodgson, et al., 2001; Lyons, 1997; Hendryx, et al., 2003). Incorporating several of these risk factors, Hendryx, et al. (2001) found that regression based risk adjustment produced different results from stratification-weighted adjustment for 22% of outpatient agencies being compared on rates of rehospitalization. These results indicate that utilization based QIs such as rehospitalization are responsive to RA.

Risk adjustment has also been used to compare different types of mental health services. Greenberg & Rosenheck (2006) used hierarchical model adjustment to compare changes in Global Assessment of Functioning (GAF) between specialized and general outpatient mental health services. Using baseline GAF score, diagnosis, and time between assessments as RAVs they found greater improvement in GAF for outpatients receiving specialized services compared to general services. Several direct limitations to these findings include the GAF as an outcome measure and a lack of possibly more relevant RAVs such as mental health symptoms, adherence, and chronicity of illness. This study also echoes concerns that the balancing of RAVs among comparison groups is needed to make comparisons of quality (Shahian & Normand, 2008). However, recalling that within inpatient and outpatient treatment settings there may be different levels of specialization and acuity, the importance of accounting this variation through RA may influence the impact of selection bias on quality comparisons.

Few national or international QI initiatives for mental health have implemented RA for comparing QIs. Logistic regression was used to risk adjust readmission rates, seclusion, and restraint use across 240 hospitals using the National Association of State Mental Health Program Directors performance measurement system in the U.S.(Schacht & Hines, 2003). The RAVs were unique for each QI but included patient sociodemographic characteristics and several unit or facility characteristics such as bed capacity and unit security. The use of facility characteristics as RAVs is

problematic as these variables should be considered contextual variables explaining differences in quality of care and not adjustment variables.

Several applications of risk adjustment for inpatient mental health services in the U.S. have focused on state-specific evaluations. Risk factors such as symptom severity, concurrent substance abuse, and demographics were chosen for QIs measuring adolescent inpatient and outpatient mental health services in Arkansas (Phillips, Hargis, Kramer, et al., 2000). Linear regression adjustment revealed that adjusted rankings were only moderately different from unadjusted rankings for most providers, with larger differences among a small number of providers (Phillips, Kramer, Compton, et al., 2003). In Florida, regression based risk adjustment led to significant differences in rankings of 50% of inpatient mental health facilities assessed based on change in GAF scores (Dow, Boaz, & Thornton, 2001). While the unadjusted and adjusted GAF scores were strongly correlated ($r = 0.89$), the rankings for several facilities shifted by 6 to 8 ranks. These results emphasize that while risk adjustment may not have a large effect for all providers, it can still affect the rankings of select providers immensely.

While RA has had limited use for comparing QIs across mental health providers, RA has been used to compare types of and relationships between mental health QIs. Such studies did not compare performance but were directed at improving measures of quality. For instance, several studies have looked at the relationship between patient satisfaction and other QIs. The use of hierarchical regression to adjust patient sociodemographics as well as medical and psychiatric diagnoses identified that higher satisfaction scores were related to better administrative QI scores (Druss, Rosenheck, & Stolar, 1999). Multinomial regression identified a moderate relationship between technical quality of care (e.g., appropriate mediation use) and satisfaction after adjusting physical health, psychiatric illness severity, and sociodemographics (Edlund, Young, Kung, et al., 2003). Finally, hierarchical regression adjustment of patient and facility characteristics found a different impact of service changes over time on patient satisfaction compared to unadjusted results among mental health inpatients (Greenberg & Rosenheck, 2004).

Previous sections of this chapter have identified the potential for using propensity scores in risk adjusting QIs. While this technique has not been used to compare mental health services, it has been used to evaluate the impact of appropriate treatment on outcomes. For instance, adherence to recommended guidelines of antipsychotic medication use was found to improve health status, reduce the prevalence of side effects, and improve the perception of care among acutely ill schizophrenia patients (Dickey, Normand, Eisen, et al., 2006). In this study patient risk factors were balanced between those who received appropriate treatment and those who did not. In a similar study, Medicaid patients with schizophrenia not enrolled in managed care in Massachusetts were found to have an equal likelihood of receiving appropriate treatment based on standard guidelines compared to those in managed care, after balancing patient covariates between these groups using propensity scores (Dickey, Normand, Hermann, et al., 2003). The use of newer methods of risk adjustment in these studies indicates their potential for adjusting QIs for comparing quality across service providers.

1.8.3 Effectiveness Indicators for Mental Health Services

The relationship between outcome and length of stay is not extensively discussed in literature on QIs for mental health. The research introduced so far has only focused on outcomes based on change but has not incorporated the magnitude of change over time. Indicators related to length of stay have been developed for evaluating quality of rehabilitation care. Using the Functional Independence Measure (FIM) an efficiency QI has been developed that measures improvement in FIM scores over length of stay (LOS; Granger, Cotter, Hamilton, & Fiedler, 1993), Specifically, the change in FIM score between admission and discharge is divided by the length of stay (LOS; in days) of the admission with higher scores indicating greater change in functioning over less time. Lower scores, or negative scores, could indicate less improvement over longer periods of time or decline over time. Therefore, the efficiency FIM score promotes improvement in FIM scores over the shortest LOS. FIM efficiency has been used widely including research on stroke outcomes (Bates & Stineman, 2000), as a QI for

rehabilitation services (Uniform System for Medical Rehabilitation, 2008), and in investigations on facility characteristics that influence outcomes in rehabilitation (Woo, Chan, Sum, et al., 2008).

Measures similar to FIM efficiency may also be relevant for mental health services as differences in outcomes have been described in relation to length of stay. For instance, psychiatric inpatients with longer LOS have been found to have lower depressive symptoms at discharge compared to those with shorter LOS (Lieberman, Wiitala, Elliott, McCormick, & Goyette, 1998). Conceivably, better quality of care could be concluded for a facility that achieves a positive outcome (e.g., improvement in depression) in a shorter episode of care than a second facility after adjusting for risk factors for that outcome. However, the results from Lieberman et al. indicate that the shorter LOS is not, necessarily, better in terms of outcome. Rather than efficiency, a more appropriate outcome may be effectiveness of services. Effectiveness refers to the achievement of outcomes as a result of treatment process (Schinnar, Kamis-Gould, Delucia, & Rothbard 1990). Therefore, optimal effectiveness of mental health services could show high ratios of patient improvement over time in hospital. Rather than rewarding outcomes achieved in the shortest amount of time, effectiveness would reward greater improvements in the time in which the person was receiving services.

1.9 The Status of Mental Health Quality Measurement in Canada

A number of accreditation, organizational, government, and research agencies have initiated QI development activities in Canada. At the National level, Accreditation Canada (formerly the Canadian Council for Health Services Accreditation or CCHSA) has put quality measurement at the forefront of their accreditation process. Accreditation Canada developed the Achieving Improved Measurement (AIM) which uses quality measurement to guide the accreditation process (CCHSA, 2003). AIM includes dimensions of quality such as organizational responsiveness, system competency, and client/community focus. The AIM program provides a set of general organizational standards (e.g., facility environment) as well as health-sector-specific standards according to which organizations are quantitatively assessed. Within mental health, nine mental health standards encourage organizations to

be open to learning (e.g., through research and quality improvement), promote well-being in both their staff and clients (e.g., empowering clients, health promotion, address needs), and to be goal oriented (e.g., achieve positive outcomes). The AIM program has yet to adopt a set of QIs for mental health. Some QIs have been reviewed but lack a national definition, data source, and evidence of reliability and validity. As such, Accreditation Canada called for a shift to a focus on improved quality measurement through the development of more rigorous QIs that emphasize outcome and can be used in comparative reporting.

The Canadian Institute for Health Information (CIHI) has led a number of National quality reporting initiatives. One of CIHI's primary functions is as a National data repository charged with ensuring the standardized definition, coding, and collection of data related to the health of Canadians and the healthcare services they receive (www.cihi.com). CIHI provides comparative reports specific to each health sector where data is collected. In conjunction with Statistics Canada, CIHI held a national consensus conference to establish a set of population health indicators (CIHI, 1999; 2005). An indicator framework was developed to include measures of health status, determinants of health, health system performance, and community and health system characteristics. Within health system performance, eight domains were proposed: acceptability, accessibility, appropriateness, competence, continuity, effectiveness, efficiency, and safety (CIHI, 2005). CIHI has also led efforts to develop QIs specific to mental health and addictions services at a health system level (CIHI, 2001a/b). These include:

- Hospital separation rates;
- Percentage of all hospital separations for mental illness/addiction services;
- Total patient days per 100,000 population;
- Average length of stay;
- Percentage of total days stay for mental illness/addiction;
- Suicide rates.

Using CIHI's health system performance framework, McEwan and Goldner (2001) developed a resource toolkit of QIs for mental health services. Descriptions of 56 QIs for mental health were provided and meant to be applied across the mental health system using various levels of measurement (e.g., program, system, or client level) and utility (e.g., policy development, clinical processes). From these indicators, McEwan and Goldner then selected a set of 12 indicators for use at the health system level which measure 5 domains:

- Increased access to services and supports for persons with severe mental illness;
- Expand community based services to correct community/institutional balance;
- Ensure comprehensive range of services/supports;
- Include consumers and families as partners in service planning, delivery, evaluation;
- Improve the quality of life for persons with severe mental illness.

Although CIHI has developed reports for population health indicators and concepts from these reports have been applied to a QI toolkit for mental health services, few reports have been developed specifically measuring mental health service quality at a National level. Since mental health services are managed by provincial and regional governance, accountability and quality are typically managed provincially.

1.10 Accountability and Quality in Ontario

In Ontario, the Ministry of Health and Long Term Care (MoHLTC) has established accountability agreements with health care providers in all health sectors. The agreements are contracts between the MoHLTC and health service providers establishing the roles and responsibilities of providers and frameworks for evaluating their performance. In 2003 the *Mental Health Accountability Framework* was released as a guideline for monitoring the accountability, efficiency, and effectiveness of mental health services (Ontario MoHLTC, 2003). A central component of this framework was to conceptually describe a set a of performance domains and performance indicators for use in hospital report cards, service improvement initiatives, accreditation, operating plans, and other operational or

quality review activities. Eight performance domains and 69 performance indicators were established (Appendix C, Table 1). The majority of these indicators describe the process and structure of service delivery with the goal that clinical outcome based indicators be added in the future. The responsibility for establishing service accountability agreements with health service providers now rests with each LHIN. To date, little public reporting has been done using the QIs established in the Mental Health Accountability Framework.

In 2005, the Ontario Health Quality Council (OHQC; www.ohqc.ca) was established in response to *The Commitment to the Future of Medicare Act* established by the MoHLTC (http://www.e-laws.gov.on.ca). The OHQC was established to support continuous quality improvement and public reports on the quality of health care including access to publicly funded health services, health human resources, consumer and population health status, and health system outcomes. The OHQC produces annual reports on the health care system as well as other specific reports. The most recent annual report included QIs from almost all health sectors, although information on mental health was limited to a financial indicator of information technology spending (OHQC, 2009). In partnership with Ontario's Joint Policy and Planning Committee (JPPC) the OHQC reported a review of Provincial accountability agreements finding that more development is needed to establish meaningful QIs, including those for mental health, and that improvements are needed in performance targets and alignment of accountability with quality improvement and public reporting (OHQC & JPPC, 2008). The OHQC reports draw from information available from a variety of available data and reporting infrastructures including the Ontario Hospital Reports and CIHI.

Several provincial report cards for quality of mental health services have been developed by the Hospital Report Research Collaborative (HRRC; www.hospitalreport.ca). The HRRC is funded by the Ontario Hospital Association (www.oha.ca) and MoHLTC to develop methodology and balanced scorecards for measuring quality of healthcare in Ontario. In 2001, the HRRC published a feasibility study for applying a balanced scorecard to mental health hospitals (Lin et al., 2001). The framework outlined 40 QIs to evaluate system integration and change, clinical utilization and outcomes,

satisfaction, and financial performance at a health system level. While no strategy for risk adjustment was developed, three domains of risk adjusters were suggested: sociodemographics, special population mix, and patient severity. A follow-up report in 2004 evaluating progress of mental health system reform measured 24 additional QIs (Lin, Durbin, Koegl, Murray, Tucker, Daniel, et al., 2004). The report identified needs for further follow-up in several quality domains including appropriate service use, integration of post-discharge care, evidence based practice, and client-centered care. While the 2004 report measured whether or not hospitals collected outcome measures from clinical and patient perspectives, no indicators were included actually reporting on clinical outcomes. The 2004 report made specific recommendations for measuring clinical outcome QIs using risk adjustment to enhance future comparisons between hospitals and regional groups. A third report on mental health was released for 2007 adding QIs on patient satisfaction and expanding the analyses to comparing QIs at the LHIN level (Lin, Durbin, Zaslavaska, et al, 2008). No new QIs were added measuring clinical outcome and risk adjusted QIs were not reported. The 2004 report recommended the addition of clinical outcomes based on the availability of clinical assessment data among all hospitals in Ontario. The next section describes this data and the opportunity for its use to develop QIs for mental health.

1.11 The Applications and Implementation of the RAI-MH in Ontario

The Resident Assessment Instrument for Mental Health (RAI-MH) is a comprehensive assessment system that includes over 300 items on psychiatric inpatients' sociodemographic, health, service utilization, and functional characteristics, and includes summary scales and resource utilization measures (Hirdes, Marhaba, Smith, et al., 2000; Hirdes, Smith, Rabinowitz, 2002). A copy of the RAI-MH is available in Appendix D. The RAI-MH is part of a suite of instruments developed by interRAI, a collaboration between researchers and clinicians from over 30 countries devoted to improving health care for vulnerable populations. The goal of InterRAI is "to promote evidence-based clinical practice and policy decisions through the collection and interpretation of high quality data about the characteristics and outcomes of persons served across a variety of health and social services settings"

(www.interrai.org). To achieve this goal a suite of assessment systems has been developed for a variety of health sectors and implemented internationally (Hirdes, Ljunggren, Morris et al., 2008). These include instruments such as the RAI 2.0 for long term care facilities (Morris et al., 1997; Hawes, Morris, Phillips, Fries, Murphy, & Mor, 1997), the interRAI-Home Care (Morris, Fries, Steel, et al., 1997), the interRAI-Acute Care and interRAI-Post Acute Care (Gray, Bernabei, Berg, et al., 2008), the Palliative Care assessment system or interRAI-PC (Steel, Ljunggren, Topinkova, et al., 2003), and the interRAI-Intellectual Disability for adults with intellectual disability (Martin, Hirdes, Fries, & Smith, 2007). An instrument similar to the RAI-MH has also been developed for community mental health, the InterRAI-Community Mental Health (InterRAI-CMH) and a shorter instrument has been developed for assessment and risk appraisal for emergency psychiatric settings, the Emergency Screener for Psychiatry (InterRAI-ESP). Both the interRAI-CMH and interRAI-ESP have received extensive pilot testing and psychometric evaluation but have yet to be provincially mandated. More information on the RAI-MH is available in the Methodology section of this dissertation.

1.11.1 Care Planning Applications

All interRAI instruments have a common approach and applications for their use. These include applications for guiding care planning, embedded summary scales of symptoms, functioning, and risk, algorithms designed to measure resource utilization, and a set of specific QIs (Hirdes, Fries, Morris, et al., 1999). The care planning applications of interRAI assessments are referred to as Clinical Assessment Protocols (CAPS). The CAPs are designed to assist clinical teams in identifying key issues or opportunities for improvement that can be used to organize and prioritize services with the person (www.interrai.org). Using information gathered from the assessment, certain combinations of item responses trigger the CAP. The RAI-MH includes a set of 32 mental health assessment protocols (MHAPs) triggering a wide range of issues including difficulty with social functioning, pain, financial or medication management, vocational functioning, and substance-use (Martin, Hirdes, Morris, et al., 2009). Currently, research is underway to refine the MHAPs (which will change to CAPs) to improve

the sensitivity and specificity of the triggered issues. The list of current MHAPs and new CAPs can be found in Appendix E.

1.11.2 Summary Scale and Outcome Applications

All interRAI assessments also include embedded summary scales for capturing the severity and outcomes of various symptoms, functioning, and risks. A number of embedded scales are common across most interRAI assessments including scales for depressive symptoms (Burrows, Morris, Simon, et al., 2000), cognitive performance (Morris, Fries, Mehr, & Hawes, 1994), activities of daily living (Morris, Fries, & Morris, 1999), health instability (Hirdes, Frijters, & Teare, 2003), and pain (Fries, Simon, & Morris, 2001). Other scales are specialized for certain instruments such as the index of social engagement for the RAI 2.0 (Mor, Branco, Fleishman, et al., 1995) or the Aggressive Behaviour Scale for interRAI-LTC and RAI-MH (Perlman & Hirdes, 2008). For the RAI-MH, 13 embedded scales and 3 risk algorithms are available including scales on depressive and positive symptoms, aggressive behaviour, ADL and instrumental ADL (IADL) functioning, cognitive performance, pain, potential problems with substance use, and risks of harm to self, others, and inability to care for self. More information about the scales embedded in the RAI-MH is available in the methods section.

1.11.4 Resource Utilization and Funding Applications

Information collected on a number of interRAI assessment systems can also be used to describe case mix and resource utilization. Extensive development and application of the Resource Utilization Groups (RUG-III) based on the RAI 2.0 has occurred in the U.S., Canada, and internationally (Fries, Schneider, Foley et al., 1994; Hirdes, Botz, Kozak, & Jepp, 1996; Ikegami, Fries, Takagi et al, 1994; Carrillo, Garcia-Altes, Peiro et al., 1996; Bjorkgren, Hakkinen, Finne-Soveri, et al., 1999; Topinková, Neuwirth , Mellanová, et al., 2000). The RUG III system uses clinical characteristics of the person to account for variable costs of care and support the allocation of health care resources.

The RAI-MH also includes a measure of case mix and resource utilization called the System for Classification of Inpatient Psychiatry (SCIPP; Hirdes, Fries, Botz, Ensley, Marhaba, & Perez,

2003). The SCIPP was developed based on a staff-time measurement study of 34 inpatient mental health hospitals/units from 3 Provinces. Clinical staff completed an RAI-MH and recorded the amount of direct and in-direct time spent on each patient's care. The SCIPP is an algorithm that divides patients into 47 groups based on clinical diagnosis as well as different patient characteristics (see Appendix F). The SCIPP uses a hierarchical grouping methodology where ordered diagnostic groups are assigned based on the presence of given psychiatric diagnoses. For instance, a person with a mood disorder and a psychotic disorder such as schizophrenia would be classified in the schizophrenia diagnostic group. Individuals are then further divided within diagnostic groups based on characteristics such as the presence of behaviours, self harm, or hallucinations. This grouping methodology accounted for 26% of variable costs. Each group in the SCIPP is assigned a case mix index (CMI) score ranging from 0.26 to 2.17. A score below 1.0 indicates the patient is less resource intensive than the average inpatient while a score above 1.0 indicates the patient is more resource intensive. For instance, the most resource intensive group includes patients with schizophrenia, a length of stay less than 3 days, and a behaviour disturbance. Their CMI is 2.17 which indicates they are 117% more resource intensive than the average patient. The SCIPP-CMI is included in the inpatient mental health funding formula used by the Ontario MoHLTC (JPPC Technical Working Group, 2008).

1.11.4 Quality Measurement Applications of interRAI Assessment Systems

Most interRAI assessment systems include applications for quality measurement. The most extensive work on QIs has been completed for long term care based on the RAI 2.0 (e.g., Zimmermann, Karon, Arling, et. al, 1995; Mor, Angelelli, Jones, et al., 2003; Rantz, Popejoy, Mehr, et al., 1997; Phillips, Zimmerman, Bernabei, & Jonsson, 1997; Jones, Hirdes, Poss et al., in press). Extensive work has also been done to develop QIs for home care based on the RAI-HC (Hirdes et al., 2004; Dalby et al., 2005) and interRAI-PAC (Fries, Morris, Aliaga, & Jones, 2003). The interRAI series of QIs for each sector follow a similar approach to quality measurement. InterRAI QIs focus on domains related to a range of issues and conditions important to the person's quality of life and the

appropriate delivery of care. For instance, home care QIs based on the RAI-HC include 15 prevalence indicators such as inadequate meals, falls, social isolation, and abuse or neglect as well as 5 measures of incidence/failure to improve in symptoms such as communication and skin ulcers (Hirdes et al., 2004). The later QIs measuring incidence/failure to improve, called double barreled QIs, are a unique form of outcome QI that have been applied to home care, post-acute, and LTC. Double barreled QIs combine the incidence of symptoms and failure to improve in symptoms that should improve as equally adverse events into a single QI.

In many settings where interRAI QIs are applied consideration for risk adjustment is needed. For instance, in LTC and home care service recipients mainly consist of older adults. Outcomes for certain conditions among older adults such as dementia may include maintenance of functioning and prevention of decline rather than improvement. Therefore, for doubled barreled QIs, substantial consideration for risk adjustment has been applied to adjust for conditions such as dementia where maintenance of functioning is a positive rather than adverse outcome. Several applications of risk adjustment have been applied to LTC QIs using stratification (Zimmerman et al., 1995), logistic regression (Berg et al., 2001; Morris et al., 2002), and hierarchical modeling (Arling, Kane, Lewis, & Mueller, 2005). Recently, a new method for adjusting LTC QIs has been developed to include weighted stratification and regression based adjustment (Jones et al., in press). For each LTC QI, a specific risk variable was selected and stratified into low, medium and high risk categories. Within each stratum, QIs are calculated using logistic regression adjusting for other covariates. Stratum QI scores are then weighted and combined producing a single QI score for each LTC facility. For many home care QIs (HCQIs), risk adjustment was performed with regression modeling using generalized estimating equations (GEE) to evaluate risk adjustment variables (Hirdes et al., 2004) and to compare HCQIs across regions in Canada (Dalby et al., 2005). Using several outcome based HQCIs, logistic regression was used to identify specific risk adjusters for post acute QIs that differ from the general home care population (Fries et al., 2003). Due to the comprehensive inventories of QIs based on

interRAI assessment systems and the complexity of the populations where these systems are applied, risk adjustment is an essential component for quality measurement.

1.11.5 Opportunity for Quality Measurement Applications for the RAI-MH

Conceptual initiatives have taken place to develop QIs based on the RAI-MH. In 1999, the Ontario JPPC developed a set of 35 QIs for mental health (JPPC QIs) based on Version 1 or the RAI-MH. The JPPC QIs (see Appendix G for list) are organized into domains for behaviour and emotional patterns, cognitive patterns, nutrition/eating, physical functioning, clinical management, resource use, sexual violence, and accidents and include a mix of prevalence, remission or incidence based process and outcome QIs. The JPPC QIs use combinations of items or sub-scales within the RAI-MH to highlight potential problems with the quality of care provided by inpatient services. They were originally created over a series of consensus groups with clinicians, quality experts, and researchers to, first, identify the domains of quality of care of importance in psychiatry, and second, to identify the items in the RAI-MH version 1.0 that could be used to point to those problem areas.

At the time of the development there were limited data to empirically test the JPPC QIs. Of the data that were available, none were longitudinal. As such, incidence based indicators could not be evaluated. The majority of JPPC QIs were developed based on clinical relevance and their potential to measure outcomes upon the availability of longitudinal data. Evaluations are needed to examine the empirical relevance of QIs based on the RAI-MH for measuring outcomes across inpatient care. Since the inception of the JPPC QIs, the RAI-MH version 1 has been revised to version 2, with a number of items being added, revised, or removed. Therefore, a number of JPPC QIs may no longer be measurable. Finally, the JPPC QIs are not conceptually consistent with QIs based on other interRAI assessment systems that include double barreled QIs for adverse events and risk adjustment.

Revision of the JPPC QIs is needed to produce a set of mental healthcare quality indicators (MHQIs) based on the RAI-MH version 2. The next section will describe the provincial implementation of the RAI-MH in Ontario and the opportunity to use data from this implementation to

revise a set of MHQIs based on the RAI-MH, evaluate applications for risk adjustment, and apply the MHQIs and risk adjustment to comparisons of quality among Ontario hospitals and LHINs with inpatient mental health beds.

1.11.6 Implementation of the RAI-MH in Ontario

The Ontario *Mental Health Accountability Framework* recommended the implementation of the RAI-MH to improve the availability of clinically relevant data for measuring quality. In 2005, the RAI-MH was mandated for use among all persons receiving treatment in a designated mental health bed in Ontario and is completed at least twice per admission lasting longer than three days. Therefore, detailed longitudinal clinical information is now available on all persons receiving inpatient mental health services in Ontario. Specific details describing the completion of the RAI-MH can be found in the methods section.

The provincial implementation was managed by CIHI who established the Ontario Mental Health Reporting System (OMHRS) based on RAI-MH data. The OMHRS team works with representatives from all hospitals with inpatient mental health beds to provide training on the completion of the RAI-MH and the utility of the RAI-MH information. The OMHRS also includes quarterly reporting of facility level summaries of data quality, patient characteristics, outcomes, and quality of acute and specialized inpatient services (www.cihi.ca/omhrs). Specifically, the reports include results for MHAPs, change in selected summary scales from admission to discharge, and several JPPC QIs. The JPPC QIs included are the prevalence of rehospitalization, prevalence of physical restraint and acute control medication use, prevalence of pain without pain management, prevalence of signs of substance use without therapy, and prevalence of self-injury. Results are stratified by acute, long-stay, geriatric, and forensic status at the facility, peer, and provincial levels.

Measuring clinical outcomes and applying risk adjustment across all treatment facilities in a given mental health system has been a challenge in international quality measurement initiatives (Hermann et al., 2006). In Ontario, the release of the *Mental Health Accountability Framework* to

guide structures and processes of care and the provincial implementation of the RAI-MH as a clinical information assessment system for inpatient mental health services have created new opportunities for developing and implementing quality measurement strategies that include clinical outcomes and adjustment for patient characteristics consistently across all inpatient service providers. With the infrastructure in place for the standardized collection of clinical data and public reporting of quality, both from OMHRS, the Hospital Report series, and the OHQC there is an opportunity to develop a system for measuring quality of mental health services.

2.0 PURPOSE OF DISSERTATION

The purpose of this dissertation is to develop a set of QIs for evaluating the quality of inpatient mental health services using the Resident Assessment Instrument-Mental Health. The development of RAI-MH mental healthcare quality indicators (MHQIs) will be completed in four phases that will refine MHQIs based on the JPPC QIs, assign risk adjusters to each MHQI, develop indicators of the effectiveness of inpatient mental health services, and apply the MHQIs and effectiveness indicators to comparisons of inpatient mental health service quality between LHINs. The four phases are as follows:

2.1 Refinement of MHQIs based on the RAI-MH

The first phase will involve the refinement of a set of feasible and meaningful MHQIs based on the RAI-MH. To develop a set of MHQIs the following questions will be explored:

- Are there JPPC QIs that should be maintained, modified, or deleted based on version 2 of the RAI-MH now in use in Ontario?
- Is information available on the RAI-MH that could be used to develop new MHQIs?
- Are the new MHQIs empirically meaningful across Ontario inpatient mental health hospitals?

2.2 Evaluation of Risk Adjustment for the RAI-MH MHQIs

The second phase in the development of a set of MHQI based on the RAI-MH is to evaluate the use of risk adjusters to use the MHQIs for comparing quality. To evaluate risk adjustment of MHQIs a list of potential RAVs based on the RAI-MH will be evaluated with the following questions:

- Is the prevalence of potential RAVs different among inpatient hospitals in Ontario?
- Are potential RAVs meaningfully related to MHQIs?
- Does risk adjustment of MHQIs using meaningfully related RAVs have an impact on comparisons of quality among inpatient hospitals in Ontario?

2.3 Development of Effectiveness Indicators based on the RAI-MH

The third phase in the development of MHQIs based on the RAI-MH is to develop indicators that identify outcomes in relation to time. Using the FIM efficiency measures described in section 1.8.3

as a guide, this phase will explore the creation of effectiveness quality indicators (EQIs) and their utility for comparing inpatient MH hospitals in Ontario. Specifically, this phase will seek to answer the following questions:

- How can EQIs be created for MHQIs based on the RAI-MH?
- Should EQIs include restriction of the denominator?
- Are there differences between hospitals on EQI scores?
- Is there a relationship between baseline scores in measures used to calculate EQIs and MHQI scores?
- Does risk adjustment of EQIs have an impact on comparisons of quality among inpatient MH hospitals in Ontario?

2.4 Application of the MHQIs and EQIs for comparing Ontario LHINs

The final phase of this dissertation will focus on the application of the MHQIs, risk adjustment and effectiveness indicators to the comparison of quality of inpatient services among LHINs in Ontario. This phase will answer the following questions:

- Do LHINs differ in the means and distributions of patient characteristics and risk adjustment variables identified in phase 3?
- Are there differences in the prevalence and rates of MHQI scores among LHINs?
- Does risk adjustment have an impact on comparisons of MHQIs among LHINs in Ontario?
- Are there differences in unadjusted and adjusted EQI scores among LHINs?

3.0 METHODOLOGY

3.1 The Resident Assessment Instrument-Mental Health

Under the Provincial mandate in Ontario, the RAI-MH is completed at admission, upon a change in clinical status, every 90-days in hospital, and upon discharge for every person admitted to an inpatient mental health bed. The RAI-MH assessment is completed by nurses and care teams who are familiar with the patient and work on the unit where the patient is admitted. Information gathered to complete the RAI-MH comes from the most reliable source within the clinical judgment of the assessor. This can include clinical observation, chart reviews, referral information, informant (e.g., family member) information, and discussions with the patient. Staff who complete the RAI-MH receive training from CIHI on how to properly complete the assessment instrument and use its components. Each hospital that completes the RAI-MH also has designated RAI-MH coordinators who oversee the quality, collection, and submission of RAI-MH data to CIHI. The OMHRS team at CIHI monitors the quality of the data submitted, returns data that is not complete and requires resubmission, and has clinical experts on staff for ongoing support to hospitals.

This study is interested in the use of a variety of individual items, the SCIPP CMI, and summary scales drawn from the RAI-MH data to measure MHQIs and include as possible RAVs. The RAI-MH items have demonstrated strong reliability. For inter-rater reliability, the average agreement is 83% for all items (Hirdes et al., 2002). In more recent reliability research only 15% of items were found to have Kappas below 0.60 with only 3 items having Kappas below 0.40 (Hirdes et al., 2008). The average weighted Kappa for all items was 0.70, which Landis and Koch (1977) describe as "substantial agreement".

Definitions of summary scales representing different clinical and risk domains embedded in the RAI-MH can be found in Appendix H. A number of these scales are included on all interRAI instruments and have been psychometrically evaluated in different settings. The Depression Rating Scale (DRS) was originally developed using the RAI 2.0 among LTC residents in the U.S. (Burrows,

Morris, Simon, Hirdes, & Phillips, 2000). The DRS is the sum of 7 items measuring negative statements, persistent anger, unrealistic fears, repetitive health complaints, anxious complaints/concerns, sad/pained facial expressions, and crying/tearfulness. Scores of 3 or higher generally indicate possible depression.

The Cognitive Performance Scale (CPS) is based on items for short-term memory cognitive decision making, ability to make self understood, and eating. An algorithm is used to compute a categorical scale that describes cognitive performance as intact to very severely impaired. The CPS has been found to be strongly correlated (r = 0.86) with the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975), a gold standard for cognitive assessment, and has shown strong accuracy (area under the ROC curve = 0.96) indentifying persons with cognitive impairment (CPS; Morris, Fries, Mehr, & Hawes, 1994),

The Activities of Daily Living-Hierarchy (ADL-H) uses four items (personal hygiene, toileting, locomotion, and eating) to categorize stages at which ADLs can no longer be performed. By assigning lower scores to ADLs that typically decline sooner (e.g., toileting) and higher to late loss ADLs (e.g., eating) the items create a 7 point scale ranging from independent (0) to total dependence (6). The ADL-H scale has been found to be positively correlated with nursing time and sensitive to change over a 12 month period (Morris, Fries, & Morris, 1999). A second scale, the Instrumental Activities of Daily Living Capacity (IADL) scale sums seven items (transportation, managing medications and finances, ability to do housework, phone use, and shopping) to create scores ranging from 0 to 42.

The Aggressive Behaviour Scale (ABS) has been recently developed as a summary measure of the number and frequency of verbally abusive, physically abusive, socially inappropriate, and aggressive resistance of care behaviours (Perlman & Hirdes, 2008). The ABS was found to be positively related (r = 0.72) to the aggressive subscale of the Cohen Mansfield Agitation Inventory (Cohen-Mansfield, 1986) and to have strong internal consistency (Chronbach's Alpha = 0.80).

The Pain Scale in the RAI-MH is based on the frequency and intensity of pain. Higher scores

indicate a greater frequency and severity of pain with scores ranging from 0 (no pain) to 4 (severe daily pain). The Pain Scale has been found to be highly predictive of pain identified on the Visual Analogue Scale (Fries, Simon, Morris, Flodstrom, & Bookstein, 2001).

Several new scales also exist specific to the RAI-MH. Work on the validation of these scales is ongoing but there is preliminary evidence of their reliability and validity among MH settings. These include three risk based scales called the Severity of Self-harm scale (SOS), the Risk of Harm to Others scale (RHO), and the Self-Care Index (SCI). Each risk scale is derived by an algorithm combining symptoms and behaviours producing scores of 0 to 6, with higher scores indicating greater risk. These three scales have been found to be highly related to psychiatric nurses' ratings of level of risk of danger to self, danger to others, and inability to care for self. For instance, 80% of individuals who scored 6 out of 6 on the SOS scale (severe risk of self-harm) were deemed by clinical nursing staff to have a moderate to imminent risk of harming themselves while 50% had severe to very severe/imminent risk.

A new RAI-MH scale has also been developed to measure positive symptoms of psychosis called the Positive Symptoms Scale-Long (PSS). The PSS is the sum of the following 8 items that are scored from 0 (symptom not present) to 3 (symptom observed daily in the last 3 days): Hallucinations, command hallucinations, delusions, abnormal thought process, inflated self-worth, hyper-arousal, pressured speech, and abnormal/unusual movements. The PSS score ranges from 0 to 24 with higher scores indicating a greater number and frequency of positive symptoms. In pilot testing, the PSS was found to have good internal consistency (Cronbach's alpha = 0.73) and is strongly related to the Positive and Negative Syndrome Scale ($r=0.65$, $p < 0.0001$), a gold standard in the assessment of positive symptoms (Kay, Fiszbein, & Opler, 1987).

Finally, the Depression Severity Index (DSI) is a new measure of depressive symptoms based on the RAI-MH. The DSI scores range from 0 to 15 based on the sum of the following 5 items that are scored from (symptom not present) to 3 (symptom observed daily in the last 3 days): Sad, pained facial expressions, negative statements, self-deprecation guilt/shame, hopelessness. Higher scores indicate a greater number and frequency of depressive symptoms. In a pilot study of psychiatric inpatients

assessed with the RAI-MH the DSI was found to have good internal consistency (Cronbach's alpha = 0.81).

3.2 Data Samples

Two RAI-MH data sets will be used for phase 1 analyses to establish a list of MHQIs. For phases 2 to 4, only the second data set, OMHRS data, will be used.

Dataset 1: Pilot-data: The first data set, referred to as the Pilot-data, was collected as part of the *Innovations in Data, Evidence, and Applications (ideas)* for Mental Health project funded by the Ontario MoHLTC Primary Healthcare Transition Fund. The *ideas for* Mental Health project was designed to improve the clinical and quality applications of the RAI-MH. The Pilot-data consist of two assessments collected from 1,056 patients from 7 volunteer hospitals with inpatient mental health beds in Ontario between November, 2004 and April, 2005. Hospitals were recruited through a letter of invitation sent to the Executive Director, Head of Psychiatry, and Head of Research of facilities with mental health beds. Ten psychiatric units or hospitals across Northern, Eastern, South-western, and Central Ontario agreed to participate in this project. However, 3 hospitals failed to provide appropriate data and were removed from the data set.

The Pilot-data are based on consecutive admissions (or those scheduled for routine re-assessment) of adults aged 18 and over in a designated psychiatric bed in the participating hospitals. As participating hospitals were using the RAI-MH as part of regular clinical practice for the duration of the study (and beyond, due to mandate), patient consent was not required. Clinical staff at participating hospitals were asked to assess 100 patients (though smaller units were permitted to contribute fewer) at two consecutive points in time (e.g., admission and discharge) from a mixture of acute, long stay, forensic and geriatric psychiatry beds. Time 1 and time 2 assessments needed to be completed a minimum of 6 days apart so that observation periods did not overlap. Sites were reimbursed $60.00 for each patient on whom two RAI-MH assessments (e.g., admission and discharge) of acceptable quality (e.g., less than 10% missing data) were completed. This process received approval from the Office of

Research Ethics at the University of Waterloo and the research ethics board of participating facilities, where applicable.

Since the Pilot data were obtained before the RAI-MH was provincially mandated, staff from participating hospitals received a one-day training session by a clinical member of the research team (Registered Psychiatric Nurse) on the completion of the RAI-MH assessment. Training covered instruction on the proper completion and coding of all items on the RAI-MH using case studies and examples. All participating facilities were also given RAI-MH manuals as reference guides for completing the assessments. Sites completed the RAI-MH using either an electronic software solution or on a paper and pencil, scannable form. Among sites using software, the anonymized data were submitted electronically (in comma separated value form) to the research team. Sites that used the paper-based, scannable version of the RAI-MH submitted the forms to the University of Waterloo research team who scanned the data into a secure server at the University of Waterloo.

Dataset 2: Ontario Mental Health Reporting System (OMHRS) data: The second dataset, referred to as OMHRS-data, includes all RAI-MH assessments completed in Ontario from October 1, 2005 to March 31, 2007. The OMHRS data consists of 41,019 unique cases (i.e., patient episodes of care). Since the development of MHQIs is concerned with clinical outcomes that reflect change in a person's status, cases that did not include at least two assessments completed at least 6 days apart (each assessment observation period is 3 days) were deleted. After deletion, the OMHRS data included 30,046 cases. There is no mandate for how the RAI-MH is collected so some facilities use electronic collection based on software approved by CIHI while other facilities complete the assessments using paper and abstract the information into an electronic medical record for submission to CIHI. The data are submitted electronically by the hospitals to CIHI every 3 months. CIHI then removes all identifying information including the patient name, health card number, and postal code. Facility identifiers are removed and replaced by scrambled identifier so individual assessments can be grouped by facility but the actual facility names are not identified. CIHI creates an individual identifier for each person to be used for identifying and linking assessments (e.g., admission and discharge). CIHI also produces a LHIN

identifier so that assessments can be grouped by LHIN. Once the data are anonymized and unique identifiers are assigned CIHI sends a copy of the data to interRAI through its Canadian Collaborating Centre at the Homewood Research Institute and University of Waterloo.

3.3 Design and Analyses

Each phase used retrospective, secondary data analyses of observational data found in the Pilot and OMHRS data sets. Descriptive statistics for both datasets were generated to describe demographic (mean age, age distribution, gender, marital status, education), admission status (reason for admission, prior history of mental health service use, patient type: acute, long stay, forensic, or geriatric, and involuntary status), and diagnostic variables. Specific analytic procedures were performed for each phase of the research proposed for this dissertation and are outlined below. The study design, including the secondary analysis of the Pilot and OMHRS data and the analyses outlined below, was approved by the Office of Research Ethics at the University of Waterloo.

3.3.1 Phase 1: Refinement and Development of MHQIs based on the RAI-MH

The review of the JPPC QIs and the development of new MHQIs based on the RAI-MH considered the feasibility and meaningfulness of each QI. Feasibility was determined by the ability of the QI to be measured with version 2 of the RAI-MH, to be used as an indicator of clinical outcome whenever possible, and to be risk adjusted when deemed appropriate. Meaningfulness was determined by how representative the QI was among inpatient mental health facilities and by the relevance of the MHQI to mental health clinical and quality experts. With these considerations in mind, the revision of the MHQIs involved several stages: 1) Development of a potential list of candidate MHQIs, and 2) Quantitative and qualitative evaluation of MHQIs to determine their meaningfulness to inpatient mental health hospitals in Ontario.

3.3.1.1 Development of Candidate MHQIs

The analyses began with a review of the original JPPC QIs using version 2.0 of the RAI-MH. The overall goal was to preserve the quality domains (e.g., depressive symptoms) established by the

JPPC QI working group while revising each QI definition to reflect clinical outcome, where appropriate. The RAI-MH items or scales used to calculate the QI needed to be present on version 2 of the RAI-MH. If the items or scale needed for the JPPC QI were not included on the RAI-MH version 2, and alternative items or scales indicative of the QI domain could not be identified, the MHQI was deleted. For JPPC QIs where version 2 RAI-MH was available, the *QI measures* (i.e., the RAI-MH information used to calculate the QI) were reviewed for their ability to be measured using a scale rather than a single item. Single item based JPPC QIs were considered for modification if a scale-based measure could be used to assess a comparable domain. For instance, the PSS scale could be used to calculate changes in symptoms of psychosis rather than using the single hallucination item to measure changes in hallucinations. Scale based measures capture greater variability in the QI domain of interest, may be more sensitive to change, and less likely to game (Morris et al., 2003). Item based QIs were considered for events of specific clinical relevance, particularly those that can be linked to care planning activities. In addition to the PSS, the DSI which was developed based on data from mental health settings and may be a more appropriate measure of depressive symptoms rather than the DRS which was developed in long term care settings.

New MHQIs or MHQI domains were considered based on available information from the RAI-MH and their added value or clinical relevance. For example, it was possible to measure a MHQI for interpersonal conflict by summing together these four items: patient's persistent hostility to other patients/staff, friends/family, persistent frustration of staff when dealing with patient, and hostility of family/friends toward patients. Other new scales were also considered for modification of JPPC QIs.

The operationalizations of the candidate MHQIs (JPPC QIs that were retained for modifications and new MHQIs) were then reviewed with preference for definitions measuring change in clinical status rather than prevalence. To remain consistent with other interRAI QI initiatives in LTC (Jones et al., in press), home care (Hirdes et al., 2004), and post-acute care (Fries et al., 2003), clinical outcome MHQIs were defined in two ways: Rate of improvement and rate of incidence or failure to improve (i.e., double-barreled). Prevalence QIs were also considered for instances where the presence

of the QI domain at a given point of time could represent a quality problem (e.g., inpatient violence). Eligibility for inclusion in the denominator of MHQIs was first determined at the individual level and then aggregated by hospital to produce a MHQI score. For improvement QIs, only patients with potential to improve in a MHQI measure were eligible. For instance, only patients who expressed aggressive behaviour in the 3 days prior to the initial assessment (e.g., ABS > 0) would be eligible for a QI measuring improvement in aggressive behaviour. All patients were considered eligible for prevalence and double-barreled QIs because any of them could experience an event (e.g., physically restrained) or incidence of an event (e.g., develop aggressive behaviour). All those who were eligible were assigned a 1 if the quality event occurred (e.g., ABS score improved) and 0 if it did not. At the hospital level, the denominator for each MHQI was calculated as the total number of patients in a hospital who were eligible for the MHQI. The numerator was calculated by summing the total number of persons who experienced the quality event among those who were eligible to experience the MHQI.

3.3.1.2 Evaluation of Candidate QIs among OMHRS Hospitals

Quantitative evaluations of JPPC QIs and derivation of new MHQIs were performed using guidance from consultations of an expert group of mental health and quality indicator experts from interRAI. The consultation participants (Appendix I) convened twice to develop a consensus as to the potential MHQIs' statistical and clinical meaningfulness for hospitals and patient groups. The experts were recruited by telephone and/or email among members of interRAI who are familiar with psychiatric practice and QI development using instruments such as the RAI-MH. The goal was to recruit research and clinical experts with technical statistical experience deriving QIs, practical experience implementing QIs into health sectors, and clinical experience evaluating QIs in psychiatric practice. Two consultations took place as teleconferences as members were based in various locations across North America.

Quantitatively, a MQHI needed to consistently demonstrate that differences in scores do exist between hospitals in the Pilot and OMHRS data and that rates of the MHQIs are not so rare or common

as to limit the utility of the MQHI. To evaluate these criteria an initial review of the JPPC QI rates was carried out in the Pilot-dataset for the total sample and by hospital. The JPPC QIs were analyzed for variability between hospitals where the expectation was that rates or prevalence should show "healthy" variation among hospitals, a process used in other interRAI QI initiatives (Berg et al., 2002; Hirdes et al., 2004). Healthy variation was observed if rates generally differed among hospitals. Statistically significant differences were not used as the designation of a statistical difference (or lack of difference) between hospitals as a MHQI score does not necessarily imply the presence (or lack of) a quality problem. If no variability in MHQI scores was identified, the MHQI may not be sensitive in detecting differences in quality or there may be systematic quality issues among all hospitals. Finally, all rates/prevalence should be above 5% and below 95% for the majority of hospitals. It may be argued that a QI with rates below 5% may still be clinically meaningful. While this may be true, rates consistently below 5% may indicate that the QI is a measure of a sentinel event. A sentinel event is a rare, but often serious, event that may have drastic consequences for individual, other patients, and treatment staff (Berg et al., 2001). Using sentinel events as QIs to compare quality may not be meaningful for understanding differences in quality of care. A prime example of a sentinel event for mental health is a completed inpatient suicide. While such an event may represent a quality problem and be clinically significant, the infrequency and rarity of the event among all treatment facilities makes the use of this event as a QI unreliable. In most instances, facilities would already have rigorous procedures in place to monitor and evaluate such events. In fact, in the U.S., JCAHO has a set of policies and procedures in place for dealing with sentinel events (http://www.jointcommission.org/SentinelEvents).

The results of the initial review were then discussed in the first consultation with mental health and quality measurement experts. The experts were presented with a list of the QIs including definitions for numerator and denominator groups. The experts were asked to review the quantitative results in terms of the variation in rates among hospitals in the Pilot data and the appropriateness of rates. A general discussion of the clinical relevance of each QI was also included. Following these

discussions for each QI, a vote was taken as to whether the QI should be deleted, modified and reevaluated, or included as a candidate MHQIs.

Following the initial review of JPPC QIs and consultation with the expert group, candidate QIs were modified and new MHQIs were defined. The modification of JPPC QIs involved re-defining them to match the criteria established in the previous section (e.g., scale based, interRAI measurement standards, etc.). For instance, the prevalence of pain without pain as a focus of intervention was modified to measure improvement and incidence/failure to improve in pain. New MHQIs that were developed using the new scales and item combinations outlined previously were also defined based on previously specified criteria. The modified and new MHQIs were then re-calculated and re-evaluated using the empirical criteria in the Pilot Data. Also, to evaluate the consistency of the candidate MHQIs they were calculated using the OMHRS-data and further evaluated based on the variation of rates between the hospitals and the consistency to which rates fall above 5% and below 95%. Using the OMHRS data, the relationship between MHQIs that were modified using new scales (i.e., DSI and PSS) was examined by performing Spearman's Rho correlations between the MHQIs using the original items or scales and those using the new scales. Construct validity of the new MHQI could be supported if the MHQIs were significantly correlated at the 0.05 level of significance, and the correlation is of meaningful value (e.g., greater than 0.70).

After the candidate MHQIs were modified, re-evaluated in the Pilot data, and replicated in the OMHRS data a second consultation took place to review the modified MHQIs. This consultation followed the same procedure as the initial consultation where rates and variability were reviewed and discussed followed by a vote for a final list of MHQIs. This final list of MHQIs was then presented at the most recent meeting of the interRAI Network of Excellence in Mental Health (iNEMH) held in North Bay, Ontario in November, 2008. The iNEMH is an international group of researchers and clinicians who meet annually to review mental health research using interRAI assessments internationally. The iNEMH members include psychiatrists, psychiatric nurses, psychologists, and researchers from 9 countries with specialties in geriatric psychiatry, forensic psychiatry, epidemiology,

and health service research. The iNEMH were asked to provide general feedback and comments regarding the potential utility of the candidate MHQIs among international settings.

The means and distributions of the final set of MHQIs were then analyzed. The sample mean, standard deviation, median, and range were between the 1st and 3rd quartiles were calculated. Also, the percent coefficient of variation (COV) was calculated by dividing the sample standard deviation by the sample mean and multiplying by 100 for each MHQI score. The COV provides a standard index of the amount of variability in MHQI scores between hospitals.

3.3.2 Phase 2: Evaluation of Risk Adjustment for Candidate MHQIs.

3.3.2.1 Development of Candidate RAVs

The OMHRS-data were used to determine RAVs, distributions of RAVs among hospitals, and to evaluate the impact of risk adjustment on MHQI comparisons among hospitals. To perform these three stages of analyses, a candidate list of RAVs will be established based on information available from the RAI-MH. Candidate RAVs included age, gender, forensic status, psychiatric diagnoses (from section Q1 on the RAI-MH), all embedded scales on the RAI-MH, the SCIPP CMI, a SCIPP diagnosis variable, the presence of current violence, any history as the victim of physical, emotional, or sexual abuse, and presence of interpersonal conflict. Interpersonal conflict is the same variable described in section 3.3.1.1, paragraph 2.

The SCIPP diagnosis variable scores patients from 0 to 6 based on the hierarchical diagnostic grouping of patients, the first stage of the SCIPP calculation. A score of 0 indicates the patient is in the lowest diagnostic category (‘other’) while a score of 6 indicates the person is in the highest diagnostic category (schizophrenia and other psychoses). The ‘other’ diagnostic category could include anxiety disorders, somatoform disorders, or other diagnoses. The SCIPP diagnosis was included as it is also a measure of concurrent psychiatric diagnoses. Patients grouped in the highest category could have schizophrenia/other psychosis as well as a mood or substance-use disorder. As patients are grouped into categories below 6 they are less likely to have a concurrent disorder. Patients grouped in the lowest

category would have only 1 diagnosis.

The presence of current violence is defined as any violence in the 3 days prior to or included the assessment date based on four items: violent actions (item d2a), intimidation of others (item d2b), and violent ideation (d2c). Violent actions include acts with purposeful, malicious, or vicious intent (e.g., stabbing or choking). Violent intimidation includes threatening gestures or stance, shouting angrily, and explicit threats of violence. Violent ideation includes reports of premeditated thoughts, statements, or plans to commit violence. Each item is coded from 0 (never) to 4 (any instance in the last 3 days). Scores of 4 were re-coded to a 1 and all other scores were recoded to 0. The items were then summed with scores ranging from 0 to 4 with any score greater than 0 indicating current violence (i.e., violence in the last 3 days).

History of emotional, physical, or sexual assault/abuse was identified by the life events and history section on the RAI-MH (Section J1). Each item measures the most recent event from 0 (never) to 3 (in the last 7 days). The items were summed with scores above 0 indicating the presence of any assault/abuse.

To examine the potentially differential relationship different levels of RAVS may have on MHQIs a number of continuously scored RAVs were collapsed into categories. For age, categories were created for those under age 25, 25 to 44, 45 to 64, and 65 or more. All scales embedded in the RAI-MH (except for the SCIPP CMI) were collapsed into 3 categories: scores of 0, scores of 1 and 2, and scores of 3 or more. These categories, similar to dummy variables, were created for convenience of interpretation. Although most scales have different ranges, these categories make sense for identifying patients with no symptoms or functional problems, symptom or functional problems that are less than daily or mild, and symptoms or functional problems that are daily, include multiple symptoms, or are more severe. For instance, items used to measure the DSI are scored 3 if they are present daily and less than 3 if not daily. Therefore, a score of 3 or more on the DSI either means that an indicator item is present daily or that multiple items are present on a non-daily basis. Additional categories were created for the ABS and PSS for their evaluation as risk adjusters for acute control medication (ACM) and

physical restraint use. Scores of 6 or more on the ABS (at least 2 behaviours occurred daily or > than 2 non-daily behaviours) were included rather than scores of 3 or more in order to prevent the identification of restraint among less severe behaviours. Similarly, scores on the PSS greater than 13 (person had at least 4 symptoms that were present daily or > 4 non-daily symptoms) were included rather than scores greater than 3.

3.3.2.2 Selection of RAVs for Risk Adjustment of MHQIs

Quantitative evaluation of potential RAVs began with an examination of the distribution of potential RAVs among OMHRS hospitals. The mean and standard deviation of RAVs across hospitals and the median and range of RAV means among hospitals were reviewed. Kruskall Wallis Analysis of Variance tests were conducted to determine if differences in means across hospitals were statistically significant ($p<0.05$ as the criterion for significance). Kruskall Wallis tests were performed due to the non-normal distribution of RAVs among hospitals.

After the analysis of RAVs among hospitals, bivariate analyses between all candidate RAVs and each MHQI were conducted and used to construct multivariate models. Dependent variables for the bivariate and multivariate selection of RAVs included the incidence/failure to improve MHQIs and time 1 prevalence for ACM and Restraint QIs. The use of incidence/failure to improve to choose RAVs was used because the denominators included all patients and the rationale that variables found to increase the odds of incidence/failure to improve could be assumed to decrease the odds of improvement.

Candidate RAVs for each MHQI were selected if Spearman's Rho correlations were statistically significant at the 0.05 level and the correlation coefficient was greater than 0.10. A coefficient of 0.10 has been used in other analyses of risk adjusters (Morris et al., 2003) with the rationale that a potential RAV should account for at least 1% of variance in a bivariate analysis. Although this bivariate correlation threshold is low, the goal was to provide a liberal threshold for considering RAVs in multivariate analyses.

Candidate RAVs identified in bivariate analyses were then evaluated in a multivariate context. First, the OMHRS data were randomly split into two smaller data sets of equal size. The split was performed by assigning every second case in the unsorted OMHRS data to a test dataset with the remaining cases assigned to a replication dataset. To evaluate randomization, frequency analyses were performed to ensure that hospitals and selected patient characteristics were equally distributed in each dataset. The results of these analyses are in Appendix J.

Using the test data, the relationship between candidate RAVs and each MHQI were evaluated with regression models using Generalized Estimating Equations (GEE). Regression models using GEEs were chosen as they can control for correlated observations within hospitals and between observations collected at different points in time (Liang & Zeger, 1986). Regression models using GEE are considered marginal, or population averaged, models that account for clustering of observations (i.e., correlation of responses) within hospitals by including a hospital as a source of random error in each model (Ballinger, 2004; Hu, Goldberg, Hedeker, et al., 1998). The GENMOD procedure in Statistical Analysis Software (SAS) version 9.1 with the REPEATED statement was used to specify the GEE procedure. The scrambled facility number in the OMHRS-data was entered as the clustering variable using the Subject option and an exchangeable correlation structure was specified. The exchangeable correlation structure assumes the correlations are identical and is recommended unless drastic differences in the correlation matrices are expected (Agresti, 2007)

All variables significant at the bivariate level were entered into a GEE regression model for each MHQI. Different combinations of the RAVs were examined to rule out order-of-entry, deletion effects, and multicolinearity (Leigh, 1988; Hosmer & Lemeshow, 2000). For inclusion in a final risk adjustment model, variables needed to be statistically related to the MHQI (i.e., parameter estimates with p-values less than 0.05) with odds ratios greater than 1.3 or below 0.77. These thresholds were established to ensure the presence of the RAV had a reasonable influence over the MHQI of interest. While some variables may have statistically significant odds ratios, ORs less than 1.3 or greater than 0.77 could be said to have a relatively small impact on the likelihood of the QI. However, to protect

against over-adjustment of QIs, variables that were strongly related to the MHQI and were considered conceptually similar to the MHQI were excluded. Over adjustment refers to the use of spurious RAVs resulting in suppression of variance and a lack of differences in QI rates between comparison groups (Dalby et al., 2005).

A number of methods to evaluate goodness of fit for regression models using GEE have been proposed in simulation studies (e.g., Evans & Hosmer, 2004). However, these methods are not routinely implemented in SAS output and general consensus on a goodness of fit statistic for GEE models has not been established. Therefore, final risk adjustment models using GEEs identified in the test data were subjected to logistic regression. Using logistic regression, the discriminatory power of the model was evaluated using the *c* statistic (Hanley & McNeil, 1982) The c statistic measures how well the model discriminates those who experience an event (e.g., outcome) from those who do not (Cook, 2007). A *c* statistic of 0.5 indicates the model is no more discriminating than chance while a statistic of 1.0 indicates perfect discriminatory power. Ideally, RAV models should have a *c* statistic greater than 0.70. Once the final RAVs were identified from multivariate models in the test data, the models were applied to the replication data. Models that performed similarly to the test data were retained for risk adjustment.

3.3.2.3 Application of Regression Adjustment to MHQIs

The final set of analyses involved the application of indirect risk adjustment to MHQIs using logistic regression. Since the purpose was to evaluate the need and impact of risk adjustment for MHQIs, and not to evaluate the type of risk adjustment to use, logistic regression was chosen as it is the most common form of risk adjustment used when comparing quality of mental and other health sectors. Risk adjustment of the MHQIs followed a similar process used for adjusting interRAI QIs for home care (Dalby et al., 2005) and long term care (Morris et al., 2003). The first step in adjustment involved calculating a patient level expected MHQI score. To do this, the MHQI is the dependentvariable in a logistic regression equation that is calculated as follows:

$$1 / 1 + e^{-x}$$

where X is the combination of logistic regression coefficients and predicted variables from the following expression:

$$X_0 + X_{RAV1} * RAV_1 + \ldots..X_N * RAV_N$$

where X_0 is the logistic regression intercept, X_{RAV1} is the regression coefficient for the first RAV and RAV_1 is the patient RAV score. The expected values of patients were then pooled for each hospital to create the hospital level expected MHQI score. The grand MHQI mean was calculated by pooling the all patient observed scores in the OMHRS data.

The final risk adjusted score is calculated by standardizing the observed MHQI score using the expected score and the grand mean MQHI score across all hospitals as follows:

$$\text{Adjusted MHQI} = \frac{1}{1 + e^{(-1*(\text{Ln (obs/1-obs)} - \text{Ln (pred/1-pred)} * \text{Ln (grand/1-grand)}))}}$$

where *Ln* = natural logarithm, obs = the hospital observed MHQI score, pred = the hospital's predicted MHQI score, and grand = the observed MHQI score for the entire sample. The adjusted MHQIs based on this technique can be interpreted as the estimated MHQI score for a hospital if that hospital accepted patients with an average level of risk based on the population (Morris et al., 2003, Dalby et al., 2005). The population in these analyses was all cases from the OMHRS dataset. Estimations of expected individual scores were calculated with SAS using the PROC LOGISTIC procedure.

3.3.2.4 Evaluation of the Impact of Risk Adjustment

The impact of regression adjustment was evaluated in several ways. First, the distributions of unadjusted and adjusted MHQI scores among hospitals were analyzed using scatter plots. The scatter plots illustrated the relationship between the unadjusted and adjusted MHQIs (i.e., the more linear the scatter the less impact of risk adjustment), showed the distribution of unadjusted and adjusted scores among hospitals, and showed the degree of change in hospital scores with the distributions.

The second method for analyzing the impact of risk adjustment included differences in absolute and quintile rankings based on unadjusted and adjusted MHQIs. Hospitals were assigned

absolute ranks by sorting the hospital MHQI scores from highest to lowest thus producing 70 ranks. Ranks were also assigned by grouping hospital MHQI scores into quintiles resulting in 5 rank groups consisting of 13 or 14 hospitals per quintile. Changes in rankings could then be analyzed by examining the number of hospitals that changed, improved, or declined in absolute or quintile rank based on unadjusted versus adjusted MHQI scores. Change in quintile ranks represents movement of a hospital rank from one quintile to another. The percentage improvement in ranks excluded hospitals with the highest possible rank (1) while percentage decline in ranks excluded hospitals with the lowest rank (either 70 for absolute or 5 for quintile) from the denominator.

3.3.3 Phase 3: Development of Effectiveness Quality Indicators

Using the OMHRS data, Effectiveness Quality Indicator (EQI) score was created for the each of the following MHQI domains: depressive symptoms, aggressive, disruptive, and violent behaviours, symptoms of psychosis, cognitive performance, ADL functioning, financial management, medication management, pain, and interpersonal conflict. The measures used to calculate MHQIs for each of the 11 domains were used for EQIs. For instance, the DSI was used to calculate an EQI for depressive symptoms. As well as the MHQI measure, the time between assessments (assessment interval) was used as the index in which effectiveness was gauged. The assessment interval was measured in 7 day intervals by subtracting the time 1 assessment date from the time 2 assessment date and dividing by 7. The 7 day interval was chosen because using single days would yield very small effectiveness scores given the largest potential range of the numerator is 24 points (based on the PSS scale) and the average time between assessments was greater than 24 days.

Calculation of EQIs followed a similar process as the FIM Efficiency measure from rehabilitation medicine. Using each MHQI measure and assessment interval, a gain score was calculated by subtracting the time 2 MHQI measure score from the time 1 MHQI measure score. Since higher scores on all MHQI measures indicate a worse condition (e.g., more symptoms, worse functioning, etc.) the gain score is positive if the score improved, negative if the score declined, and 0

if the score did not change. The EQI gain score was then divided by the assessment interval. An example calculation of the Depression EQI is as follows:

$$\text{DSI Effectiveness} = \frac{\text{DSI Gain}}{\text{Assessment Interval}}$$

where:

DSI Gain = DSI score $_{\text{Time 1}}$ - DSI score $_{\text{Time 2}}$, and

Assessment Interval = (Time 2 Assessment date – Time 1 Assessment date) / 7

Two different types of inclusion criteria were examined for calculating EQIs: Inclusion of all scores on the MHQI measure or the exclusion of cases where the baseline MHQI measure score was 0. The second option would, essentially, include the same cases as the denominators for the improvement MHQIs. Two forms of inclusion criteria were evaluated to determine if EQIs should be based only on patients who are identified with symptoms at admission or all patients, regardless of symptoms.

The inclusion criteria were compared by calculating the mean with 95% confidence intervals for all EQI scores across patients and by examining the distribution of scores among hospitals. The distribution included the mean, standard deviation, COV, median, and range between the 5th and 95th percentiles. To examine the impact of excluding patients with MHQI measure scores of 0 at time 1, the incidence of MHQI measure scores was calculated among all patients with 0 scores at time 1. Also, Spearman correlations were calculated between hospital EQI scores based on the two inclusion criteria to determine if EQI scores for each inclusion type were related at the hospital level. The mean and distribution of EQIs based on the final method chosen was also examined by patient types (acute, long stay, forensic, or geriatric) to determine how effectiveness might differ by patient type.

Once the final EQI definition was determined risk adjustment was considered. Among studies that have used the FIM efficiency QI in rehabilitation, risk adjustment considered sociodemographics and diagnostic information as well as the baseline FIM score to adjust for improvement by chance (e.g., Woo et al., 2008). The RAVs chosen for the MHQIs were used for risk adjustment of the EQIs.

In addition, the baseline MHQI measure scores used to calculate the gain score were also evaluated for use as RAVs. Since the EQIs are based on the measurement of magnitude of change over time, rather than change itself, the likelihood that higher scores at time 1 improve by a greater magnitude than lower time 1 scores was considered. This issue is similar to the concept of the law of initial values where, pertaining to change scores, higher scores (i.e., more extreme) are deemed less likely to decline and more likely to improve compared to lower (i.e., less extreme) scores (Oken & Heath, 1963). To examine this potential, the coefficient of variation and distribution of time 1 hospital MHQI measure scores and Spearman Rho correlations between time 1 MHQI measure scores and EQI gain scores were examined. Also, Spearman's Rho correlations were calculated for the relationship between time 1 MHQI measure scores and EQI scores. If the time 1 MHQI measure scores were found to be unequally distributed among hospitals and significantly related to gain and overall EQI scores then they were included in risk adjustment for EQIs.

Risk adjustment of EQIs was performed with multivariate linear regression since the EQI scores are continuous. Risk adjustment using linear regression is less complex than logistic regression adjustment since the information is not logarithmically transformed. Otherwise, the process is very similar to linear regression. First, an expected EQI score was calculated for each patient using the following equation:

$$\hat{y} = a + b_1X_1 + b_2X_2 + \ldots\ldots\ldots + b_nX_n$$

Where $\hat{y}$ = the predicted EQI score, a = the estimated intercept, b_1 = the parameter estimate of the first RAV, X_1 = the patient score on the first RAV, and so on until the n^{th} RAV.

Second, following the same approach used to calculate a hospital's MHQI score, predicted EQIs for each patient were then pooled to create a hospital EQI score. Third, the predicted hospital EQI score was then combined with its observed EQI score and standardized on the population average EQI score (i.e., the average EQI score across all patients in the OMHRS data) to produce the hospital risk

adjusted EQI score. This was achieved by subtracting the predicted EQI score from the observed EQI score and multiplying by the population average EQI score.

Following adjustment, the impact of risk adjusting EQI scores was evaluated using the same process as the evaluation of risk adjustment of MHQI scores. The distributions of unadjusted and adjusted EQI scores were plotted, hospitals were absolute and quintile ranked, and patterns of change in ranks were calculated.

3.3.4 Phase 4: Comparison of MHQIs and EQIs among LHINs

The OHMRS-data were used to compare MHQI and EQI rates between LHINs. CIHI assigns a LHIN number to each patient row of data so that patient data can be grouped by LHINs. The prevalence of demographics (mean age, age distribution, gender, marital status, education), psychiatric service (patient type, prior history of involvement with mental health treatment, admission status), and common mental health diagnoses was calculated among all LHINs. All variables were dichotomized to equal 1 (yes) or 0 (no). The mean of each variable within each LHIN produced the prevalence of that variable in the LHIN and allowed the prevalence to be compared with other LHINs using Kruskall Wallis tests.

Before examining the unadjusted and adjusted MHQI and EQI results among LHINs, a comparison of the distribution of RAVs between LHINs was performed. This comparison used the same procedure as that used in section 3.2 when examining RAVs among hospitals. Risk adjustment was applied to all MHQI prevalence/rates and EQI scores among LHINs using the same procedures outlined in sections 3.2 and 3.3. The ranges in rates/prevalence of unadjusted MHQIs were compared to adjusted MHQIs to determine how risk adjustment influenced MHQI distribution. To examine the impact of risk adjustment on LHIN MHQI rates/prevalence, changes in rankings were used following analyses from previous phases. The median and range in unadjusted and adjusted EQI scores were calculated along with the number of LHINs that improved or declined in rankings following adjustment.

4.0 RESULTS

4.1 Refinement of the Mental Health Quality Indicators

4.1.1 Sample Characteristics

Table 2 shows selected patient characteristics for the Pilot and OMHRS data. In both datasets, patients had similar age and sex distributions with about 50% under the age of 44 and male. About a third of patients were married or had a partner. In the OMHRS data, 29% of patients had less than high school education compared with 35% in the Pilot data. Just over half of patients had a mood disorder while about a third had schizophrenia/other psychosis. A higher prevalence of dementia was found in the Pilot data and a slightly higher prevalence of substance use disorders was found in the OMHRS data.

Most patients had a history of involvement with mental health services in both sets of data. About 55% and 60% in the Pilot and OMHRS data, respectively, were in contact with community mental health services in the 30 days prior to the current admission. Over 50% of patients in both data sets had prior admissions in the prior two years and about 70% had any prior admissions in their lifetime. About 20% had six or more previous admissions to inpatient mental health services in their lifetime. For the current admission, the majority were acute patients in the OMHRS and Pilot data, although slightly more were considered long-term or geriatric among Pilot Data.

Table 2. Prevalence of selected patient characteristics in the pilot and OMHRS data.

Characteristics		Pilot Data % (n)	OMHRS Data % (n)
Age categories	Under 25	21% (220)	12% (3550)
	25-44	35% (365)	41% (12239)
	45-64	31% (332)	34% (10294)
	65+	13% (139)	13% (4103)
Gender (male)		51% (540)	50% (15023)
Marital Status	Never Married	38% (404)	47% (14180)
	Married	24% (253)	26% (7848)
	Partner	5% (56)	3% (1009)
	Widowed	6% (64)	6% (1737)
	Separated	12% (129)	8% (2459)
	Divorced	13% (141)	10% (2963)
Education	None	12% (129)	3% (986)
	8th grade or less	6% (62)	7% (2046)
	9 – 11	17% (181)	19% (1953)
	High school	21% (226)	24% (7298)
	Technical/trade	4% (43)	3% (925)
	Some college/univ.	15% (162)	17% (5185)
	Post Secondary	13% (136)	14% (2225)
	Unknown	11% (117)	12% (3633)
Involuntary admission status		28% (296)	24% (7385)
Last Contact with Community Mental Health	None in last year	45% (461)	40% (12213)
	31 days or more	18% (184)	19% (5613)
	30 days or less	37% (380)	41% (12370)
Psychiatric Admissions in Last 2 years	None	47% (484)	46% (13871)
	1-2	37% (379)	35% (10562)
	3 or more	16% (162)	19% (5763)
Lifetime Psychiatric Admission	None	28% (290)	29% (8873)
	1-3	38% (388)	35% (10645)
	4-5	15% (153)	14% (4306)
	6 or more	19% (193)	21% (6372)
Patient type:	Acute	72% (764)	79% (23881)
	Longer term	18% (191)	13% (3962)
	Geriatric	8% (80)	5% (1382)
	Forensic	2% (21)	3% (971)
Provisional psychiatric diagnoses:	Dementia	12% (130)	7% (2096)
	Mood disorder	54% (570)	51% (15412)
	Psychoses	39% (415)	37% (11073)
	Substance-related	21% (225)	26% (7893)

4.1.2 Prevalence and Rates of JPPC QIs in Pilot Data

The original JPPC QIs were calculated in the Pilot Data for each hospital and across diagnostic groups (Table 3). Derivation of four of the original MHQIs was no longer possible with RAI-MH version 2.0. Items were no longer available to calculate the prevalence of sexual violence (as perpetrator) and prevalence of fire setting. Medication data were not available to calculate prevalence of extrapyramidal symptoms and prevalence of psychotropic drug underuse. The prevalence of substance use without the offer of therapy and prevalence of smoking without the offer of therapy were excluded because the RAI-MH version 2.0 no longer distinguishes between therapy offered for alcohol/drug use and smoking therapy (i.e., both are included in 1 addictions item). The prevalence of rehospitalization to the same facility 30-days after discharge was also excluded as it could not account for rehospitalizations to other facilities in the region in the same timeframe.

Several MHQIs were deleted because their rates/prevalence were below 5% across all hospitals in the Pilot data. These include incidence measures of inpatient weight loss (mean= 2%) and weight gain (mean = 2%) as well as prevalence measures of dehydration (mean = 1%), inpatient suicide-attempt (mean = 2%), inpatient self-injury (mean = 2%), and inpatient falls (mean = 2%).

Table 3 Prevalence and incidence rates of JPPC QIs among pilot data hospitals.

	Hospital						
JPPC QIs	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Remission of depressive symptoms	80%	54%	70%	51%	91%	66%	60%
Incidence of depressive symptoms	2%	7%	12%	38%	12%	13%	5%
Remission rate of aggressive behaviour disturbance	69%	43%	68%	67%	80%	56%	30%
Incidence of aggressive behaviour disturbance	8%	4%	9%	16%	1%	5%	13%
Remission rate of disruptive behaviour disturbance	100%	67%	83%	55%	83%	91%	42%
Incidence of disruptive behaviour disturbance	8%	2%	7%	5%	3%	2%	9%
Prevalence of violent behaviour	15%	14%	0%	0%	0%	0%	8%
Remission rate of hallucinations	68%	53%	45%	68%	75%	58%	31%
Incidence of hallucinations	15%	2%	10%	4%	1%	1%	15%
Improvement in cognitive impairment	57%	36%	27%	29%	50%	57%	26%
Incidence of cognitive impairment	17%	5%	8%	17%	0%	2%	10%
Incidence of weight loss	0%	0%	0%	50%	2%	0%	4%
Incidence of weight gain	3%	3%	0%	0%	10%	4%	6%
Prevalence of dehydration	0%	0%	1%	5%	1%	1%	1%
Improvement of ADL functioning	78%	26%	59%	44%	0%	80%	27%
Incidence of ADL functioning	3%	1%	8%	7%	0%	3%	8%
Improvement in financial management IADL	55%	25%	13%	23%	33%	47%	22%
Incidence in financial management IADL	12%	4%	10%	26%	1%	0%	20%
Improvement in medication management IADL	57%	36%	14%	28%	70%	54%	18%
Incidence in medication management IADL	11%	6%	11%	14%	2%	2%	24%
Prevalence of rehospitalization	9%	3%	9%	10%	0%	8%	2%
Prevalence of unauthorized leaves of absence	57%	23%	29%	67%	60%	40%	36%
Prevalence of inpatient suicide attempts	3%	3%	0%	3%	0%	6%	0%
Prevalence of self-injury (non-suicidal)	1%	2%	1%	0%	0%	4%	0%
Prevalence of pain without analgesic use or pain management	25%	83%	40%	29%	44%	64%	67%
Prevalence of smoking/tobacco use without an offer of therapy	79%	87%	58%	72%	72%	98%	77%
Prevalence of signs of substance abuse without therapy	57%	78%	40%	57%	19%	75%	40%
Prevalence of chemical restraint use	8%	5%	11%	24%	4%	10%	20%
Prevalence of physical restraint use	0%	3%	5%	3%	0%	2%	11%
Prevalence of seclusion room use	18%	18%	11%	58%	8%	8%	27%
Prevalence of falls	3%	0%	3%	2%	1%	2%	1%

*** The following JPPC QIs could not be calculated due to missing items on the RAI-MH version 2.0: Prevalence of extrapyramidal symptoms, Prevalence of psychotropic medication underuse, prevalence of fire setting, prevalence of sexual violence (perpetrator)**

4.1.3 Expert Review of JPPC QIs and Suggestions for Refinement

A series of consultations was held with an international group of 4 mental health clinicians and quality experts who further evaluated the JPPC QI rates and provided suggestions for refinement of MHQIs. The expert panel was first presented with the results from table 3 and was asked to discuss the relevance of each QI as a descriptor of quality of inpatient mental health and appropriateness of the QIs. Relevance of the QI was based on whether the group deemed the domain meaningful to mental health (e.g., represents an important issue to patient recovery, negative consequences to the patient if the quality issue was not addressed, applicable to large proportion of patients) and feasibility of the QI (e.g., interpretability of QI, reasonable expectation that issue could be addressed through care).

Following the initial discussion, a series of alterations were made to the potential QIs and subsequent consultations were held (3 in total). The alterations included variations in inclusion or exclusion criteria and various coding options for the numerator. For instance, the prevalence of inpatient falls was evaluated as an incidence measure among patients who had not fallen in the 30 days prior to admission, but had a fall indicated on their follow-up RAI-MH assessment. The reviewers did not recommend this item as the rates across hospitals were too low.

After reviewing the original JPPC QI descriptions, rates among the pilot data, and various iterations of potential QIs the expert reviewers made several recommendations. First, the group agreed that QIs with rates below 5% should be deleted. Second, the group felt seclusion room use should be dropped as there were concerns that the large variability in seclusion room practices makes it difficult to determine when seclusion room use is appropriate. The group also commented that there could be ambiguity around the definition of a seclusion room (e.g., is locking a patient in their room considered seclusion room use?) and that variations in rates may be due to variations in the availability of seclusion rooms within each hospital.

Several modifications were also suggested by the expert group. First, the group agreed that all QIs should measure improvement instead of complete remission. Second, the group noted that QIs measuring an increase in symptoms/behaviours or a decrease in function are missing a proportion of patients who experience no change in symptoms/behaviours/functioning where an improvement might be expected. Therefore, the group felt that the addition of failure to improve to QIs measuring incidence or decline is important. Finally, QI's measuring financial and medication management should be defined as having no difficulty in functioning if the patient has full capacity or requires only set up help to complete either IADL (the JPPC version includes only full capacity).

The expert group felt that each MHQI domain should include two QIs. The first QI should measure rates of improvement among those who can improve and the second measures rates of incidence and failure to improve as 1 QI. Several prevalence MHQIs were also suggested to include physical restraint use and acute control medication use in the 3 days after admission and the 3 days prior to follow-up assessment. The final list of MHQI definitions approved by the expert reviewers can be found in table 4.

The MQHI domains were also reviewed at an annual meeting of the interRAI Network of Excellence in Mental Health (iNEMH). During the iNEMH meetings the MHQI domains were discussed in terms of their relevance at an international level, both for fit within different countries' models of mental health services and for making international comparisons of quality. The iNEMH membership agreed that the MHQIs should focus on outcomes that reflect positive and adverse events and that a diverse group of domains be included, beyond changes in symptoms or behaviours. Internationally, aspects of daily and social functioning were deemed important for insuring the overall recovery of the individual. Similar to the expert panel, risk adjustment was also a key issue among the iNEMH to enhance the utility of the MHQIs for making comparisons within and between mental health systems.

4.1.4 Definitions of Modified or New MHQIs

Using the expert group suggestions, all outcome MHQI domains in table 3 include an indicator of improvement as well as an indicator of incidence/failure to improve. Several new MHQIs were also developed. In addition to the prevalence of violence (including thoughts, actions, and intimidation) two outcome MHQIs were added for changes in violence between assessments. Two outcome MHQIs were defined to measure changes in pain rather than including pain as a prevalence measure. A new MHQI domain and two new MHQIs were developed to measure changes in interpersonal conflict. Four RAI-MH items measuring hostility toward friends/family (item o2c) and other patients/staff (o2d) as well as hostility toward patient by friends/family (02e) and staff frustration dealing with patient (02f) were summed to measure interpersonal conflict.

The depression MHQI domain was also revised to include scores on the DSI rather than the DRS for both MHQIs. The DSI was chosen because it was derived from a mental health sample rather than a long term care sample and has been found to have somewhat better psychometric properties among mental health patients than the DRS. To improve consistency with other MHQIs, the denominator for improvement on the DSI is based on scores of 1 or more at admission rather than 3 or more used on the DRS MHQI. In the OMHRS data, the rate of improvement on the DRS (83.4%) was significantly related to the rate of improvement on the DSI (77.1%), $r = 0.65$, $p < 0.0001$.

The hallucinations MHQIs were replaced with MHQIs for positive symptoms using the PSS. The denominator for improvement includes all scores greater than 0 on first assessment and all patients with non-missing values on the PSS are included in the denominator of for incidence/failure to improve. The rate of improvement in hallucinations (72.5%) was significantly related to the rate of improvement in the PSS (78.4%) in the OMHRS data, $r = 0.69$, $p < 0.0001$.

Table 4. RAI-MH Mental Health Quality Indicator (MHQI) definitions and coding rules.

Domain	Type of Indicator	Numerator	Denominator
Depression	Improvement	Patients whose initial Depression Severity Index (DSI) score is greater than their DSI score at next assessment.	All patients where their initial DSI score is greater than 0.
	Incidence/Failure to Improve	a) Patients whose initial DSI score is less than their DSI score at next assessment, OR b) Patients whose initial DSI score is greater than 1 and their DSI score at the next assessment equals their DSI score at initial assessment	All patients.
Aggressive Behaviour	Improvement	Patients whose initial Aggressive Behaviour Scale (ABS) score is greater than their ABS score on their next assessment	Patients whose initial ABS score is greater than 0.
	Incidence/Failure to Improve	a) Patients whose initial ABS score is greater than 0 and is less than their ABS score at next assessment. b) Patients whose initial ABS score is greater than 0 and their ABS score at the next assessment equals their ABS score at initial assessment	All cases
Disruptive Behaviour	Improvement	Patients whose initial Disruptive Behaviour Sum (DBS)[1] score is greater than their DBS score on their next assessment	All patients whose initial DBS score is greater than 0.
	Incidence/Failure to Improve	a) Patients whose initial DBS score is greater than 0 and less than their DBS score at next assessment. b) Patients whose initial DBS score is greater than 0 and their DBS score at the next assessment equals their DBS score at initial assessment	All cases
Violent Behaviour	Prevalence of Inpatient Violence	Patients whose Current Violence Sum (CVS)[2] is greater than 0 at their second assessment[3]	All cases at second assessment.
	Improvement	Patients whose initial CVS is greater than their CVS on their next assessment	All patients whose CVS is greater than 0 on their initial assessment.
	Incidence/Failure to Improve	a) Patients whose initial CVS score is greater than 0 and less than their CVS score at next assessment. b) Patients whose initial CVS score is greater than 0 and their CVS score at the next assessment equals their CVS score at initial assessment	All cases

Domain	Type of Indicator	Numerator	Denominator
Positive Symptoms	Improvement	Patients whose score on Positive Symptom Scale-Long (PSS) on the initial assessment is greater than their PSS score on their next assessment	All patients whose initial PSS score is greater than 0
	Incidence/Failure to Improve	a) Patients whose initial PSS score is less than their PSS score at next assessment. b) Patients whose initial PSS score is greater than 0 and their PSS score at the next assessment equals their PSS score at initial assessment	All cases
Cognition	Improvement	Patients whose initial Cognitive Performance Scale (CPS) score is greater than their CPS score at next assessment.	All patients whose CPS score is greater than 0 at initial assessment.
	Incidence/Failure to Improve	a) Patients whose initial CPS score is less than their CPS score at next assessment, OR b) Patients whose initial CPS score is greater than 0 and their CPS score at the next assessment equals their CPS score at initial assessment	All cases
ADL Functioning	Improvement	Patients whose initial Activities of Daily Living Hierarchy Scale (ADLH) score is greater than their ADLH score at next assessment.	All patients whose ADLH score is greater than 0 at initial assessment.
	Incidence/Failure to Improve	a) Patients whose initial ADLH score is less than their ADLH score at next assessment, OR b) Patients whose initial ADLH score is greater than 0 and their ADLH score at the next assessment equals their ADLH score at initial assessment	All cases.
Capacity to Manage Finances	Improvement	Patients whose initial financial management (item g2d) score is greater than their financial management score at next assessment.	All patients whose g2d score is greater than 0 at initial assessment
	Incidence/Failure to Improve	a) Patients whose initial g2d is greater than their next g2d score, OR b) Patients whose initial g2d score is greater than 1 and their g2d score at the next assessment equals their g2d score at initial assessment	All cases
Capacity to Manage Medication	Improvement	Patients whose initial medication management (g2b) score is greater than their medication management score at next assessment.	All patients whose g2b score is greater than 0 at initial assessment
	Incidence/Failure to Improve	a) Patients whose initial g2b score is less than their g2b score at next assessment, OR b) Patients whose initial g2b score is greater than 1 and their g2b score at the next assessment equals their g2b score at initial assessment	All cases

Domain	Type of Indicator	Numerator	Denominator
Pain	Improvement	Patients whose initial Pain Scale (PS) score is greater than their PS score at next assessment.	All cases where the initial PS score is greater than 0.
	Incidence/Failure to Improve	c) Patients whose initial PS score is less than their PS score at next assessment, OR d) Patients whose initial PS score is greater than 0 and their PS score at the next assessment equals their PS score at initial assessment	All cases
Acute Control Medications	Prevalence at time 1	Patients who are given an acute control medication at least once (k5 > 0) in the 3 days prior to initial assessment.	All cases at initial assessment.
	Prevalence at time 2	K5 > 0 in the 7 days prior to second assessment.	All cases at second assessment.
Physical Restraint Use	Prevalence at time 1	Patients who are physically restrained[4] at least once in the 3 days prior to initial assessment.	All cases at initial assessment.
	Prevalence at time 2	Patients who are physically restrained at least once in the 7 days prior to second assessment.	All cases at second assessment.
Interpersonal Conflict	Improvement	Patients whose initial Interpersonal Conflict Sum (ICS)[5] score is greater than their ICS score on their next assessment	All patients whose initial ICS score is greater than 0.
	Incidence/Failure to Improve	c) Patients whose initial ICS score is greater than 0 and less than their ICS score at next assessment. d) Patients whose initial ICS score is greater than 0 and their ICS score at the next assessment equals their ICS score at initial assessment	All cases

[1] The Disruptive Behaviour Sum is the sum items measuring wandering, inappropriate sexual behaviour, and elopement; [2] The Current Violence Sum is the sum of items measuring violent actions, intimidation, and ideation where present in the last 3 days; [3] A second assessment represents observations while a patient is current in-hospital and can be completed upon a change in status, after 90 days in their current stay, and at discharge; [4] Physical restraint use includes mechanical restraint, chair prevents rising, and physical/manual restraint by staff; [5] The Interpersonal Conflict Sum is the sum of 4 items measuring patient hostility toward friends/family, staff/other patients, family hostility toward patient, and staff consistently frustrated with patient.

4.1.5 Distribution of New MHQIs among Hospitals in Pilot and OMHRS Data

The unadjusted distribution of MHQI rates across facilities in the Pilot data are found in table 5 and in table 6 for the OMHRS data. Several MHQIs were retained because of their clinical importance, even though their rates were below 5% including prevalence of inpatient violence (means = 4.1% Pilot, 3.5% OMHRS), incidence or failure to decline in violence (means = 3.7% Pilot, 3.3% OMHRS), and prevalence of physical restraint use at 2nd assessment (means = 3.5% Pilot, 3.3% OMHRS). Aside from these, the two lowest MHQIs were prevalence of physical restraint use at time 1 (8.9% Pilot, 7.5% OMHRS) and incidence/failure to improve in disruptive behaviour (8.8% Pilot, 7.7% OMHRS). The highest MHQI rates were for improvement in depression (86% Pilot, 80% OMHRS), improvement in violent behaviour (78.2% Pilot, 74% OMHRS), and improvement in aggressive behaviour (80.8% Pilot, 71.7% OMHRS).

Unadjusted rates of improvement for depression based on the DSI were 78% in the Pilot data and 72% in the OMHRS data. The coefficient of variation (CV) for both data sets were low compared to other QIs but there was about a 28% difference between the first and third quartiles in the Pilot data and a 14% difference among facilities in the OMHRS data. Hospitals tended to vary more on rates of incidence/failure to improve in depressive symptoms with CVs of 41% in the Pilot data and 51% in the OMHRS data. Unadjusted rates of improvement in positive symptoms were 76% in both the Pilot and OMHRS data. Greater dispersion was found for rates of incidence/failure to improve in positive symptoms with CVs of 50% (Pilot) and 45% (OMHRS) compared with rates of 16% (Pilot) and 20% (OMHRS) for rates of improvement.

For the new QI domain measuring interpersonal conflict, rates of improvement were about 46% (Pilot) and 44% (OMHRS) while rates of incidence/failure to improve were 22% (Pilot) and 18% (OMHRS). The CVs for improvement were 37% (Pilot) and 47% (OMHRS) and were lower for rates of incidence/failure to improve (30% and 37%). There was about 20% difference between quintiles 1 and 3 for improvement in interpersonal conflict in the Pilot facilities and about a 30% difference in

OMHRs facilities. Rates of incidence/failure to improve ranged from about 18% to 25% among 1st and 3rd quartiles in Pilot and 13% to 22% in OMHRS hospitals.

A large amount of variation in MHQI rates existed between facilities in both the Pilot and OMHRS data. The CVs ranged from 16% to 134% in the Pilot data and 20% to 112% in the OMHRS data. Coefficients of variation tended to be larger among QIs measuring incidence/failure to improve and domains measuring control procedures and changes in physical (ADL) or daily functioning (e.g., Financial Management). The highest CVs tended to be among QIs with rates below 10%. For example, in the Pilot data the highest CVs were for the prevalence of physical restraint use at time 1 (84%) and time 2 (134%) where the prevalence rates were 8.9% and 3.5%, respectively. Among QIs with rates above 10% in the OMHRS data, the most variation was found for incidence/failure to improve in ADL functioning (71%), prevalence of restraint use at time 1 (58%), and incidence/failure to improve in financial management (54%). The lowest variation was for rates of improvement in positive symptoms (20%), depressive symptoms (24%), and aggressive behaviour (28%).

In both the Pilot and OMHRS data, the largest interquartile ranges were found for QIs that did not have the highest CVs. In the Pilot data, incidence/failure to improve in medication management (36% difference), incidence/failure to improve in financial management (35% difference), and incidence/failure to improve in cognition (34% difference) showed the greatest interquintile range. In the OMHRS data, improvement in ADL functioning showed the greatest interquartile range (33% difference) followed by improvement in cognitive functioning (figure 1; 31% difference) and improvement in interpersonal conflict (31% difference). A large amount of variation also exists in the first and third quartiles among many QIs. For instance, within the first quartile of hospitals in Figure 1 (hospitals 1 to 16), the rate of improvement in cognition ranged from 3% to just over 30% while the rates ranged from 65% to almost 90% between hospitals 51 and 67.

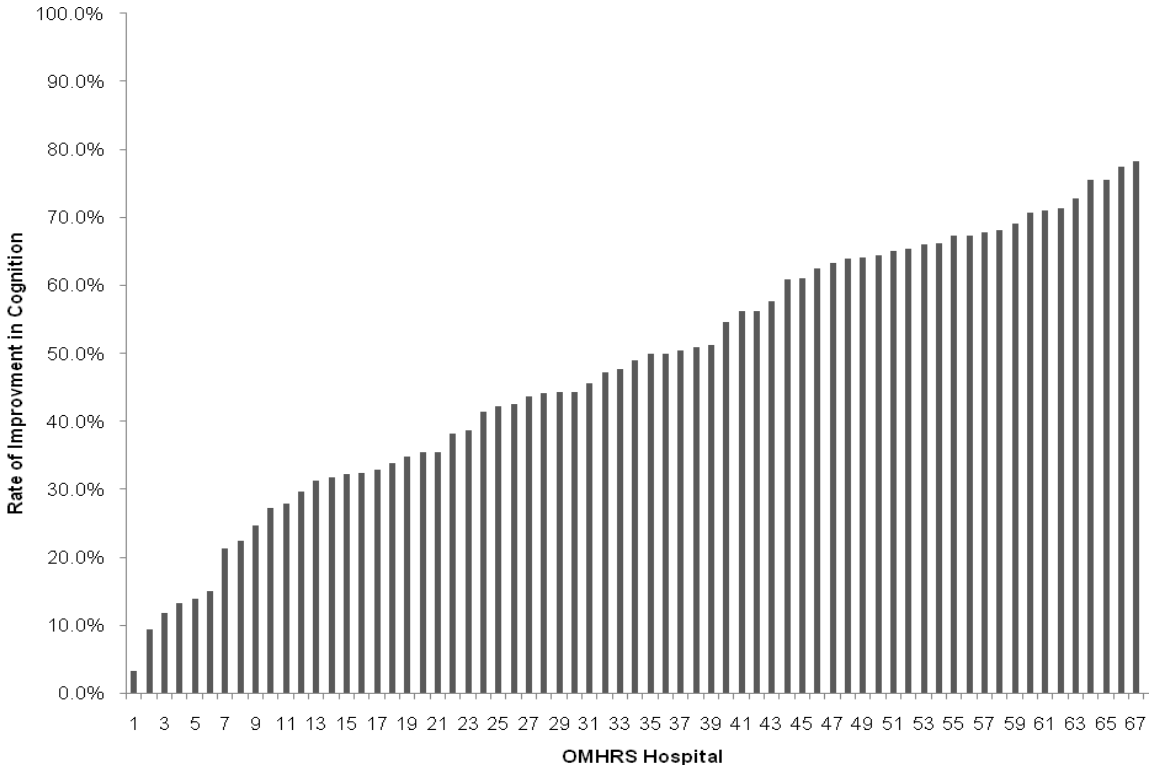


Figure 1. Range of unadjusted improvement in cognition scores among OMHRS hospitals.

Table 5. Means and distributions of unadjusted MHQI results among facilities in the Pilot data.

MHQIs		Pilot Data					
		Mean	SD	CV (%)	Median	Quartile 1	Quartile 3
Depressive Symptoms	Improvement	78.4%	14.0%	18	83.0%	62.4%	90.4%
	Incidence/failure to improve	21.4%	8.7%	41	16.8%	13.9%	30.8%
Aggressive Behaviour	Improvement	79.3%	12.3%	16	85.7%	73.7%	87.9%
	Incidence/failure to improve	12.3%	7.2%	59	11.3%	8.8%	13.9%
Disruptive Behaviour	Improvement	78.9%	14.9%	19	82.1%	68.7%	82.7%
	Incidence/failure to improve	8.8%	5.1%	58	8.0%	4.6%	12.2%
Violent Behaviour	Prevalence of Inpatient violence	3.9%	2.4%	62	4.8%	2.0%	5.1%
	Improvement	79.0%	12.4%	16	75.0%	68.3%	80.0%
	Incidence/failure to improve	3.7%	2.1%	57	4.2%	2.0%	4.7%
Positive Symptoms	Improvement	75.6%	11.8%	16	80.7%	60.9%	87.1%
	Incidence/failure to improve	20.2%	10.1%	50	20.9%	11.2%	24.7%
Cognition	Improvement	46.1%	13.9%	30	39.2%	37.3%	47.9%
	Incidence/failure to improve	33.8%	18.2%	54	34.5%	16.6%	50.7%
ADL Functioning	Improvement	57.3%	18.5%	32	56.5%	41.0%	59.9%
	Incidence/failure to improve	15.8%	10.2%	65	10.8%	8.8%	25.3%
Financial Management	Improvement	35.1%	15.3%	44	30.2%	25.1%	33.8%
	Incidence/failure to improve	31.6%	22.1%	70	21.8%	15.5%	50.3%
Medication Management	Improvement	45.5%	21.2%	47	39.9%	29.5%	54.6%
	Incidence/failure to improve	36.1%	23.9%	66	29.9%	18.2%	54.2%
Pain	Improvement	50.3%	22.6%	45	50.0%	39.4%	52.8%
	Incidence/failure to improve	17.3%	12.8%	74	13.0%	11.0%	17.9%
Acute Control Medication	Prevalence at first assessment	22.1%	14.4%	65	21.7%	12.1%	24.4%
	Prevalence at second assessment	11.7%	7.7%	66	9.4%	6.3%	14.7%
Physical Restraint	Prevalence at first assessment	8.9%	7.5%	84	4.2%	3.4%	14.5%
	Prevalence at second assessment	3.5%	4.7%	134	2.1%	1.0%	3.3%
Interpersonal Conflict	Improvement	46.2%	17.2%	37	50.0%	33.2%	53.4%
	Incidence/failure to improve	22.4%	6.7%	30	22.6%	18.1%	24.7%

Table 6. Means and distributions of unadjusted MHQI results among facilities in the OMHRS data.

MHQIs		OMHRS Data					
		Mean	SD	CV(%)	Median	Quartile 1	Quartile 3
Depressive Symptoms	Improvement	72.1%	17.1	24	79.2%	70.1%	83.8%
	Incidence/failure to improve	24.9%	12.6	51	20.7%	17.4%	29.8%
Aggressive Behaviour	Improvement	71.7%	20.2	28	78.8%	60.9%	86.7%
	Incidence/failure to improve	12.5%	7.1	57	10.5%	7.2%	16.5%
Disruptive Behaviour	Improvement	72.0%	22.5	31	82.1%	69.5%	87.3%
	Incidence/failure to improve	7.7%	5.4	70	6.3%	4.3%	9.9%
Violent Behaviour	Prevalence of Inpatient violence	3.5%	3.4	97	2.1%	1.6%	4.3%
	Improvement	74.0%	24.1	33	81.5%	65.3%	91.5%
	Incidence/failure to improve	3.3%	3.4	103	1.9%	1.4%	3.9%
Positive Symptoms	Improvement	75.8%	14.9	20	79.8%	70.2%	86.4%
	Incidence/failure to improve	19.0%	8.6	45	17.2%	13.2%	23.2%
Cognition	Improvement	48.2%	19.2	40	49.0%	33.9%	65.1%
	Incidence/failure to improve	27.0%	13.7	51	25.8%	16.6%	33.2%
ADL Functioning	Improvement	54.5%	20.1	37	55.6%	38.6%	72.1%
	Incidence/failure to improve	11.4%	8.1	71	9.3%	6.2%	12.5%
Financial Management	Improvement	35.0%	18.2	52	31.9%	22.5%	46.9%
	Incidence/failure to improve	25.3%	13.6	54	21.0%	16.4%	32.1%
Medication Management	Improvement	37.2%	18.1	49	34.2%	26.6%	48.4%
	Incidence/failure to improve	31.6%	15.2	48	29.6%	20.8%	38.9%
Pain	Improvement	47.5%	20.0	42	50.0%	35.7%	62.0%
	Incidence/failure to improve	17.0%	7.8	46	16.3%	11.8%	19.6%
Acute Control Medication	Prevalence at first assessment	17.3%	10.1	58	15.4%	9.9%	22.9%
	Prevalence at second assessment	8.2%	7.5	91	5.7%	3.5%	10.3%
Physical Restraint	Prevalence at first assessment	7.5%	5.7	76	6.6%	3.1%	9.7%
	Prevalence at second assessment	3.3%	3.7	112	2.2%	1.0%	3.6%
Interpersonal Conflict	Improvement	43.8%	20.8	47	42.9%	28.8%	59.7%
	Incidence/failure to improve	18.5%	6.8	37	18.1%	13.2%	22.4%

4.2 Application of Risk Adjustment to MHQIs

4.2.1 Distributions of Potential Risk Adjustment Variables among OMHRS Hospitals.

Table 7 shows the distribution of various diagnoses among OMHRS hospitals. Mood disorders were most common followed by schizophrenia/other psychoses and substance-related disorders. About 75% of hospitals had an 8% prevalence of dementia and a 14% prevalence of personality disorders. Mood disorders showed the highest interquartile range with about a 20% difference between the first and third quartiles. Among the top three diagnoses, the largest variation in prevalence among hospitals was for substance related disorders (CV = 58%) compared to schizophrenia (CV = 32%) and mood disorders (CV =26%). Among other diagnoses with prevalence greater than 5%, the greatest variation was for dementia (CV =99%), anxiety (CV = 66%) and personality disorders (58%). On average, about 42% of patients in each facility had at least 2 mental health diagnoses ranging by 19% between the first and third quartiles of hospitals.

Table 7. Means and distributions of the prevalence of diagnoses among OMHRS Hospitals.

	Mean	Sd	Median	Q1	Q3
Disorders of Childhood/Adolescence	1.6%	1.4%	1.2%	0.5%	2.2%
Dementia	7.9%	7.2%	6.0%	3.9%	8.3%
General Medical Condition Related	2.8%	2.6%	2.2%	1.2%	3.3%
Substance-Use Disorders	19.9%	11.0%	19.2%	11.3%	24.3%
Schizophrenia or other Psychosis	37.4%	12.3%	35.3%	30.6%	44.2%
Mood Disorders	54.1%	14.0%	54.0%	45.3%	65.3%
Anxiety Disorders	11.0%	7.1%	9.4%	6.3%	14.3%
Somatoform Disorders	0.8%	1.3%	0.5%	0.1%	1.0%
Factitious Disorders	0.1%	0.3%	0.0%	0.0%	0.1%
Dissociative Disorders	0.4%	0.8%	0.2%	0.0%	0.5%
Sexual Identity Disorders	0.2%	0.4%	0.0%	0.0%	0.4%
Eating Disorder	1.4%	2.9%	0.7%	0.3%	1.6%
Sleep Disorder	0.8%	1.2%	0.4%	0.0%	1.0%
Impulse Disorder	1.8%	1.7%	1.1%	0.6%	2.6%
Adjustment Disorder	3.6%	2.8%	2.9%	1.5%	4.8%
Personality Disorder	10.6%	6.1%	10.1%	6.1%	14.0%
Concurrent Diagnoses	42.4%	13.5%	43.8%	32.3%	50.9%
Medical Diagnoses	31.5%	18.0%	30.3%	19.9%	42.6%

Table 8 shows the mean prevalence of selected patient characteristics and RAI-MH embedded scales among hospitals in the OMHRS data. On average, the mean age of patients in OMHRS hospitals was about 47 and about 72% of patients were between 25 and 64. A large amount of variation in the prevalence of forensic patients among hospitals is due to the fact that 20 out of 70 hospitals had forensic patients with 10 of those hospitals having more than 30 forensic patients. The distributions of the DSI and the DRS were very similar, with more than half of hospitals having over 50% of patients with scores of 3 or more. On average, about 60% of patients in hospitals had positive symptoms and over 30% had PSS scores higher than 3. The larger average prevalence of patients with IADL scores greater than 3 compared the prevalence of those with ADL scores greater than 3 is attributable to the greater range in the IADL summary scale (0 to 42) compared to the ADL hierarchy scale (0 to 6). About a third of patients among hospitals, on average, had scores of 3 or more on the SCI, RHO, and SOS risk scales. The prevalence of patients with interpersonal conflict or current violence was about 10% among hospitals. About 36% of patients among OMHRS hospitals had been the victim of abuse.

Kruskall-Wallis tests revealed significant differences between hospitals on the means of all variables listed in table 8 ($p<0.0001$). A large amount of variation existed (CV > 50%) between hospitals on the prevalence of patients older than 65, with DSI scores of 0, with CPS scores of 3 or more, and with ADL scores of 1 or more. Interquartile ranges were greater than 20% for the prevalence of DSI, DRS, IADL, SCI, and SOS scores greater than 3 and the prevalence of patients who experienced any prior abuse.

Table 8. Means and distributions of the prevalence of potential risk adjustment variables among OMHRS hospitals.

	Mean	Std	Median	Q1	Q3
Mean Age	46.6	5.8			
under 25	11.6%	4.9%	11.3%	9.0%	13.6%
25 to 44	38.1%	8.2%	38.9%	36.1%	42.3%
45 to 64	33.8%	7.4%	33.8%	29.8%	38.2%
65 or more	16.4%	13.8%	13.7%	9.8%	17.5%
Male	48.0%	9.8%	46.2%	43.3%	52.2%
Forensic	3.7%	12.9%	0.0%	0.0%	0.2%
Depression Severity Index	3.9	1.1			
3 or more	56.0%	14.3%	56.1%	45.3%	67.4%
1 to 2	19.2%	6.1%	18.4%	15.3%	21.9%
0	24.8%	12.4%	24.3%	16.5%	30.5%
Depression Rating Scale	3.3	0.8			
3 or more	53.4%	13.9%	52.6%	43.9%	64.4%
1 to 2	28.9%	6.5%	29.2%	24.8%	32.8%
0	17.7%	10.0%	15.7%	9.6%	25.0%
Positive Symptom Scale - Long	3.8	1.2			
3 or more	30.1%	10.7%	29.8%	22.5%	36.7%
1 to 2	30.7%	5.2%	30.3%	27.7%	34.5%
0	39.2%	11.6%	39.5%	29.8%	47.1%
Cognitive Performance Scale	0.9	0.4			
3 or more	11.6%	7.0%	9.9%	7.6%	14.4%
1 to 2	32.5%	11.1%	31.4%	26.8%	37.6%
0	56.0%	15.4%	57.9%	50.2%	65.7%
ADL Hierarchy	0.5	0.3			
3 or more	8.3%	6.6%	6.3%	4.9%	8.7%
1 to 2	12.1%	6.1%	11.0%	7.7%	14.3%
0	79.6%	10.3%	82.1%	75.7%	86.0%
IADL Summary	5.7	3.2			
3 or more	41.8%	19.2%	37.0%	28.6%	51.2%
1 to 2	8.8%	4.5%	8.2%	6.5%	10.3%
0	49.5%	19.7%	52.2%	41.4%	62.9%
Self Care Index	2.1	0.5			
3 or more	31.1%	10.5%	30.0%	21.7%	39.5%
1 to 2	47.4%	7.6%	46.3%	43.4%	51.5%
0	21.5%	10.6%	19.8%	14.0%	27.1%
Risk of Harm to Others	2.0	0.5			
3 or more	32.4%	10.0%	31.9%	26.7%	39.4%
1 to 2	43.8%	8.1%	44.0%	38.2%	47.7%
0	23.8%	9.8%	23.4%	18.4%	29.5%

	Mean	Std	Median	Q1	Q3
Severity of Self-Harm	2.3	0.5			
3 or more	34.6%	11.2%	34.2%	25.5%	43.0%
1 to 2	45.3%	9.4%	44.1%	38.9%	50.7%
0	20.1%	9.8%	17.8%	13.6%	24.2%
Aggressive Behaviour Scale	1.3	0.5			
3 or more	19.7%	7.7%	20.4%	14.8%	23.4%
1 to 2	12.6%	4.2%	12.0%	9.6%	15.2%
0	67.7%	10.9%	68.2%	60.6%	75.2%
Pain	0.4	0.2			
3 or more	3.1%	2.6%	2.4%	1.6%	4.0%
1 to 2	21.1%	8.6%	19.7%	15.6%	25.8%
0	75.8%	10.3%	77.8%	70.2%	82.1%
SCIPP CMI among all inpatients	1.67	0.10			
Any Current Violence	9.8%	4.7%	9.8%	6.5%	12.2%
Any Abuse	35.6%	13.8%	34.9%	26.0%	44.5%
Any Conflict	10.5%	5.0%	10.3%	6.7%	13.2%

4.2.2 Evaluation of Potential Risk Adjustment Variables for MHQIs

The process for choosing MHQI risk adjustment variables began with bivariate correlation analyses. Spearman correlations were produced between the list of potential risk adjusters and each MHQI domain using the Incidence/Failure to Improve QI as the dependentvariable. Variables were retained for multivariate analysis if their correlations with the MHQI were greater than 0.10. For MHQIs where no correlations were 0.10 or greater, the top 10 significant correlations were retained. For example, all covariates had correlations less than 0.10 for MHQI2 (Incidence/Failure to Improve in Depressive Symptoms). Variables were also retained for multivariate analyses if they were included as potential risk adjusters for the JPPC quality domain or if the covariate is clinically relevant to the MHQI. Table 9 lists potential RAVs for each MHQI identified from bivariate Spearman correlation results.

Table 9. Potential risk adjustment variables selected based on Spearman correlations with MHQI domains.

DSI QIs		ABS QIs		Disruptive Behaviour		
ABS	CPS	ADL-Hierarchy	PSS	Incapable- Manage property	IADL	Anger
ADL	DSI	CPS	RHO	Substitute decision maker	Mania	CPS
IADL	Forensic	DSI	SCI	ADL hierarchy	PSS	Dementia Dx
Substance Diagnosis	SOS	IADL	Mood Dx	Difficulty managing finances	RHO	SCI
Dementia	Age 65 +	Mania	Personality	SCIPP CMI	ABS	
		Forensic		SCIPP Diagnosis		
Violence		**PSS**		**Cognition**		
ABS	PSS	Incapable- Manage property	Schizophrenia	ABS	Mania	
Any Conflict	RHO	IADL	SCI	Any Conflict	Personality	
Anger	Mania	Mood	SCIPP Diagnosis	Current Violence	RHO	
IADL						
ADL QIs		**Financial Management**			**Medication Management**	
IADL	ABS	Incapable- manage property	Current Violence	CHESS	Incapable to manage property	Current Violence	CHESS
Incapable- Manage property	Age 65 or more	Substitute decision	Dementia	CPS	Substitute decision maker	Dementia	CPS
Substitute decision maker	Any abuse	ADL Hierarchy	DSI*	SOS	ABS	DSI	SOS
RHO	Chess	Age 65 +	IADL	PSS	ADL Hierarchy	IADL	SCI
SCI	CPS	Any Abuse	Mood Dx	SCI	Age 65 or more	Mood Dx	SCI
Medical Dx	Dementia	Schizophrenia	RHO	ABS	Any Abuse	Schizophrenia	PSS
	General Medical	SCIPP CMI	Substance Dx			SCIPP CMI	RHO
	Substance Dx					Substance Dx	
	SOS						
Pain		**Acute Control Medication Use**			**Physical Restraint Use**	
Any Abuse	Pain Scale	ABS	Current Violence		Incapable to manage property	Current Violence
Age	Schizophrenia	ADL Hierarchy	DRS		Legal guardian/ substitute decision	Dementia
DRS	Medical Dx	Any Conflict	Extreme		ABS	Extreme Behaviour
Poor Self Rated Health		Command Hallucinations	Behaviour		ADL Hierarchy	IADL
Skin/Foot Condition		CHESS	IADL		Any Conflict	Mania
		CPS	Mania		CHESS	PSS
		SCI	PSS		CPS	RHO
		SCIPP CMI	RHO			SCI
		SOS	Schizophrenia			SCIPP CMI
Interpersonal Conflict						
ABS	Current Violence					
RHO	Personality					
Mania	Any Abuse					

Potential RAVs listed in table 9 were evaluated in multivariate contexts with regression models using GEE and logistic regression analyses among test and replication data. The results of GEE model replications are shown in Table 10. Almost all models predicting MHQIs performed reasonably well with c statistics greater than 0.70 for all but 3 QIs. The risk adjustment models were the weakest for the Depression QI. Only the CPS was chosen as a RAV for the Depression MHQIs as forensic status was not significant in the validation data and the odds ratio in the entire OMHRS data was less than 1.3. Interestingly, the model fit improved when facility was entered into the model. In the test data, the c-statistic from logistic regression changed from 0.55 to 0.68 when facility was entered into the model with CPS.

Table 10. Multivariate results for selecting risk adjusters based on test and replication data.

MHQI Domain	Covariate	Test Data				Replication Data			
		OR	95% CI		c	OR	95% CI		c
Depression	Forensic	1.46	1.08	1.98	0.53	*1.00*	*0.70*	*1.44*	0.54
	CPS 3+	1.50	1.30	1.73		1.37	1.20	1.57	
	CPS 1 to 2	1.05	0.96	1.14		1.02	0.92	1.13	
Aggressive Behaviour	Forensic	2.14	1.51	3.02	0.69	1.75	1.14	2.67	0.68
	CPS 3+	1.91	1.60	2.27		1.82	1.55	2.14	
	CPS 1 to 2	1.38	1.25	1.53		1.34	1.19	1.50	
	DSI 3+	0.63	0.56	0.70		0.73	0.64	0.83	
	DSI 1 to 2	0.89	0.74	1.06		0.90	0.79	1.03	
	PSS 3+	1.02	0.89	1.18		1.16	0.98	1.37	
	PSS 1 to 2	1.21	1.07	1.36		1.27	1.11	1.45	
	Any Conflict	1.45	1.25	1.68		1.39	1.21	1.61	
	dementia	1.55	1.26	1.91		1.55	1.30	1.86	
	Personality Dx	1.32	1.14	1.53		1.29	1.11	1.51	
	Mania 3+	2.02	1.74	2.36		1.82	1.58	2.09	
	Mania 1 to 2	1.55	1.37	1.75		1.43	1.20	1.70	
Disruptive Behaviour	CPS 3+	2.42	1.96	2.96	0.73	2.42	1.97	2.97	0.72
	CPS 1 to 2	1.37	1.19	1.60		1.38	1.19	1.60	
	DSI 3+	0.66	0.57	0.76		0.66	0.57	0.76	
	DSI 1 to 2	0.78	0.65	0.90		0.77	0.65	0.91	
	PSS 3+	1.46	1.21	1.76		1.46	1.21	1.76	
	PSS 1 to 2	1.29	1.11	1.50		1.29	1.11	1.50	
	Mania 3+	1.62	1.33	1.96		1.61	1.33	1.96	
	Mania 1 to 2	1.24	1.02	1.51		1.24	1.02	1.51	
	IADL 3 +	1.52	1.27	1.83		1.53	1.27	1.83	
	IADL 1 to 2	1.24	0.95	1.62		1.24	0.95	1.62	
	dementia	1.88	1.58	2.22		1.88	1.56	2.22	
Violence	CPS 3+	1.97	1.49	2.61	0.73	1.69	1.22	2.35	0.72
	CPS 1 to 2	1.22	1.01	1.48		1.25	1.04	1.51	
	Mania 3+	2.62	2.02	3.39		2.62	2.09	3.27	
	Mania 1 to 2	1.69	1.32	2.17		1.43	1.11	1.84	
	Any Conflict	2.09	1.66	2.64		1.82	1.39	2.35	
	SCIPP CMI	1.52	1.20	1.89		1.55	1.26	1.88	
	Male	1.72	1.47	1.99		1.65	1.28	2.12	

Table 10. Multivariate results for selecting risk adjusters based on test and replication data.

MHQI Domain	Covariate	Test Data				Replication Data			
		OR	95% CI		c	OR	95% CI		c
Positive Symptoms	Incapable - Property	1.38	1.19	1.59	0.62	1.36	1.18	1.56	0.63
	Forensic	2.08	1.62	2.66		1.29	1.01	1.64	
	CPS 3+	1.27	1.06	1.52		1.29	1.11	1.51	
	CPS 1 to 2	1.08	0.98	1.20		1.07	0.97	1.18	
	DSI 3+	0.72	0.64	0.81		0.70	0.63	0.78	
	DSI 1 to 2	0.95	0.81	1.11		0.89	0.77	1.06	
	SCIPP Diagnosis	1.21	1.06	1.38		1.16	1.13	1.21	
Cognitive Performance	Substance Dx	0.63	0.54	0.73	0.70	0.68	0.54	0.85	0.70
	Mood Dx	0.61	0.54	0.68		0.62	0.55	0.69	
	Age 65+	2.64	2.22	3.13		2.36	1.93	2.88	
	Age 45 to 65	1.28	1.13	1.45		1.20	1.02	1.39	
	Age 25 to 44	1.01	0.90	1.13		0.90	0.80	1.03	
	Mania 3+	1.43	1.32	1.56		1.24	1.15	1.34	
	Mania 1 to 2	1.21	1.10	1.33		1.09	0.98	1.22	
	dementia	3.07	2.44	3.87		3.21	2.69	3.82	
ADL Functioning	Incapable - Property	1.71	1.38	2.11	0.81	1.44	1.13	1.83	0.83
	Substitute Decision	1.51	1.26	1.81		1.62	1.35	1.95	
	Age 65+	2.27	1.72	2.99		2.42	1.84	3.19	
	Age 45 to 65	1.06	0.84	1.34		1.11	0.85	1.46	
	Age 25 to 44	0.71	0.57	0.89		0.70	0.55	0.88	
	CPS 3+	5.88	4.75	7.28		6.87	5.39	8.75	
	CPS 1 to 2	2.65	2.26	3.11		2.80	2.30	3.40	
	dementia	2.08	1.64	2.64		2.08	1.73	2.49	
Financial Management	Incapable - Property	1.87	1.57	2.24	0.80	2.09	1.68	2.61	0.80
	Substitute Decision	1.85	1.54	2.22		1.85	1.57	2.18	
	Age 65+	1.88	1.41	2.51		1.78	1.36	2.32	
	Age 45 to 65	0.90	0.73	1.10		0.80	0.65	0.97	
	Age 25 to 44	0.76	0.63	0.91		0.68	0.55	0.83	
	CPS 3	3.44	2.82	4.19		3.75	3.13	4.49	
	CPS 1 to 2	2.07	1.89	2.27		2.19	1.97	2.45	
	Schizophrenia DX	1.91	1.70	2.15		1.79	1.62	1.98	
Medication Management	Incapable - Property	1.50	1.24	1.80	0.78	1.54	1.27	1.87	0.79
	Substitute Decision	1.77	1.49	2.11		1.63	1.39	1.93	
	Age 65+	2.16	1.72	2.71		1.78	1.38	2.30	
	Age 45 to 65	0.99	0.84	1.16		0.83	0.68	1.01	
	Age 25 to 44	0.86	0.74	0.99		0.71	0.59	0.85	
	CPS 3+	2.83	2.33	3.45		2.83	2.29	3.51	
	CPS 1 to 2	2.03	1.84	2.23		1.89	1.71	2.09	
	Schizophrenia DX	2.26	2.01	2.54		2.23	1.89	2.63	

Table 10. Multivariate results for selecting risk adjusters based on test and replication data.

MHQI Domain	Covariate	Test Data				Replication Data			
		OR	95% CI		c	OR	95% CI		c
Pain	Medical Dx	1.80	1.61	2.01	0.67	1.82	1.65	2.01	0.68
	Poor Health	2.26	2.04	2.50		2.21	1.95	2.50	
	Age 65+	2.06	1.73	2.45		2.53	2.08	3.09	
	Age 45 to 65	1.84	1.59	2.13		2.08	1.70	2.54	
	Age 25 to 44	1.45	1.21	1.74		1.71	1.41	2.07	
	Skin/Foot Condition	1.50	1.31	1.72		1.27	1.04	1.55	
Acute Control Rx	Current Violence	1.65	1.45	1.88	0.74	1.74	1.50	2.02	0.75
	Extreme Behaviour	1.62	1.35	1.94		1.39	1.21	1.61	
	PSS 13 +	1.68	1.38	2.04		1.24	1.08	1.42	
	RHO 3+	2.16	1.80	2.58		2.43	2.04	2.89	
	RHO 1 to 2	1.28	1.13	1.44		1.36	1.13	1.64	
	ABS 6+	2.39	1.96	2.90		2.94	2.41	3.60	
	Mania 3+	2.24	1.97	2.53		2.22	1.90	2.59	
	Mania 1 to 2	1.49	1.29	1.71		1.47	1.25	1.73	
Physical Restraint	Current Violence	1.83	1.48	2.27	0.81	1.88	1.54	2.30	0.82
	RHO 3+	3.98	2.96	5.35		4.02	3.03	5.33	
	RHO 1 to 2	1.63	1.23	2.15		1.65	1.26	2.16	
	ABS 6+	4.78	3.88	5.89		4.52	3.60	5.67	
	Mania 3+	2.23	1.81	2.75		2.31	1.85	2.90	
	Mania 1 to 2	1.54	1.25	1.89		1.67	1.40	1.99	
Interpersonal Conflict	Any Abuse	1.48	1.31	1.66	0.65	1.52	1.35	1.71	0.66
	RHO 3+	1.71	1.41	2.08		1.78	1.53	2.07	
	RHO 1 to 2	1.25	1.10	1.41		1.18	1.06	1.32	
	Anger	1.31	1.25	1.37		1.32	1.26	1.37	
	Personality Dx	1.59	1.39	1.81		1.48	1.29	1.68	

With the exception of the Depression MHQI domain, all models developed in the test data were replicated in the validation data. The most common RAVs were the Mania scale and the CPS. The mania scale was particularly strong among the Acute Control Medication, Physical Restraint, and Aggressive Behaviour MHQIs. Age was also a common RAV, particularly for cognitive, Pain, ADL, and IADL based QIs. Even after controlling for cognitive status, patients older than 64 were more likely to experience decline/failure to improve in cognition, pain, ADL functioning, and financial/medication management.

The collapsed variable for the ABS was adjusted for inclusion as a RAV for acute control medication use and physical restraint. Scores of 6 or more (at least 2 behaviours occurred daily or > than 2 non-daily behaviours) were included rather than scores of 3 or more in order to prevent the use of this form of restraint on less severe behaviours from being identified by the QI. Similarly, for acute control medication use, scores on the PSS greater than 13 (person had at least 4 symptoms that were present daily or > 4 non-daily symptoms) were included rather than scores greater than 3.

Not all variables significantly related to a QI based on GEE models were included as RAVs. For instance, the ADL, SCI, and IADL scales all had odds ratios greater than 1.3 in relation to Cognition QIs. However, given the strong Speaman correlations between the CPS and the SCI (rho = 0.59), IADL (rho = 0.62), and the ADL (rho = 0.54) these variables were excluded. Similarly, the ABS, IADL, and SCI were excluded as risk adjusters for the ADL QIs given their strong relationship with either the ADL or the CPS. For medication management QIs, the PSS scale was excluded as it was collinear with a schizophrenia diagnosis and had a smaller impact on the c-statistic (0.77 if PSS excluded) than schizophrenia (0.75 when excluded). Lack of insight was also significant for financial and medication management but was excluded as it didn't improve the c-statistic and could be a gameable item.

4.2.3 Evaluation of the Impact of Risk Adjustment on MHQI Comparisons

Table 11 shows the mean unadjusted and adjusted MHQI scores among OMHRS hospitals. The average unadjusted QI Scores did not significantly differ from adjusted scores. The average unadjusted scores were slightly lower for 13 of the 27 QIs. After adjustment, less than 50% of patients in OMHRS hospitals, on average, achieved improvement in Cognition, ADL functioning, financial or medication management, pain, or interpersonal conflict. Rates of improvement were highest for depression, positive symptoms, and behaviours (aggressive/disruptive/violence). Interestingly, rates of incidence/failure to improve were also among the highest for depression and positive symptoms as well as cognition and financial/medication management.

The impact of risk adjustment is best illustrated by examining how changes in scores affect how hospitals compare. Table 12 shows patterns of change in absolute and quintile rankings of OMHRS hospitals. Almost all hospitals changed in absolute rank following adjustment, with the exception of improvement in violence where 50% changed ranks. For incidence/failure to improve in medication management and interpersonal conflict, 100% of the facilities changed absolute ranks.

Since changes in absolute ranks can be driven by potentially minute differences in QI scores, changes in quintile ranks were also examined. For instance, 91% of hospitals changed absolute ranks while 37% changed quintile ranks following adjustment of improvement in cognition. The QIs that experienced the most change in quintile rank after adjustment include incidence/failure to improve in interpersonal conflict (74%) and financial management (77%) as well as the prevalence of restraint use at time 1 (73%). Among hospitals whose quintile rank on the incidence/failure to improve in ADL functioning, 47% improved (62% of lowest ranked) and 41% declined (71% of highest ranked).

Table 11. Average unadjusted and adjusted MHQI scores among OMHRS hospitals.

MHQIs		Unadjusted Mean	Adjusted Mean	Adjusted SD	Adjusted 95% CI	
Depressive Symptoms	Improvement	72.1%	74.5%	16.2%	70.7%	78.4%
	Incidence/failure to improve	24.9%	24.4%	12.8%	21.4%	27.4%
Aggressive Behaviour	Improvement	71.7%	72.4%	20.2%	67.4%	77.3%
	Incidence/failure to improve	12.5%	12.8%	8.9%	10.7%	14.9%
Disruptive Behaviour	Improvement	72.0%	73.5%	20.2%	68.2%	78.7%
	Incidence/failure to improve	7.7%	9.3%	8.5%	7.2%	11.3%
Violent Behaviour	Prevalence of Inpatient violence	3.5%	5.0%	6.4%	3.5%	6.5%
	Improvement	74.0%	78.4%	20.6%	72.5%	84.4%
	Incidence/failure to improve	3.3%	4.6%	5.8%	3.2%	6.0%
Positive Symptoms	Improvement	75.8%	75.7%	15.0%	72.1%	79.3%
	Incidence/failure to improve	19.0%	19.5%	10.0%	17.1%	21.8%
Cognition	Improvement	48.2%	48.1%	20.0%	43.3%	52.9%
	Incidence/failure to improve	27.0%	25.3%	16.1%	21.5%	29.1%
ADL Functioning	Improvement	54.5%	50.4%	20.5%	45.1%	55.7%
	Incidence/failure to improve	11.4%	15.8%	14.3%	12.4%	19.2%
Financial Management	Improvement	35.0%	33.4%	19.2%	28.7%	38.2%
	Incidence/failure to improve	25.3%	24.7%	15.9%	20.9%	28.5%
Medication Management	Improvement	37.2%	36.1%	19.4%	31.4%	40.9%
	Incidence/failure to improve	31.6%	28.7%	16.7%	24.7%	32.7%
Pain	Improvement	47.5%	48.4%	19.6%	43.6%	53.2%
	Incidence/failure to improve	17.0%	15.8%	9.0%	13.7%	18.0%
Acute Control Medication	Prevalence -first assessment	17.3%	20.3%	17.0%	16.2%	24.3%
	Prevalence – second assessment	8.2%	8.2%	7.5%	6.4%	10.0%
Physical Restraint	Prevalence -first assessment	7.5%	13.2%	14.2%	9.8%	16.6%
	Prevalence – second assessment	3.3%	3.3%	3.7%	2.4%	4.2%
Interpersonal Conflict	Improvement	43.8%	44.1%	21.2%	38.9%	49.2%
	Incidence/failure to improve	18.5%	19.1%	9.8%	16.8%	21.5%

Note: SD = Standard Deviation, CI = Confidence Interval

Table 12. Patterns of change in absolute and quintile OMHRS hospital ranks between unadjusted and adjusted MHQIs.

		Absolute	Quintile				
						Quintile 1	Quintile 5
MHQIs		% Change	% Change	% Improve	% Decline	n improve/ quintile n	n improve/ quintile n
Depressive Symptoms	Improvement	86.5	32.9	18.1	21.4	2/13	4/14
	Incidence/failure to improve	89.4	30.4	20.0	17.9	3/13	2/14
Aggressive Behaviour	Improvement	96.9	24.7	15.4	15.3	1/13	4/13
	Incidence/failure to improve	95.7	60.9	38.2	37.5	4/13	6/14
Disruptive Behaviour	Improvement	88.0	28.9	17.0	18.7	0	3/12
	Incidence/failure to improve	94.2	66.8	38.2	44.6	7/13	6/14
Violent Behaviour	Prevalence of Inpatient	94.3	62.4	36.4	41.1	7/13	6/14
	Improvement	50.0	12.5	7.9	7.7	1/9	1/10
	Incidence/failure to improve	94.2	63.8	38.2	41.0	7/13	6/14
Positive Symptoms	Improvement	91.3	40.6	23.6	26.8	3/13	3/14
	Incidence/failure to improve	95.7	49.3	32.7	28.6	5/13	3/14
Cognition	Improvement	91.2	36.6	20.0	25.4	3/13	3/13
	Incidence/failure to improve	98.6	66.7	45.4	37.5	7/13	6/14
ADL Functioning	Improvement	94.9	38.9	20.8	27.1	1/11	4/11
	Incidence/failure to improve	95.7	62.1	47.3	41.1	8/13	10/14
Financial Management	Improvement	95.4	44.1	21.1	32.7	2/13	5/13
	Incidence/failure to improve	98.6	76.9	50.9	44.6	9/13	11/14
Medication Management	Improvement	93.9	50.0	24.5	37.7	3/13	6/13
	Incidence/failure to improve	100.0	72.9	42.7	42.9	7/13	10/14
Pain	Improvement	72.9	21.6	13.5	13.5	0	3/13
	Incidence/failure to improve	95.7	72.5	50.9	39.3	8/13	7/14
Acute Control Med.	Prevalence -first assessment	97.1	71.1	50.9	37.5	9/13	6/14
Physical Restraint	Prevalence -first assessment	97.1	72.9	54.5	41.1	10/13	7/14
Interpersonal Conflict	Improvement	87.9	19.7	11.3	13.0	2/12	2/13
	Incidence/failure to improve	100.0	73.9	49.1	42.9	9/13	9/14

Note: The sum of % Improve and % decline will not equal the % change as each has different denominators.

Figure 2 shows the unadjusted and adjusted rates of MHQIs among OMHRS hospitals. Adjustment had a small impact on depression QIs (Figures 2a & 2b) where rates shifted for 11 hospitals for both improvement and incidence/fail to improve. Rates of improvement in depressive symptoms ranged from about 50% to over 90% for all but 5 facilities. The greatest dispersion of MHQI rates after adjustment occurred for incidence/fail to improve in financial management (figure 2q). Adjusted rates ranged from less than 1% to almost 70% among OMHRS hospitals. Similar patterns of change in MHQIs occurred among all behaviour MHQIs. Among a number of QIs, rates tended to shift more among hospitals with higher scores. For instance, the highest rates of adjusted incidence/fail to improve in disruptive behaviour (40% and 43%) resulted from unadjusted scores of 18% and 20%. For the prevalence of acute control medication (figure 2v), 7 of the 10 hospitals with unadjusted rates of 30% or more had rates that increased after adjustment. Interestingly, unadjusted and adjusted rates of incidence/fail to improve in interpersonal conflict did not exceed 40%. Most hospitals that scored less than 20% still had scores less than 20% after adjustment.

Figure 2. Scatter plots of unadjusted and adjusted MHQI scores among OMHRS hospitals.

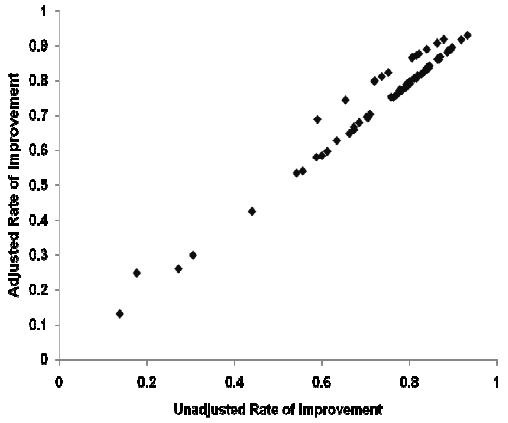


Fig. 2a. Improvement in depressive symptoms

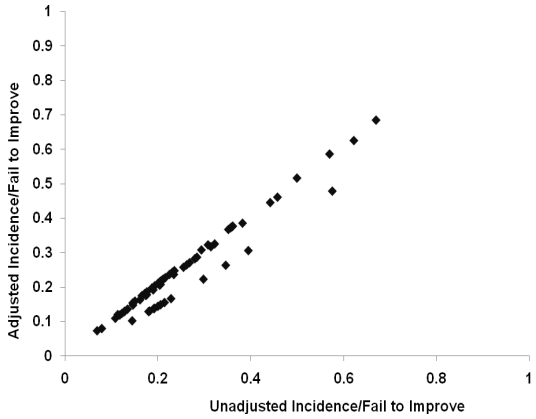


Fig 2b Incidence/failure to improve in depressive symptoms.

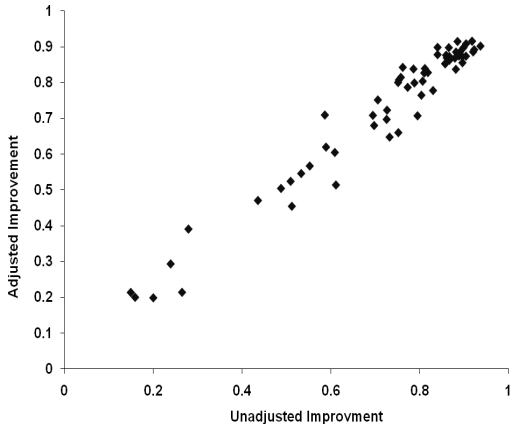


Fig. 2c Improvement in aggressive behaviour.

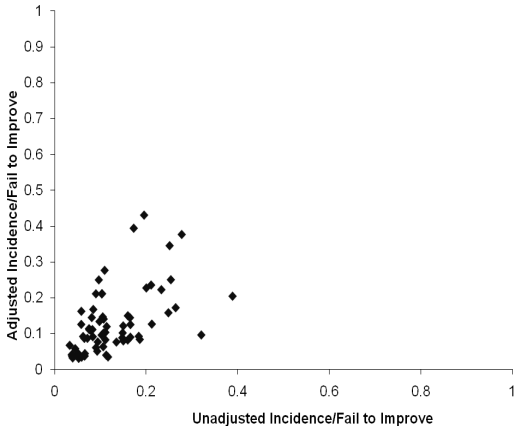


Fig. 2d. Incidence/failure to improve in aggressive behaviour.

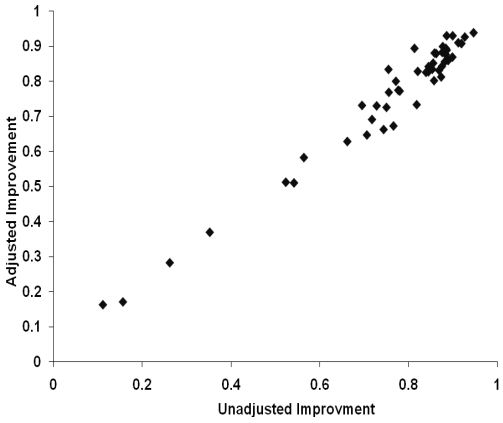


Fig. 2e. Improvement in disruptive behaviour.

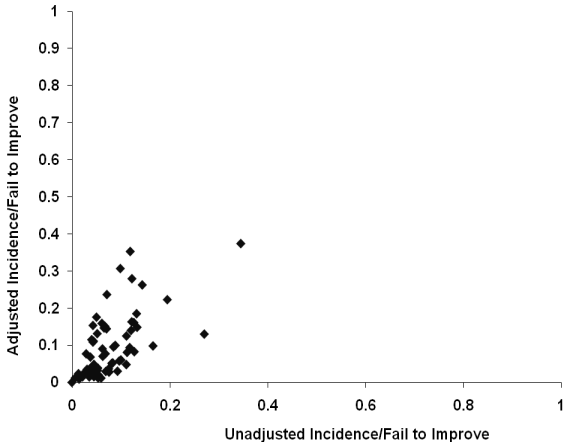


Fig. 2f. Incidence/failure to improve in disruptive behaviour.

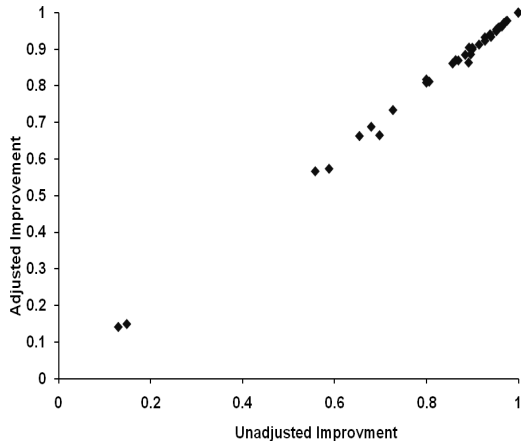


Fig. 2g Improvement in violent behaviour.

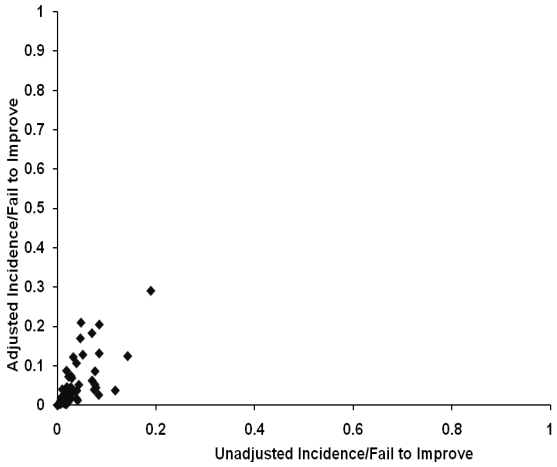


Fig. 2h Incidence/failure to improve in violent behaviour.

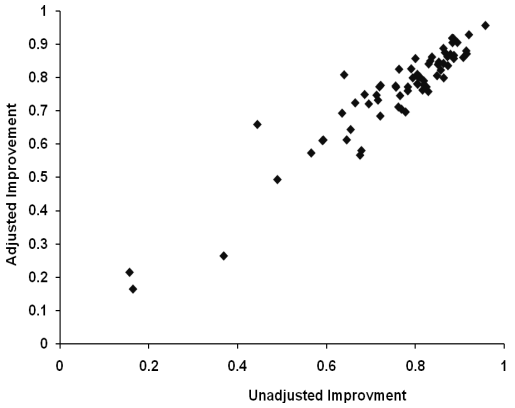


Fig. 2i. Improvement in positive symptoms.

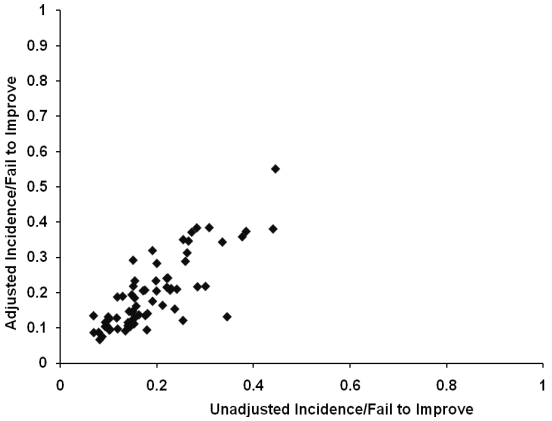


Fig. 2j. Incidence/failure to improve in positive symptoms.

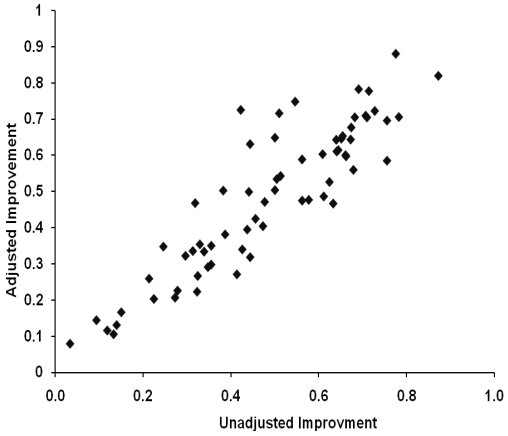


Fig. 2k. Improvement in cognitive performance.

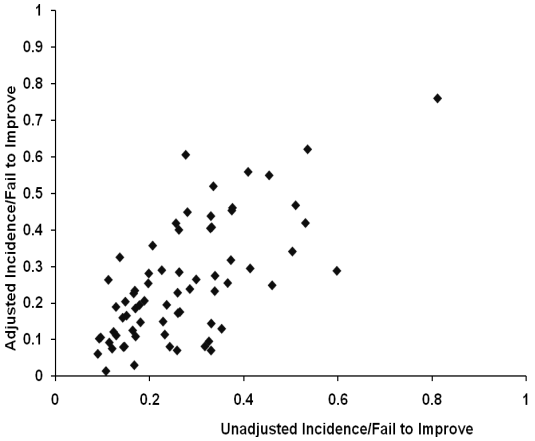


Fig. 2l. Incidence/failure to improve in cognitive performance.

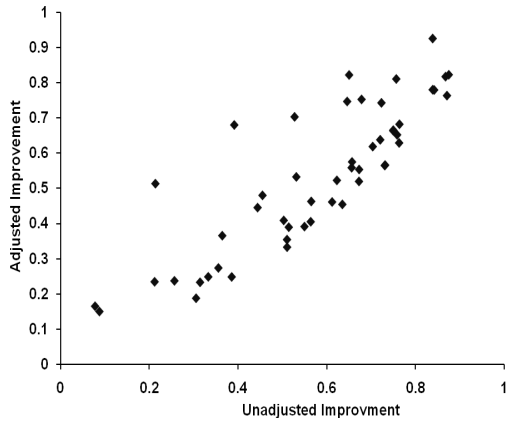


Fig. 2m. Improvement in ADL functioning.

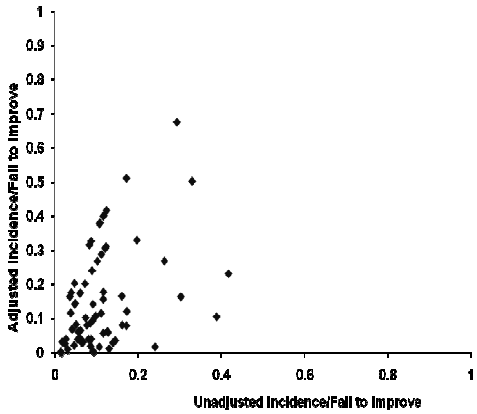


Fig. 2n. Incidence/failure to improve in ADL functioning.

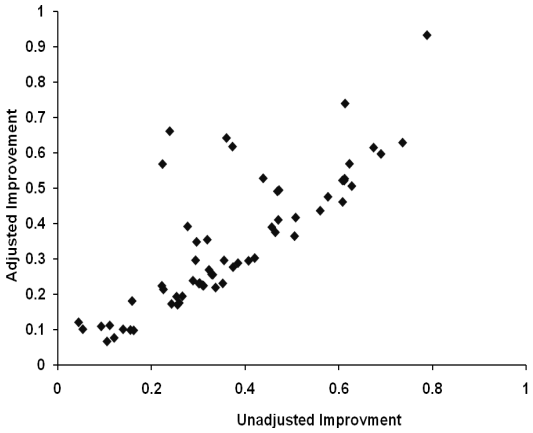


Fig. 2o. Improvement in financial management.

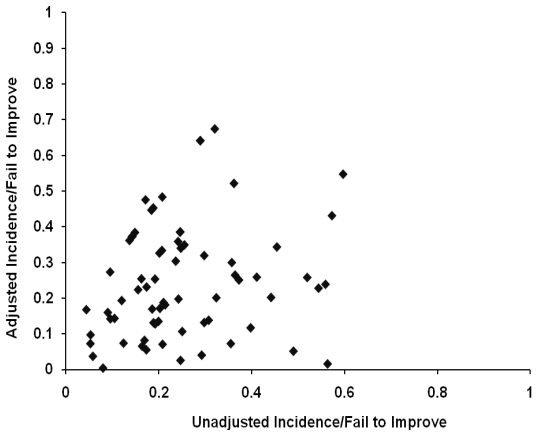


Fig. 2p. Incidence/failure to improve in financial management.

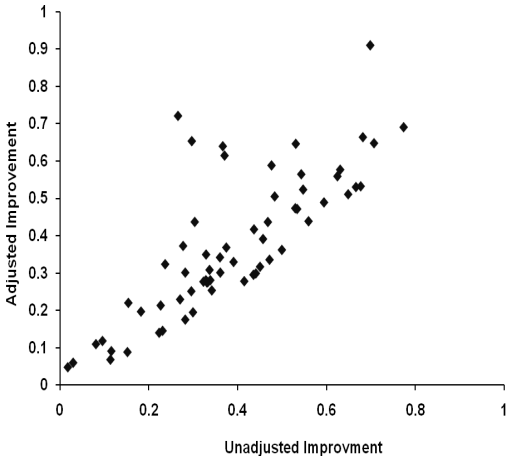


Fig. 2q. Improvement in med. management.

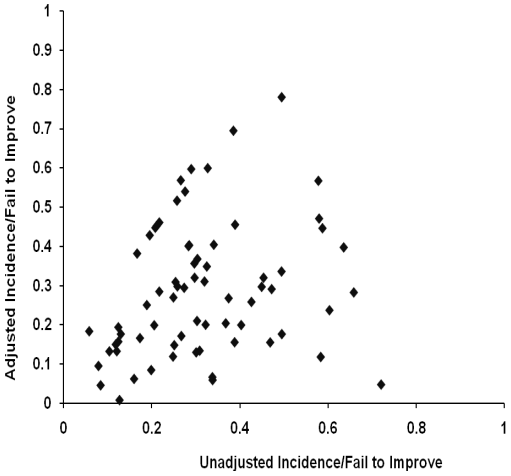


Fig. 2r. Incidence/failure to improve in med. management.

.

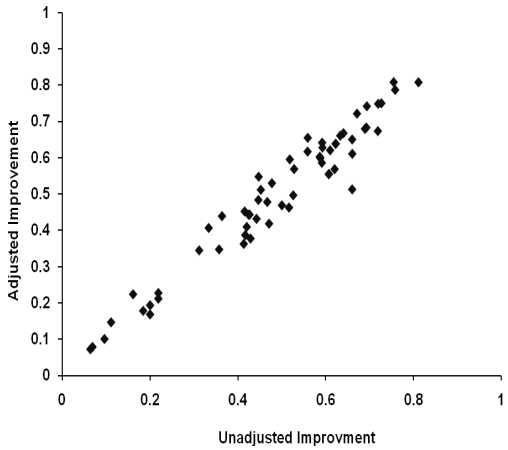


Fig. 2s. Improvement in pain.

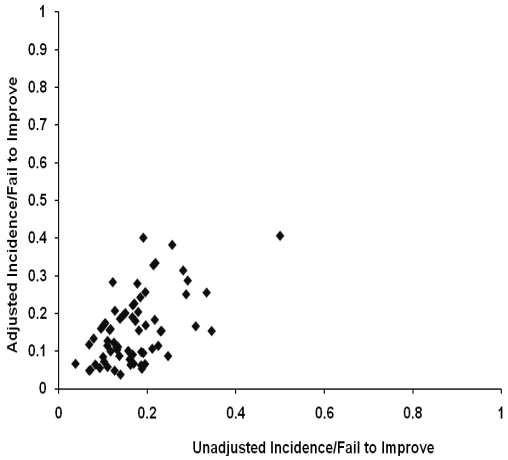


Fig. 2t. Incidence/failure to improve in pain.

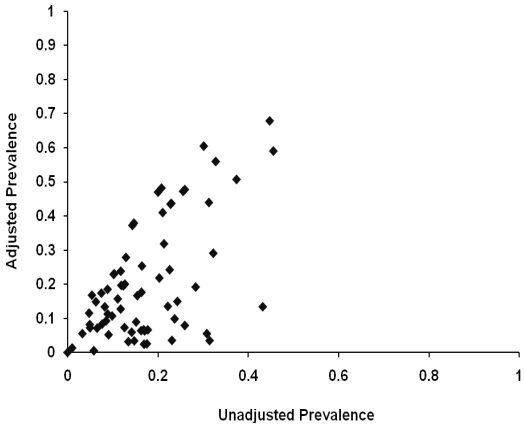


Fig 2u. Prevalence of time 1 acute control medication use.

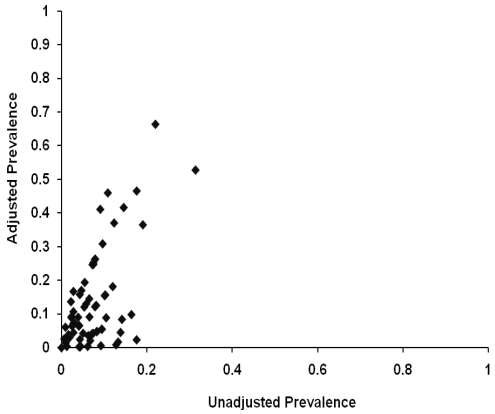


Fig 2v. Prevalence of time 1 physical restraint use.

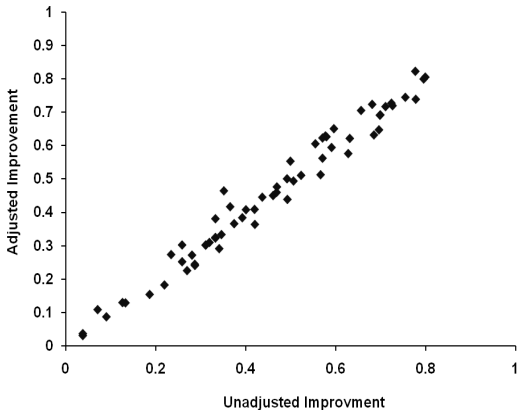


Fig 2w. Improvement in interpersonal conflict.

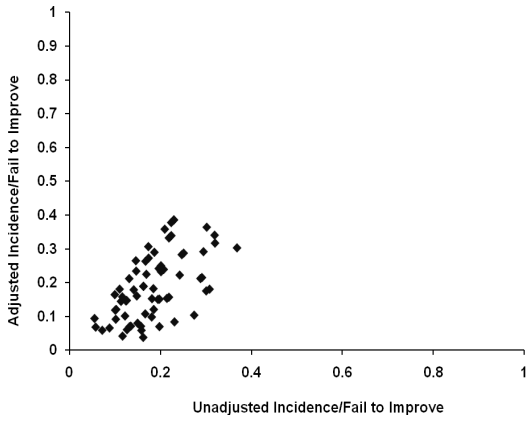


Fig 2x. Incidence/failure to improve in interpersonal conflict.

4.3 Development of Effectiveness Quality Indicators.

4.3.1 Patterns of EQI Scores among Individuals and Hospitals

Figure 3 shows the average scores for two types of EQI scores among all patients in the OMHRS data. Table 13 shows the distribution of two types of EQI scores among OMHRS hospitals. At both the individual and hospital levels, mean scores increased while variation among facilities decreased when persons with EQI measure scores of 0 were excluded. The mean scores among most EQIs more than tripled when those with a score of 0 at baseline were excluded. Behaviour based QI Measures (ABS, Disruptive, Violence) experienced larger increases in effectiveness with exclusion while symptom EQI measures such as the DSI and PSS experienced less dramatic, but still substantial, increases. For aggressive behaviour, the mean EQI score among hospitals went from 0.28 when all patients were included to 1.00 when those with an ABS baseline score of 0 were excluded. For the DSI, the average patient improved by just over 1 point per week between assessments when no patients were excluded. When those who scored 0 at time 1 on the DSI were excluded, the ratio increased to a 1.8 improvement when all baseline 0's were excluded.

Exclusion also tended to reduce the CV dramatically. For example, the CV for conflict went from 104% when none were excluded to 58% when baseline scores of 0 were excluded. This means that among those who have scores greater than 0 at baseline, improvement may be more uniform. However, the diversity in the number of patients with scores of 0 who develop scores greater than 0 at follow-up may drive variation across facilities. Regardless of the exclusion criteria, substantial variation existed in EQI scores among hospitals. For the PSS EQI, the lower 5% of facilities averaged less than 0.08 while the upper 5% averaged more than1.94 point improvement per week.

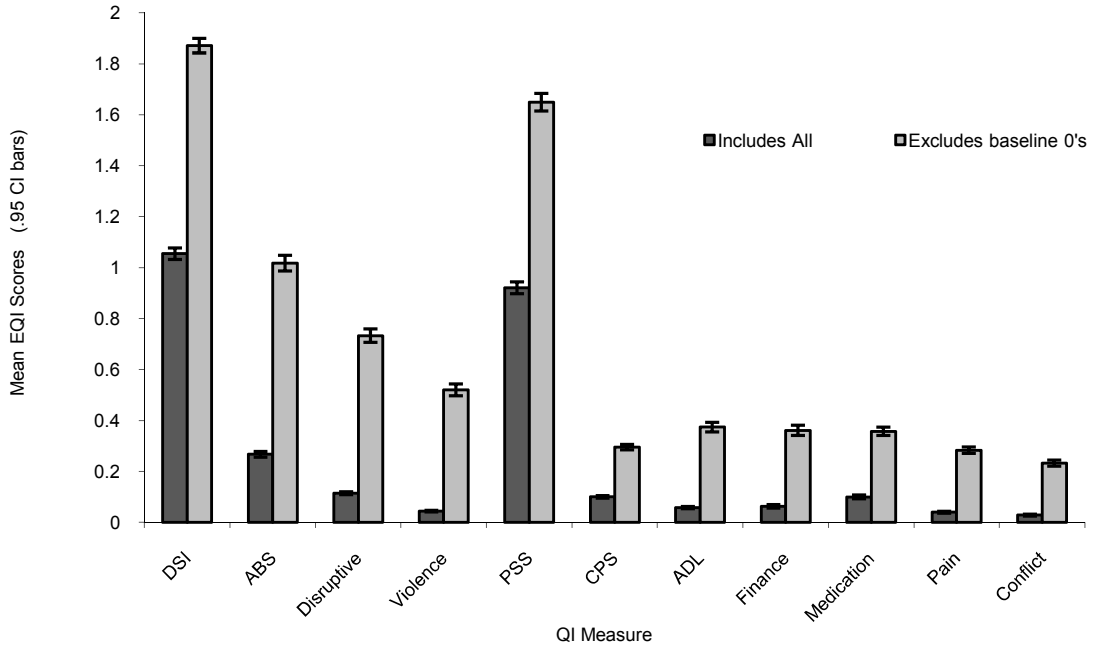


Figure 3. Mean EQI scores that include all patients compared to scores that exclude patients with baseline scores of 0 on each QI measure based on the OMHRS data

Table 13. Means and distributions of two types of EQI scores among OMHRS hospitals

EQI	Exclusion	Mean	SD	Coefficient of Variation (%)	5th Percentile	Median	95th Percentile
DSI	None	1.07	0.54	50	0.21	1.10	2.08
	Baseline 0	1.82	0.75	41	0.45	1.99	2.81
ABS	None	0.28	0.17	60	0.04	0.26	0.60
	Baseline 0	1.00	0.44	44	0.25	1.02	1.72
Disruptive	None	0.12	0.08	69	0.02	0.11	0.29
	Baseline 0	0.73	0.35	49	0.12	0.74	1.27
Violence	None	0.04	0.03	78	0.00	0.04	0.11
	Baseline 0	0.49	0.21	43	0.12	0.49	0.83
PSS	None	0.95	0.51	54	0.08	0.97	1.94
	Baseline 0	1.57	0.67	43	0.27	1.66	2.56
CPS	None	0.11	0.07	65	0.01	0.10	0.23
	Baseline 0	0.29	0.17	58	0.04	0.28	0.61
ADL	None	0.06	0.04	75	-0.01	0.06	0.13
	Baseline 0	0.38	0.22	57	0.05	0.38	0.75
Finance	None	0.07	0.07	104	-0.03	0.06	0.22
	Baseline 0	0.36	0.23	64	0.05	0.33	0.79
Medication	None	0.10	0.10	96	-0.01	0.09	0.27
	Baseline 0	0.36	0.24	67	0.03	0.32	0.82
Pain	None	0.04	0.04	93	-0.02	0.04	0.11
	Baseline 0	0.28	0.15	54	0.06	0.28	0.54
Conflict	None	0.03	0.03	104	-0.01	0.02	0.09
	Baseline 0	0.23	0.14	59	0.03	0.22	0.48

Note: Baseline 0 = cases where the QI measure scored 0 at time 1 were excluded

The percentage of patients with scores of 0 on baseline QI measures who develop scores greater than 0 at follow-up are shown in Table 14 with the correlation between both types of EQI scores. The DSI had, by far, the greatest incidence among those with no scores at time 1 followed by the PSS, medication management, and CPS. Interestingly, the correlation between the two types of DSI, PSS, and CPS EQI scores were also the strongest, even though 22%, 8%, and 6% of patients were missing from the calculation of the Baseline 0 EQI scores, respectively. Given the moderate to strong correlations between the two types of EQI scores, and that between 3% and 23% of patients who experience incidence of QI measure scores are be excluded with option 2, option 1 (denominator includes all patients) will be used for the calculation of EQI going forward.

Table 14. Incidence of MHQI measure scores among those who scored 0 at time 1 and correlations between 2 types of EQI among OMHRS Hospitals.

QI Measure	Incidence where Time 1 scores = 0 (%)	Correlation between 2 types of EQI Scores
DSI	22.6	.92
ABS	5.2	.73
Disruptive	3.0	.65
Violence	1.0	.70
PSS	8.2	.93
CPS	6.4	.84
ADL	2.2	.63
Finance	5.6	.55
Medication	7.5	.72
Pain	6.2	.78
Conflict	6.1	.74

4.3.2 Distribution of Unadjusted EQI Scores among Patient Types

Figure 4 shows the average and 95% CL for EQI scores by patient type in the OMHRS data. It is not surprising that there are large differences between patient types given the differences in lengths of stay. The average number of weeks between assessments was shortest for acute patients (mean = 3.3, .95CI =3.3,3.4), followed by long stay (mean = 6.5, .95CI = 6.3,6.8), geriatric (mean = 7.6, .95CI = 8.6,10.1), and forensic (mean = 9.4, .95CI = 8.6, 10.1). No matter how high the baseline QI measure score or the amount of change in the QI measure, forensic, long stay, and geriatric patients will be very unlikely to have unadjusted Effectiveness scores as high as acute patients. For most EQIs, scores indicated that, on average, patients in each group tended to improvement per 7 days between assessments. The exception is among the Interpersonal conflict EQI for long stay and forensic patients where, on average, scores tended to decrease slightly over time. The similarity between acute EQI scores and total EQI scores is due to the high representation of acute patients in the OMHRS data. For most EQIs, however, acute scores were still significantly higher than the average of all patient types.

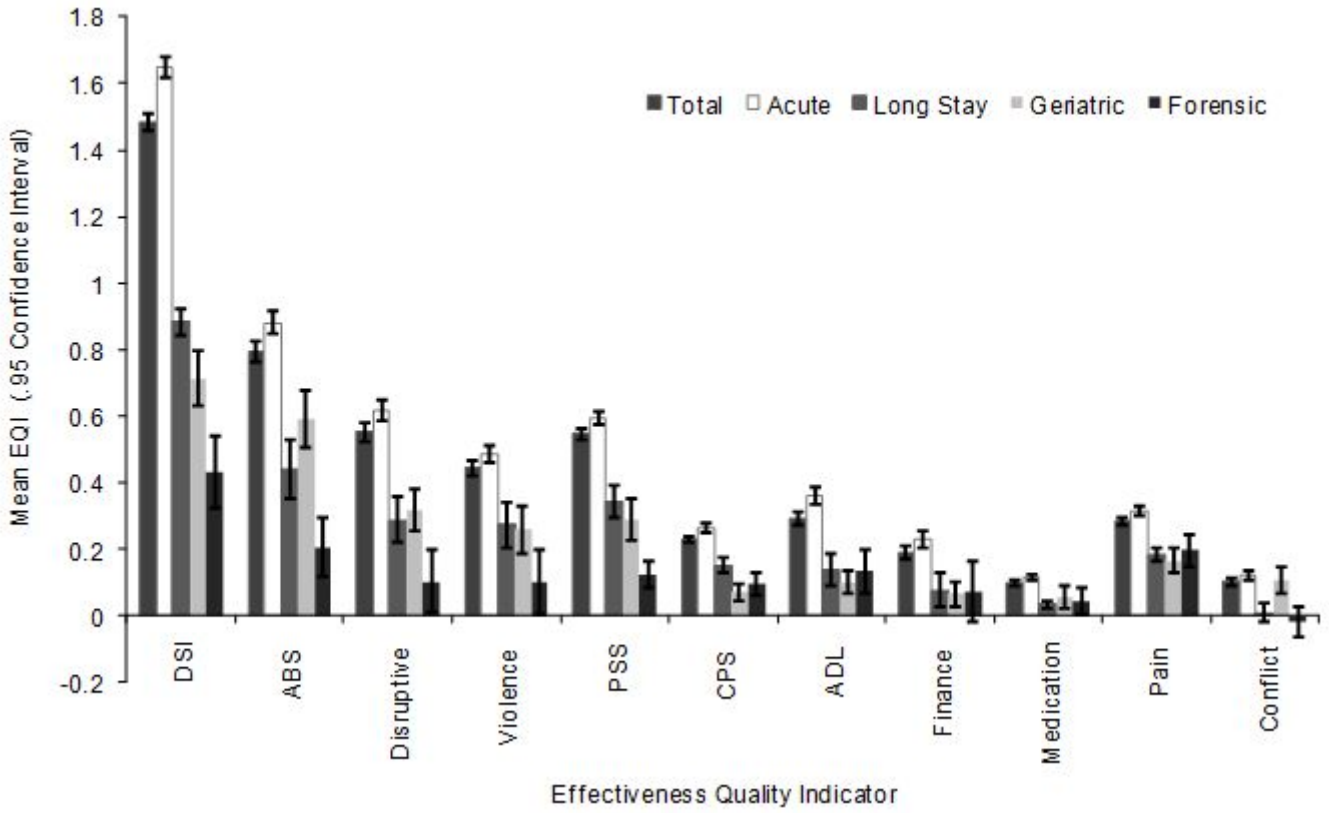


Figure 4. Mean and 95% confidence intervals of EQIs scores by patient type in OMHRs data.

4.3.3 Risk Adjustment of EQI Scores.

Baseline QI measures were examined as potential RAVs for EQI scores. A large amount of variation in baseline and change in QI measure score are evident among hospitals. Table 15 shows the mean, distribution, and correlation between baseline and change QI measure scores among OMHRS hospitals. Hospitals varied on baseline QI measures between 28% and 62%. Coefficients of variation ranged between 39% and 107% for QI measure change.

Moderate to strong correlations were found between baseline QI measures and change in QI measures over assessment time. Some of the strongest relationships were found for the DSI, PSS, ABS, Violence, and Disruptive measures. For instance, baseline PSS scores accounted for about 70% of the variance in predicting change (rho =0.84). The smallest relationships were found for the CPS, ADL-Hierarchy, Medication and Financial management, and conflict measures. For instance, baseline CPS accounted for about 11% of CPS change (rho = 0.33). Since baseline and change in QI measure scores are not evenly distributed among facilities and that change in QI measures is related to the baseline score of that measure. Therefore, baseline QI measure scores were evaluated for inclusion as RAVs for EQI scores.

Figure 5 shows the relationship between baseline QI measures and EQI scores. All correlations were statistically significant. More importantly, all were well above 0.10, the cut-off used to select RAVs for MHQIs. Interestingly, for MHQIs with weaker risk adjustment models such as DSI and PSS based QIs, the correlations between baseline and EQI scores were among the highest. The CPS and Financial/Medication management QIs had the lowest relationship between baseline and EQI scores. These results indicate that baseline QI measure score may be an important risk adjustment variable for EQI scores.

Table 15. Mean, distribution, correlation between baseline and change in MHQI measure scores for OMHRS hospitals.

MHQI Measure	Mean	SD	CV (%)	Range			Correlation
				5th Percentile	Median	95th Percentile	Spearman's rho*
DSI							
Baseline	3.92	1.08	28	2.28	3.94	5.58	0.72
Change	2.24	0.88	39	0.96	2.34	3.54	
ABS							
Baseline	1.29	0.51	39	0.44	1.32	2.13	0.72
Change	0.71	0.35	50	0.10	0.74	1.23	
Disruptive							
Baseline	0.54	0.24	44	0.18	0.54	0.95	0.78
Change	0.30	0.17	56	0.05	0.30	0.60	
Violence							
Baseline	0.16	0.09	56	0.03	0.15	0.33	0.74
Change	0.10	0.07	66	0.00	0.09	0.22	
PSS							
Baseline	3.82	1.21	32	1.90	3.91	5.96	0.84
Change	2.15	0.90	42	0.40	2.15	3.67	
CPS							
Baseline	0.89	0.38	42	0.43	0.82	1.34	0.33
Change	0.25	0.14	54	0.03	0.25	0.47	
ADL							
Baseline	0.47	0.29	62	0.17	0.38	1.03	0.42
Change	0.16	0.12	79	0.00	0.14	0.32	
Finance							
Baseline	1.36	0.75	55	0.44	1.18	2.80	0.40
Change	0.16	0.15	97	-0.04	0.14	0.46	
Medication							
Baseline	1.55	0.79	51	0.52	1.35	3.17	0.37
Change	0.23	0.25	107	-0.03	0.22	0.71	
Pain							
Baseline	0.42	0.19	46	0.17	0.38	0.81	0.65
Change	0.09	0.08	88	-0.03	0.09	0.21	
Conflict							
Baseline	0.40	0.15	38	0.17	0.39	0.69	0.45
Change	0.08	0.08	99	-0.01	0.06	0.22	

* $p < 0.0001$ for all correlations. CV = Coefficient of Variation

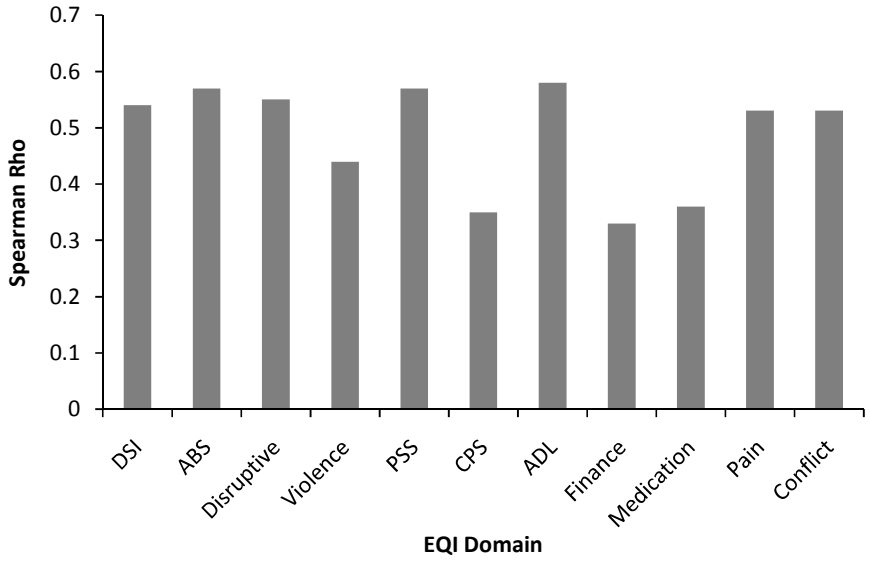


Figure 5. Spearman correlations between baseline MHQI measure scores and EQI scores.

4.3.4 Evaluation of the Impact of Risk Adjustment on EQI Scores

Adjusted EQI scores were estimated using multiple linear regression. Risk adjusters included covariates chosen for MHQIs and the baseline QI measure score. Model fit as assessed by R^2 ranged from 0.15 for the finance EQI to 0.49 for the violence EQI. Five of the 11 EQIs had R^2 greater than .40 (DSI, ABS, Disruptive, Violence, PSS), 3 were between 0.20 and 0.30 (CPS, ADL, Pain) , and 3 were below 0.20 (Financial and medication management, conflict).

Figure 6 shows the median and interquartile range of unadjusted and adjusted EQI scores among OMHRS hospitals. The median scores among hospitals were very similar between unadjusted and adjusted. The interquartile range between hospitals was reduced among adjusted EQI scores for the DSI, ABS, disruptive, PSS, and CPS scores. The distributions among EQIs with lower scores were very similar between unadjusted and adjusted EQIs.

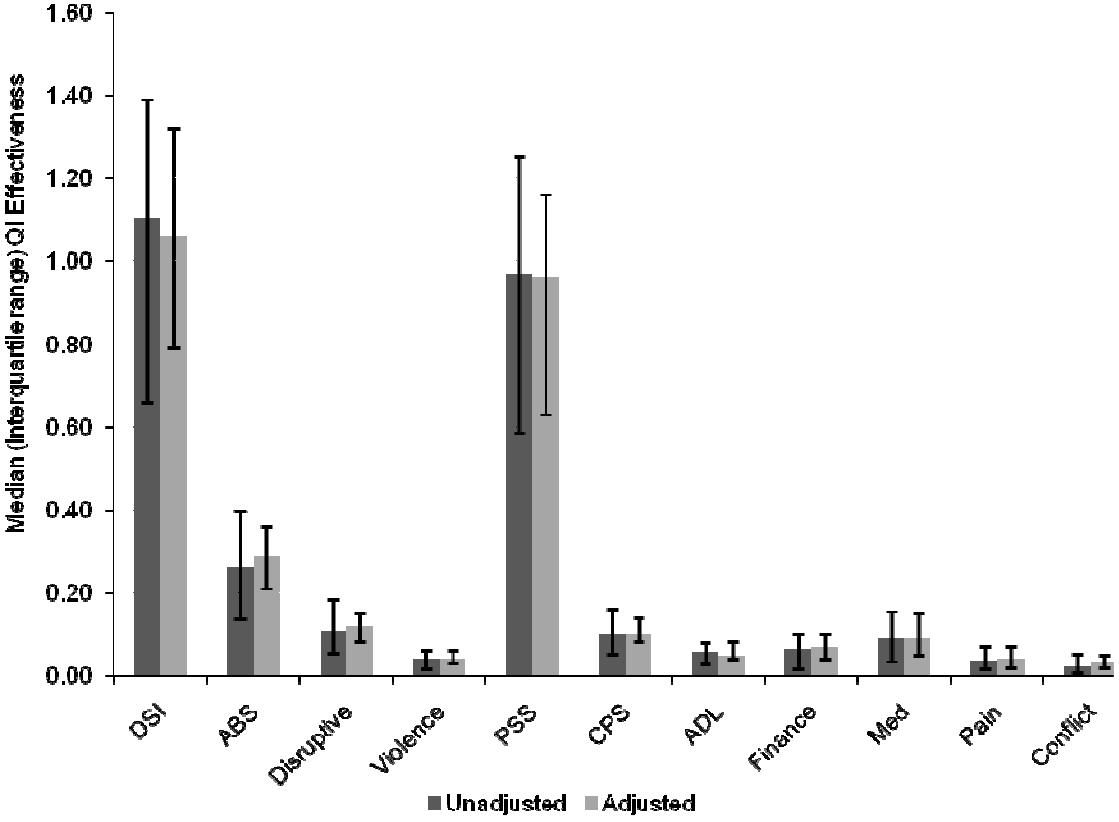


Figure 6. Median and interquartile range of unadjusted and adjusted EQI scores among OMHRS hospitals.

Unadjusted and adjusted hospital EQI scores were ranked into quintiles. The differences between quintile ranks were then compared to determine the number of hospitals who improved in rank and the number who declined following adjustment (figure 7). More than half of OMHRS hospitals were affected by adjustment, with slightly more tending to improve in quintile rank for the DSI, ABS, disruptive, and PSS effectiveness indicators. The largest numbers of hospitals that declined in rank were found for the Violence, ADL, financial management, and conflict EQIs where at least 5 more hospitals declined than improved.

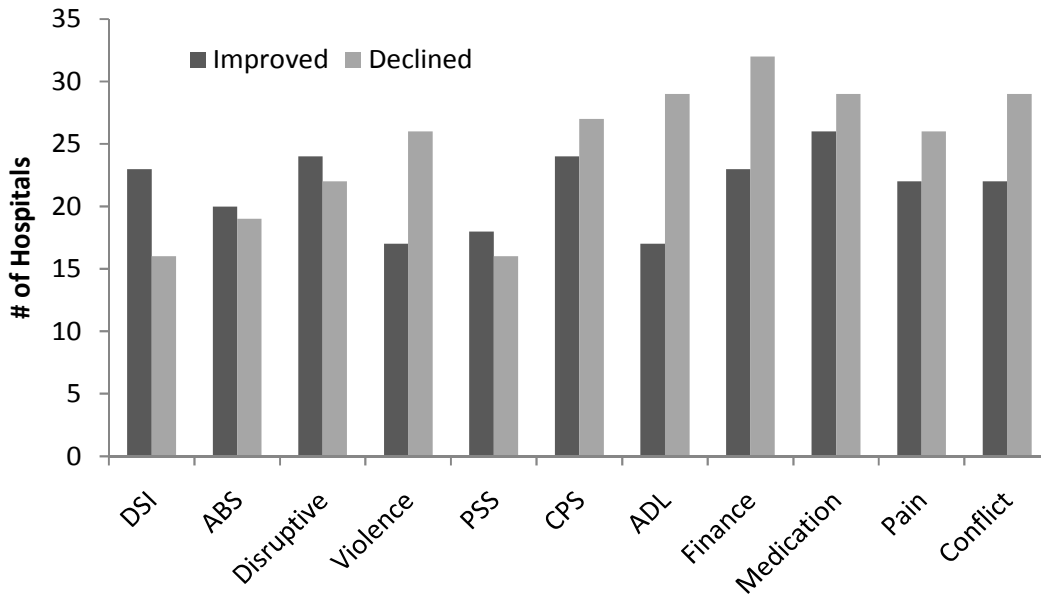


Figure 7. The number of OMHRS hospitals that changed in quintile ranks following adjustment of EQI scores.

To illustrate the movement in absolute ranks, figures 8a to 8k show the unadjusted and adjusted EQI scores among hospitals which are ranked on the unadjusted score. Almost all hospitals changed in absolute rank for all EQIs. As the figures illustrate, hospitals were most affected if they were ranked among the lowest or highest scores, with some exceptions. For most hospitals, the lowest scores tended to improve after adjustment and higher scores tended to decline after adjustment, with less movement in the middle ranked hospitals. Since the EQIs are based on measures of different scale, comparisons are not made for the amount of movement between unadjusted and adjusted scores among the EQIs. However, the pattern of scores tended to be more similar between unadjusted and adjusted scores for the DSI, PSS, and disruptive EQIs. For other scores such as the CPS, ADL, financial and medication management, and conflict EQIs, the pattern of the adjusted scores was less similar. For these EQIs, many of the lowest ranked hospitals had higher scores after adjustment the the highest ranked hospitals.

Figures 8. Line graphs comparing unadjusted and adjusted EQI scores.

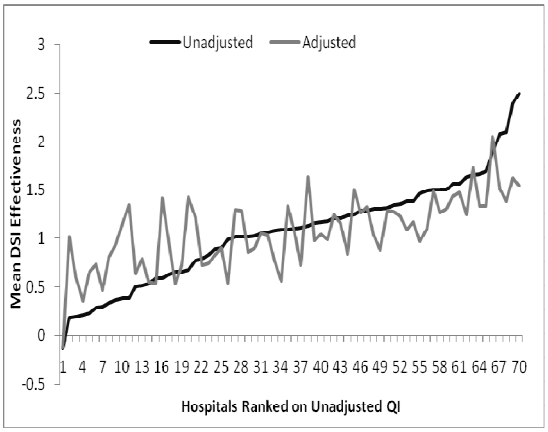


Fig. 8a Uadjusted & adjusted DSI effectiveness .

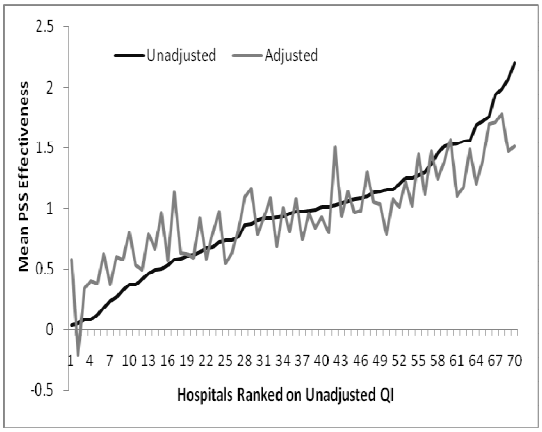


Fig. 8b Unadjusted & adjusted PSS effectiveness.

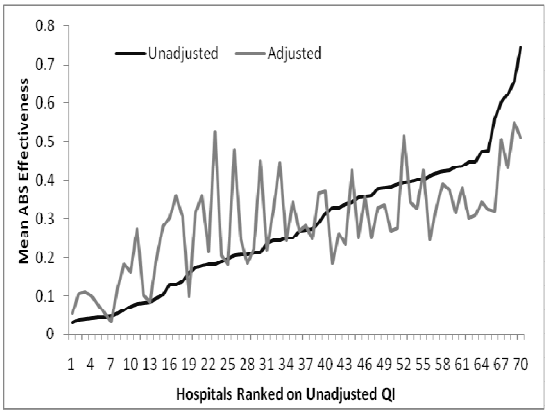


Fig. 8c Unadjusted & adjsuted ABS effectiveness.

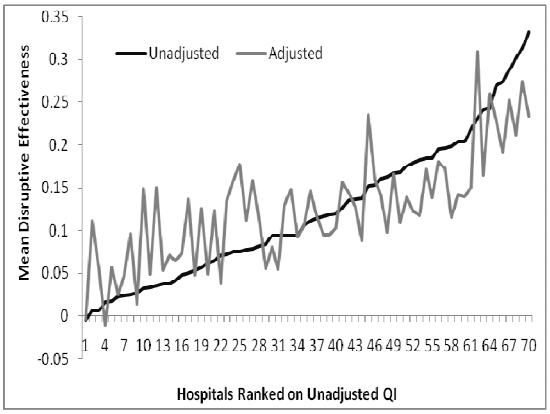


Fig. 8d Uadjusted & adjusted disruptive effectiveness

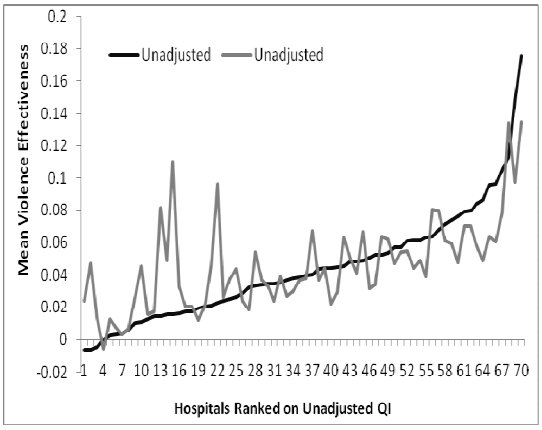


Fig. 8e. Unadjusted & adjusted violence effectiveness.

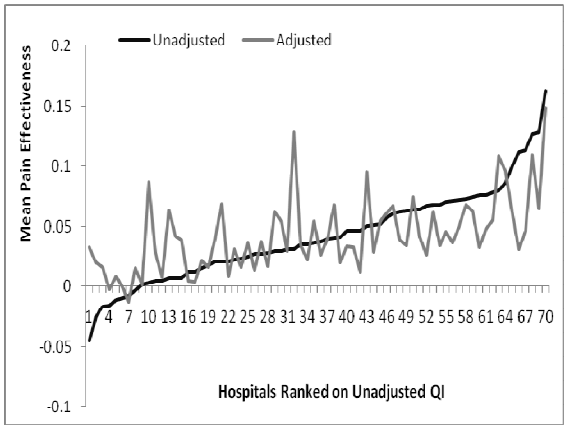


Fig. 8f. Unadjusted & adjusted pain effectiveness.

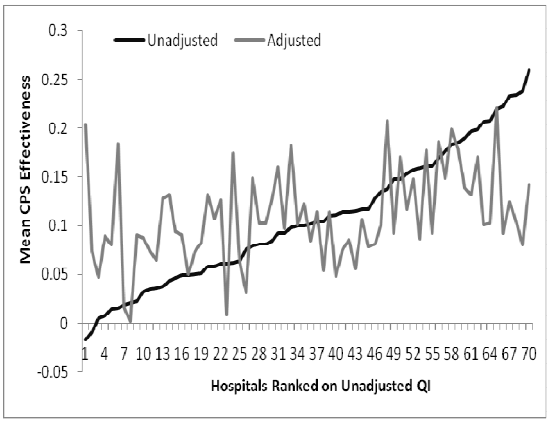


Fig. 8g. Unadjusted & adjusted CPS effectiveness.

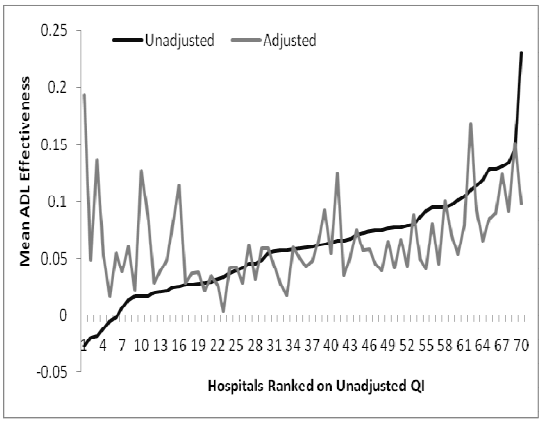


Fig. 8h. Unadjusted & adjusted ADL effectiveness.

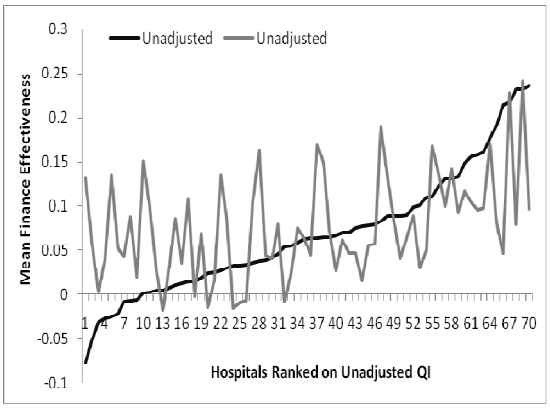


Fig. 8i Unadjusted & adjusted finance effectiveness.

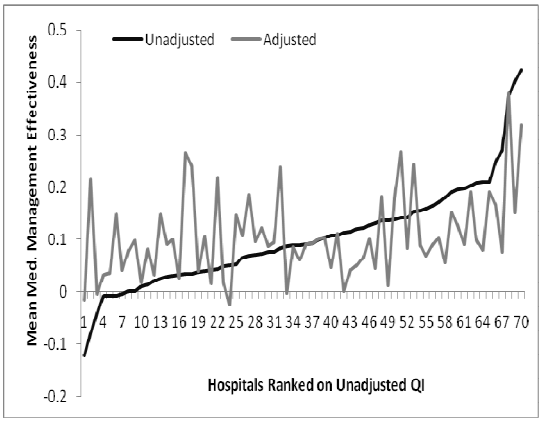


Fig. 8j. Unadjusted & adjusted medication effectiveness.

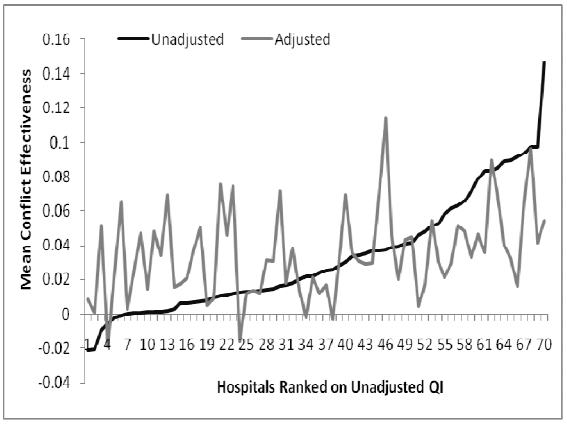


Fig. 8k. Unadjusted & adjusted conflict effectiveness.

Negative values in EQIs were found among hospitals for almost all unadjusted and adjusted QIs. A negative value indicates that patients in the hospital, on average, declined per week between assessments. Among hospitals with unadjusted negative EQI scores, almost all shifted to positive scores after adjustment. Hospitals whose scores were negative after adjustment tended to be those in the 2nd quartile of unadjusted scores. For instance, figure 39 shows that all 6 hospitals with negative scores for change in financial management were in the 2nd quartile of unadjusted scores.

4.4 Comparisons of LHINs using the MHQIs and EQIs

4.4.1 Patient Characteristics among LHINs in OMHRS Data

Table 16 shows the number of inpatient psychiatric units/hospitals and the distribution of selected patient characteristics for each LHIN in the OMHRS data. Significant differences between LHINs were identified for all characteristics in Table 15 except the percentage who were age 25 to 44 and 25 or less. All LHINs had at least 2 inpatient mental health units or hospitals to a maximum of 8. Significant differences were evident for the average age of patients between LHINs; however, the range in mean age differed by only 4 years between the lowest and highest mean age among LHINs. The greatest difference in age was among the prevalence of patients who were 65 years of age or older, ranging from 7% to 18%. The rate of involuntarily admitted patients differed significantly and ranged between 9% and 35%. With the exception of LHIN 3, the majority of all patients in each LHIN were acute. The rates of the four most common diagnoses significantly differed between LHINs. The rate of substance-related diagnoses ranged from 6% to 56%. LHINs that had among the highest rates of mood diagnoses (LHINs 4,8,13) and psychotic diagnoses (LHINs 5,7,9) also had among the highest rates of acute patients.

Table 16. Number of patients, hospitals, and prevalence of selected characteristics among LHINs in OMHRS data.

	LHIN													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
N	1250	2847	3765	2529	937	1276	5239	2022	2637	1293	3343	1179	1138	741
# hospitals with mental health beds	4	7	3	6	2	3	8	5	8	4	8	4	5	3
Age (Mean) *	45.3	46.9	43.4	45.2	43.6	46.6	44.0	47.2	46.8	46.6	46.9	43.2	45.7	44.5
% Under 25	13.0	13.1	10.0	12.4	14.1	12.7	12.0	10.5	10.2	12.1	10.7	14.6	10.1	16.9
% 25 – 44	37.8	35.3	45.7	39.0	42.6	38.1	44.2	38.1	40.3	36.5	38.8	43.3	39.6	39.1
% 45 – 64^	34.9	33.8	37.5	35.1	31.8	31.2	32.6	35.1	32.5	37.2	33.3	32.7	39.0	27.7
% 65 or more*	14.3	17.8	6.9	13.2	17.9	18.0	11.2	16.3	17.1	14.2	17.2	9.3	11.2	16.3
% Male *	47.7	49.8	54.4	48.4	47.8	44.2	52.3	42.2	47.4	53.0	48.4	57.2	45.8	51.1
% Involuntary *	27.4	22.9	11.3	33.1	34.7	14.9	30.7	25.4	30.2	18.4	24.7	9.2	27.1	28.6
Patient Type (%)														
Acute *	97.3	82.5	31.8	88.2	99.9	93.3	82.6	95.4	80.7	77.6	85.3	79.4	92.7	71.4
Long Stay *	1.4	5.7	66.7	8.8	0.1	0.3	7.6	4.0	10.0	7.0	1.7	3.9	5.0	6.6
Geriatric *	1.3	9.0	1.4	1.7	0	6.4	5.0	0.6	4.2	5.5	10.1	3.1	1.7	10.9
Forensic *	0	2.8	0	1.3	0	0	4.8	0	5.0	9.9	2.9	13.6	0.6	11.1
Diagnoses (%)`														
Dementia *	6.1	9.9	2.3	6.9	3.8	9.4	4.7	7.2	10.5	9.2	10.1	5.3	4.5	10.9
Mood *	59.2	57.9	45.8	60.6	39.3	55.6	45.3	58.3	49.7	48.9	46.6	49.4	62.1	48.6
Psychoses *	33.2	31.3	11.7	35.1	50.0	41.7	46.5	40.2	45.7	38.1	51.5	34.1	37.1	37.6
Substance *	14.1	21.3	56.0	20.7	5.9	10.7	29.6	16.6	14.0	25.5	23.0	38.5	21.1	31.7

* $p < 0.0001$ ^ $p < 0.05$,

` Diagnoses do not total 100% because patients can have more than one diagnosis.

4.4.2 Distribution of Mean and Prevalence of RAVs among LHINs

Table 17 shows the mean, standard deviation, median, and interquartile range of RAVs among LHINs. Kruskall-Wallis tests revealed significant differences between LHINS on the means of all variables listed in table 16. The average age among LHINs was 45 and about 5% more patients were under 44 compared to the percentage 45 or older. While the mean prevalence of forensic patients among LHINs was about 4%, 50% of LHINs had prevalence of forensic patients less than 3%. The CVs among LHINs in the prevalence of patients with embedded scale scores of 3 or more ranged from 16% for the DSI to over 50% for the Pain Scale. The CV for the prevalence of current violence was about 22% and 30% for the prevalence of any abuse or any conflict.

Table 17. Means and distributions of potential risk adjustment variables among LHINs.

	Mean*	Std	Median	Q1	Q3
Mean Age	45.4	1.4			
under 25	12.3%	1.9%	12.1%	10.5%	13.1%
25 to 44	39.9%	2.9%	39.2%	38.1%	42.6%
45 to 64	33.9%	2.8%	33.8%	32.5%	35.1%
65 or more	13.9%	3.3%	14.2%	11.2%	17.1%
Male	49.3%	3.9%	48.5%	47.4%	52.3%
Forensic	3.7%	4.5%	2.8%	0.0%	5.0%
Depression Severity Index	4.0	0.5			
3 or more	56.6%	8.5%	55.3%	48.3%	64.7%
1 to 2	18.7%	3.2%	18.5%	16.5%	20.9%
0	24.7%	6.7%	25.5%	18.0%	29.7%
Depression Rating Scale	3.3	0.5			
3 or more	52.6%	7.5%	49.5%	48.0%	58.5%
1 to 2	29.5%	3.3%	30.2%	26.3%	32.1%
0	17.9%	5.5%	18.4%	15.0%	21.4%
Positive Symptom Scale - Long	3.8	0.9			
3 or more	30.4%	7.7%	29.2%	27.9%	36.2%
1 to 2	30.0%	3.7%	29.5%	28.8%	32.8%
0	39.6%	10.2%	38.6%	33.2%	43.3%
Cognitive Performance Scale	0.8	0.2			
3 or more	11.2%	3.2%	10.5%	9.5%	11.8%
1 to 2	30.6%	5.2%	30.1%	28.4%	35.1%
0	58.3%	7.3%	58.1%	53.9%	59.9%

	Mean*	Std	Median	Q1	Q3
ADL Hierarchy	0.4	0.1			
3 or more	7.7%	2.4%	7.4%	6.1%	9.4%
1 to 2	11.4%	2.4%	11.4%	10.0%	12.7%
0	80.9%	4.4%	81.6%	76.7%	83.6%
IADL Summary	5.2	1.3			
3 or more	38.2%	8.7%	39.6%	35.1%	43.2%
1 to 2	9.5%	2.0%	8.9%	8.1%	10.7%
0	52.3%	9.3%	52.5%	44.4%	54.2%
Self Care Index	2.1	0.3			
3 or more	30.8%	6.9%	31.5%	27.2%	35.6%
1 to 2	46.1%	3.4%	46.1%	43.4%	47.7%
0	23.1%	7.3%	21.9%	18.1%	24.5%
Risk of Harm to Others	2.1	0.2			
3 or more	32.6%	5.8%	33.2%	31.4%	36.1%
1 to 2	44.1%	6.3%	43.6%	41.4%	45.9%
0	23.3%	3.8%	22.9%	21.1%	25.8%
Severity of Self-Harm	2.2	0.3			
3 or more	34.1%	6.0%	33.3%	30.8%	38.3%
1 to 2	44.0%	4.8%	44.6%	43.2%	46.5%
0	21.9%	7.0%	21.9%	17.3%	24.2%
Aggressive Behaviour Scale	1.3	0.3			
3 or more	19.3%	4.2%	19.7%	18.2%	22.7%
1 to 2	12.1%	2.8%	12.2%	11.0%	13.7%
0	68.6%	6.9%	67.9%	64.0%	71.7%
Pain	0.4	0.1			
3 or more	2.8%	1.6%	2.5%	2.0%	3.2%
1 to 2	20.2%	5.7%	18.3%	14.8%	25.5%
0	77.0%	6.9%	79.7%	71.2%	83.2%
SCIPP CMI	1.7	0.9			
Any Current Violence	9.9%	2.1%	10.5%	9.2%	11.4%
Any Abuse	37.2%	11.2%	34.8%	31.6%	40.8%
Any Conflict	10.1%	3.2%	10.2%	8.4%	11.9%

* Difference among LHINs all significantly differ based on Kruskall-Wallis tests $p<0.0001$

4.4.3 Rates of Unadjusted and Adjusted MHQIs among LHINs in the OMHRS Data

Table 18 shows the unadjusted MHQI rates among LHINs. At least half of all LHINs had rates of improvement above 75% for depressive symptoms, aggressive and disruptive behaviour, and violence. The median rates of incidence/failure to improve were at least 20% for cognition, financial management, and medication management. Rates of improvement for financial management (median = 32%) and medication management (median = 35%) were the lowest among all MHQIs.

Similar to the variation in rates among facilities, MHQI rates among LHINs tended to vary more for cognition and daily functioning MHQIS as well as pain and interpersonal conflict. A large amount of variation was also found among unadjusted prevalence of acute control medication (ACM) use (figure 9). LHIN 3, which had the highest rates of substance use diagnoses and long stay patients, had the lowest rates of ACM use. LHIN 9 had one of the highest prevalence of ACM use at time 1. For physical restraint use, rates at time 1 were below 10% for all but two LHINs. Interestingly, the LHIN with the highest rate of physical restraint use also had among the lowest rates of improvement in aggressive behaviour (62%) and violent behavior (59%) and the highest rate of incidence/failure to improve in aggressive behaviour (16%). Almost all patients in this LHIN (5) were acute and 50% had a schizophrenia or other psychosis diagnosis.

Table 18. Unadjusted MHQI rates among LHINs.

	LHIN													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
N	1250	2847	3765	2529	937	1276	5239	2022	2637	1293	3343	1179	1138	741
MHQI_1	79.9%	78.9%	84.6%	70.2%	76.7%	77.5%	86.4%	81.8%	87.8%	89.7%	85.8%	84.2%	87.8%	85.4%
MHQI_2	17.6%	13.4%	10.5%	18.5%	17.0%	17.4%	9.5%	15.0%	10.4%	7.6%	9.5%	12.4%	12.0%	14.6%
MHQI_3	76.3%	74.2%	84.0%	65.5%	62.6%	75.5%	77.7%	80.7%	69.8%	77.4%	77.2%	60.4%	83.6%	84.1%
MHQI_4	8.5%	13.6%	5.1%	12.2%	15.9%	10.2%	12.0%	10.1%	15.8%	11.3%	10.6%	15.6%	7.2%	13.2%
MHQI_5	76.6%	74.3%	83.3%	68.4%	62.0%	75.5%	80.2%	77.8%	64.1%	80.6%	79.8%	61.3%	79.7%	88.1%
MHQI_6	5.7%	8.5%	2.6%	6.5%	11.3%	7.1%	7.3%	6.1%	11.8%	6.8%	7.1%	8.6%	5.2%	6.5%
MHQI_7	2.9%	2.2%	1.0%	4.3%	5.5%	3.1%	4.2%	3.8%	3.9%	2.0%	1.7%	4.7%	2.7%	2.4%
MHQI_8	76.9%	87.7%	92.7%	67.6%	59.0%	75.2%	65.7%	77.5%	74.9%	89.6%	80.4%	66.0%	81.7%	94.9%
MHQI_9	2.6%	2.0%	1.0%	4.1%	5.3%	3.0%	4.0%	3.3%	3.7%	1.6%	1.6%	4.2%	2.6%	2.1%
MHQI_10	71.0%	77.8%	82.8%	62.1%	53.8%	66.7%	77.2%	78.8%	77.7%	77.3%	73.1%	75.4%	76.4%	88.3%
MHQI_11	9.4%	7.6%	2.8%	10.7%	19.9%	12.5%	9.8%	7.7%	11.6%	8.6%	9.6%	5.6%	9.1%	7.8%
MHQI_12	42.2%	47.8%	71.0%	38.4%	39.4%	48.0%	52.5%	62.3%	51.3%	53.9%	44.1%	39.1%	53.7%	62.3%
MHQI_13	24.5%	24.3%	9.8%	21.5%	31.6%	31.7%	25.6%	17.1%	29.4%	22.7%	28.1%	27.4%	23.8%	27.6%
MHQI_14	50.2%	50.3%	70.0%	48.5%	49.3%	55.3%	69.5%	63.9%	48.8%	59.5%	55.3%	42.8%	53.8%	59.3%
MHQI_15	9.6%	12.0%	3.4%	9.3%	11.6%	12.1%	7.8%	8.3%	14.4%	12.4%	11.4%	11.2%	8.1%	10.5%
MHQI_16	24.6%	31.7%	51.5%	31.6%	28.5%	27.5%	43.7%	53.4%	33.2%	34.4%	32.1%	28.0%	29.2%	53.5%
MHQI_17	33.8%	25.8%	6.3%	19.6%	27.0%	36.6%	22.0%	18.4%	31.2%	24.8%	26.4%	14.7%	19.1%	22.5%
MHQI_18	28.8%	34.2%	56.1%	34.6%	29.6%	33.2%	47.7%	49.9%	36.3%	34.8%	35.4%	28.7%	35.4%	45.7%
MHQI_19	42.1%	32.8%	9.3%	26.7%	30.7%	40.0%	27.3%	28.8%	39.7%	31.8%	30.6%	26.6%	29.9%	28.0%
MHQI_20	41.5%	45.7%	55.6%	30.1%	54.3%	44.1%	56.8%	53.0%	54.4%	61.0%	48.6%	42.2%	56.9%	63.5%
MHQI_21	15.1%	22.2%	17.6%	13.7%	7.8%	12.7%	12.9%	14.2%	14.8%	14.4%	19.4%	14.8%	22.8%	19.2%
MHQI_22	9.1%	23.7%	4.9%	9.7%	28.7%	21.2%	21.5%	16.9%	24.4%	22.2%	19.5%	8.1%	15.7%	9.7%
MHQI_23	4.1%	9.1%	1.8%	6.8%	16.8%	12.4%	9.0%	10.0%	10.1%	6.5%	8.1%	5.6%	6.4%	4.5%
MHQI_24	2.2%	7.3%	1.9%	4.3%	20.1%	7.8%	5.8%	8.5%	12.9%	7.3%	7.0%	5.0%	4.6%	6.5%
MHQI_25	1.0%	3.1%	1.1%	2.4%	8.3%	3.7%	1.8%	2.8%	4.5%	2.3%	3.6%	3.8%	1.5%	4.8%
MHQI_26	46.5%	51.8%	59.8%	28.9%	43.1%	35.8%	52.9%	55.0%	51.9%	46.8%	40.9%	35.6%	27.7%	66.8%
MHQI_27	13.3%	14.7%	13.1%	20.1%	22.6%	21.3%	16.6%	17.2%	17.0%	17.2%	18.6%	19.4%	22.9%	17.2%

Note: **MHQI_1, _2** = Improvement, Incidence/Fail to improve in depressive symptoms., **MHQI_3, _4** = Improvement, Incidence/Fail to improve in aggressive behaviour, **MHQI_5, _6** = Improvement, Incidence/Fail to improve in disruptive behaviour; **MHQI_7**= Prevalence of inpatient violence, **MHQI_8, _9** = Improvement, Incidence/Fail to improve in violence; **MHQI_10, _11** = Improvement, Incidence/Fail to improve in positive symptoms; **MHQI_12_13** = Improvement, Incidence/Fail to improve in cognition; **MHQI_14, _15** = Improvement, Incidence/Fail to improve in ADL; **MHQI_16, _17** = Improvement, Incidence/Fail to improve in financial management; **MHQI_18, _19** = Improvement, Incidence/Fail to improve in medication management; **MHQI_20, _21** = Improvement, Incidence/Fail to improve in pain; **MHQI_22, _23** = Prevalence of acute control medication use at time 1, time 2; **MHQI_24, _25** = Prevalence of physical restraint use at time 1, time 2; **MHQI_27, _28** = Improvement, Incidence/Fail to improve in interpersonal conflict.

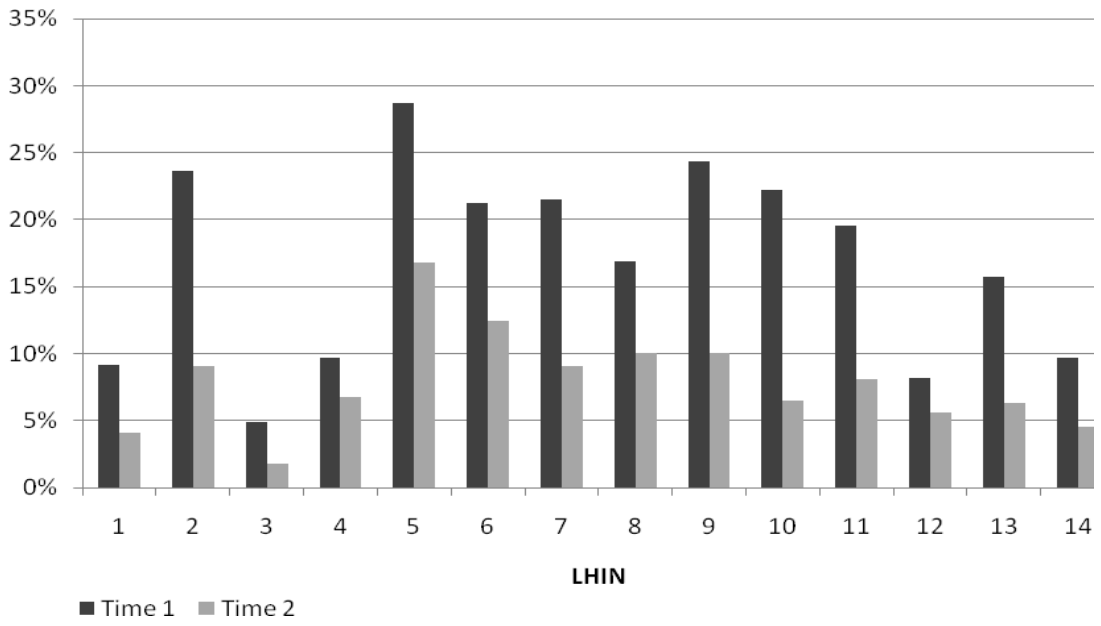


Figure 9. Prevalence of acute control medication use at time 1 and time 2 by LHIN.

Figure 10 shows the median and range of unadjusted and adjusted MHQI rates among LHINs. The least amount of change (about 1% to 3% difference) in MHQI distribution following adjustment occurred for improvement in depression, aggressive and disruptive behaviour, and violence. For other MHQIs, adjustment had a large effect on the distribution of rates. For the prevalence of physical restraint, the distribution increased by almost 20% after adjustment. The range of MHQI rates at least doubled following adjustment for the incidence/fail to improve in ADLs (MHQI 15), aggressive behaviour (MHQI 4), disruptive behaviour (MHQI 6), violence (MHQI 8), and all QIs for financial management (MHQIs 16,17) and medication management (MHQIs 18,19). Among all QIs measuring improvement, adjusted median rates were less than 50% for improvement in cognition, ADLs, financial or medication management, and interpersonal conflict, with maximum rates for financial or medication management never exceeding 60%. With the exception of ADL and interpersonal conflict, median rates of incidence/fail to improve all exceeded 20% as did incidence/fail to improve in depressive symptoms.

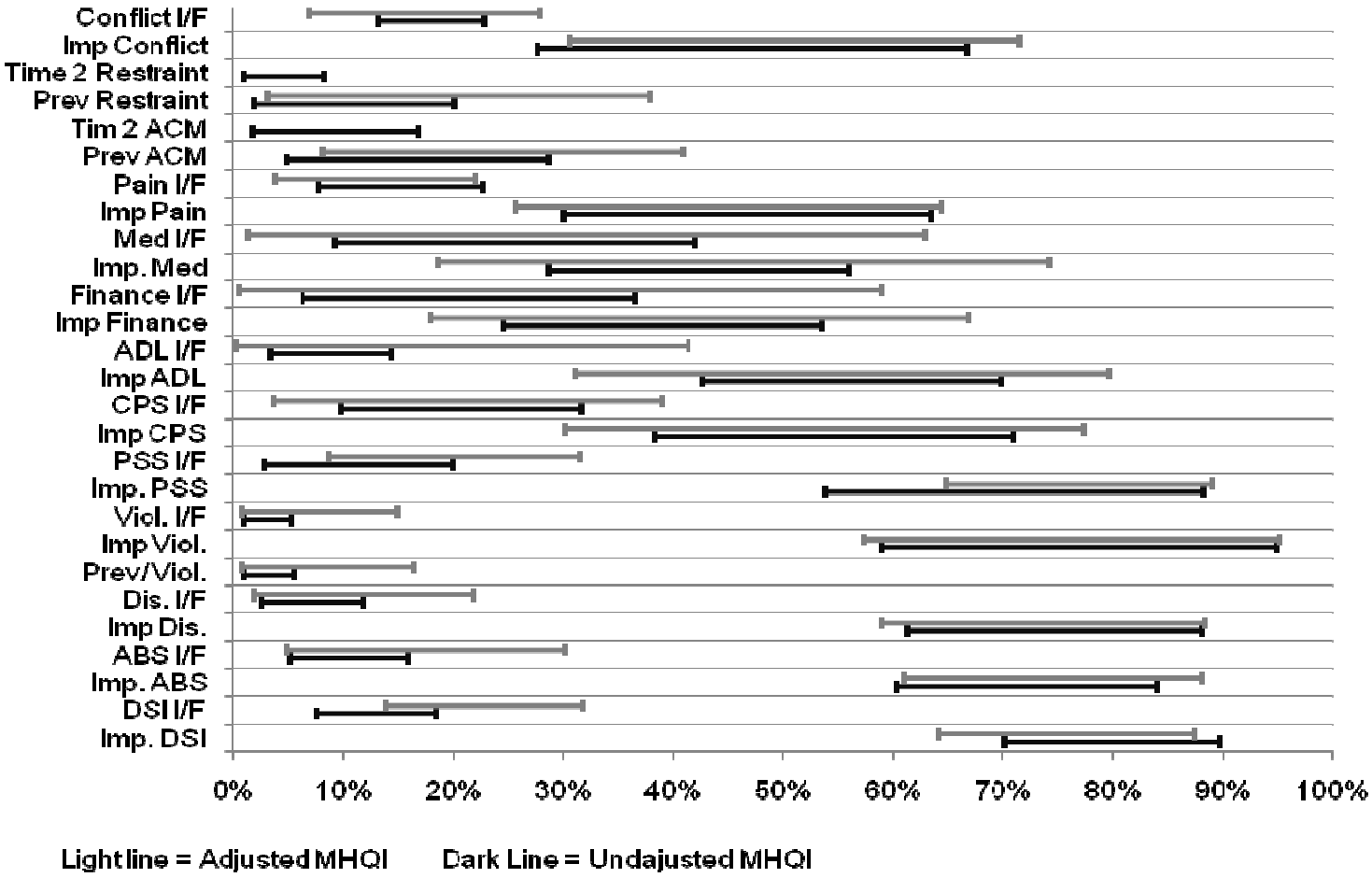


Figure 10. Range of unadjusted and adjusted MHQI Rates among LHINs.

Note: Imp. = improvement, I/F = incidence/fail to improve, Prev = prevalence, Conflict = interpersonal conflict, Restraint = physical restraint, ACM, = acute control medication, Med = medication management, finance = financial management, CPS = cognitive performance scale, PSS = Positive symptom scale, Viol. = violence, Dis = disruptive behaviour, ABS = Aggressive behaviour scale, DSI = depressive symptoms index.

Figure 11 shows the number of LHINs that improved or declined in absolute rank following adjustment of MHQIs. The lowest number of LHINS that changed in rank following adjustment was for interpersonal conflict (MHQI 26), improvement in depressive symptoms (MHQIs 1), prevalence of violence (MHQI 8), and improvement of cognition (MHQI 24). About half of the LHINS improved in rank on the incidence/failure to improve in medication management (MHQI 19) and the prevalence of physical restraint use (MHQI 24). About half of the LHINS declined in rank for the incidence/failure to improve in ADLs (MHQI 15), improvement in financial management (MHQI 16), and incidence/fail to improve in pain (MHQI 21).

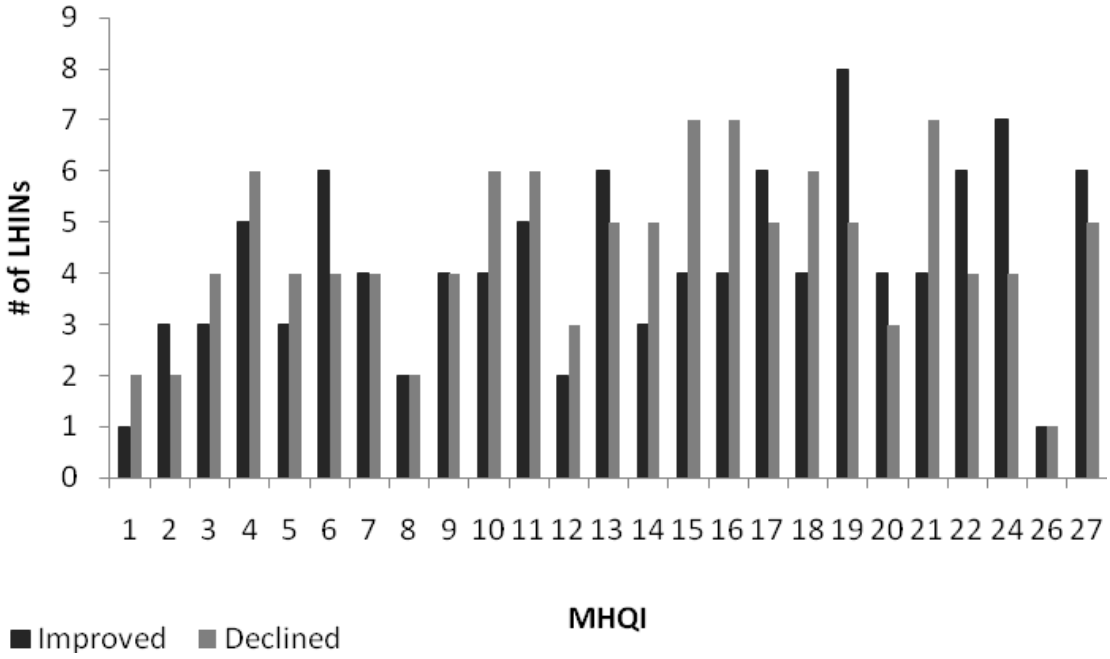


Figure 11. The number of LHINS that improved or declined in absolute rank following adjustment of MHQIs.

4.4.4 Unadjusted and Adjusted EQI Scores among LHINs

Figure 12 shows the median, minimum, and maximum EQI scores among LHINs. Median LHIN scores were almost identical between almost all QI effectiveness scores. For the DSI effectiveness indicator, the adjusted median change was about 1.0 per week between assessments, but ranged from about 0.85 to 1.39. The unadjusted and adjusted EQI scores for the CPS and ADL (both range from 0 to 6) were very similar, with the CPS slightly higher with a slightly larger range. The range of scores among LHINs was slightly less among the DSI and PSS effectiveness indicators and about the same among most other EQIs.

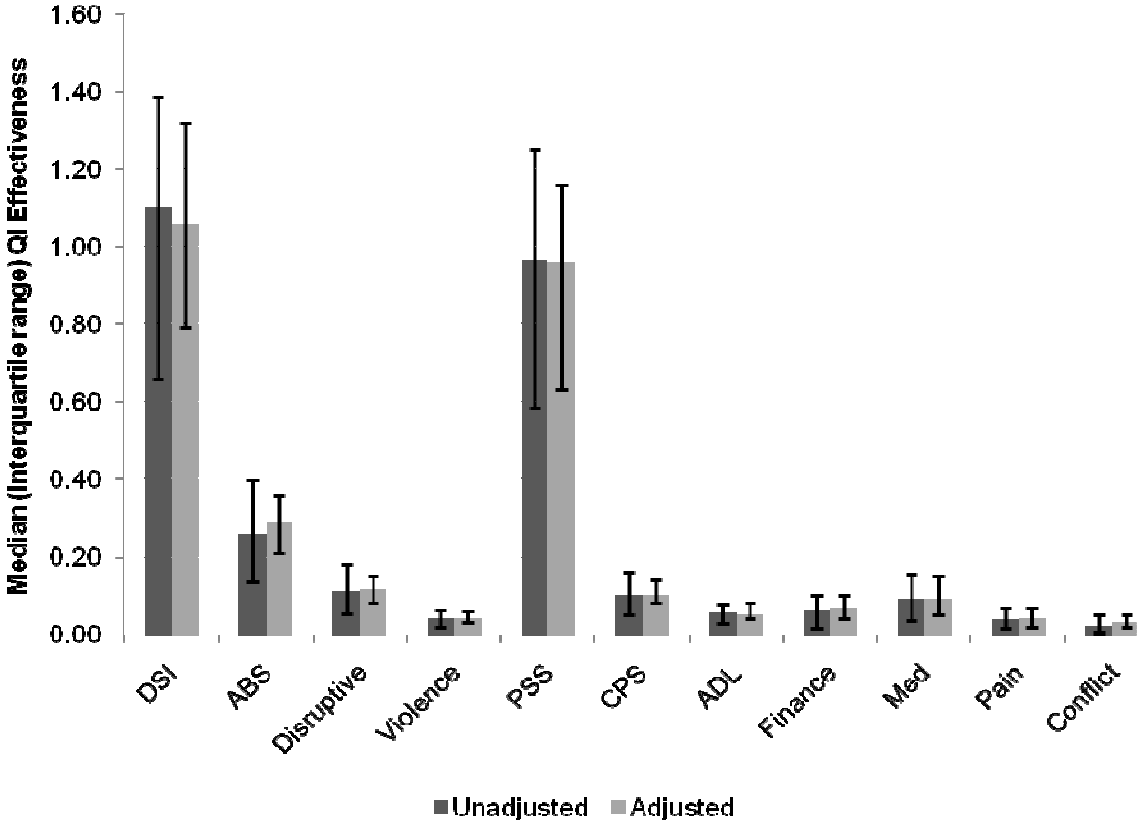


Figure 12. Median and distribution of unadjusted and adjusted EQI Scores among LHINs.

Adjustment had an impact on the rankings of EQI scores among LHINs. Table 19 shows the number of LHINs that changed in absolute and quintile ranks following adjustment of EQI scores. For most EQIs, the number of LHINs that declined in ranks tended to be the same or larger than the number who improved in ranks, with the exception of CPS, ADL, medication management, and conflict QIs (quintile ranked). Interestingly, the larger number of LHINs that changed in absolute ranks occurred for the CPS QI while the largest number that changed in quintile rank occurred for the DSI QI. The largest shift in rank for a single LHIN was by 13 ranks based on adjustment of financial management EQI.

Table 19. Number of LHINs that changed quintile and absolute ranks after adjusting EQIs.

	Quintile Rank		Absolute Rank	
	Improved	Decline	Improved	Decline
DSI	5	5	7	7
ABS	3	3	4	4
Disruptive	4	5	7	5
Violence	3	3	4	7
PSS	5	4	6	6
CPS	4	3	9	7
ADL	5	3	6	4
Finance	3	5	3	9
Meds	4	3	5	4
Pain	2	6	7	7
Conflict	5	3	6	6

5.0 DISCUSSION

This dissertation has developed a set of MHQIs based on the RAI-MH that include outcome QIs, EQIs, and risk adjustment strategies. The derivation of MHQIs based on provincially collected data and feedback from mental health clinical and quality experts supports the feasibility and meaningfulness of the MHQIs. The addition of EQIs provides additional insights into the magnitude of changes that occur among psychiatric inpatients. The availability of risk adjustment techniques for the MHQIs increases the capacity of various stakeholders to make fair comparisons of quality at hospital and regional levels. Before discussing specific implications of this work, a number of findings merit further discussion. The discussions are arranged first for the refinement of MHQIs followed by risk adjustment of the MHQIs, the derivation of EQIs, and regional comparisons of quality. Limitations are then identified followed by opportunities for future research. Finally, implications for clinical, practice, policies, service delivery, and public accountability are discussed.

5.1 Refinement of the MHQIs

Mental health services are challenged with providing care for persons with a diverse array of strengths, preferences, and needs. To optimize accountability and quality improvement, QIs need to reflect this diversity with tangible indicators that promote effective enhancements for recipients of MH services. This study has identified a set of MHQIs that reflect changes to the clinical status of the person following inpatient treatment. The original list of 35 MHQIs was reduced to 27 indicators focusing on symptoms, behaviours, daily functioning, social interactions, and safety. Most of the original MHQI domains were retained because they were identified as being clinically important by mental health stakeholders. However, the remaining MHQIs were revised to improve their utility and measurement properties. The revised MHQI definitions represent a more meaningful approach for quality.

The new MHQIs make several improvements over the initial set of JPPC indicators. First, by removing sentinel events the MQHIs focus on more prevalent mental health outcomes that are likely to

yield more stable estimates of quality. Rare events represented by sentinel indicators are not useful for quality measurement because they may be insensitive to underlying quality problems given that they tend to focus on extreme events. In addition, they do not provide a consistent picture of an organization's quality, making it difficult to determine whether action is required to address a quality concern. Second, the MHQIs provide more reasonable quality expectations for service providers by focusing on improvement rather than complete remission of symptoms, functional status, and behavioural problems. Third, the inclusion of improvement as well as incidence/failure to improve for most domains emphasizes positive achievement while also identifying opportunities for improvement. It also considers both preventive clinical strategies and interventions aimed at remediating existing problems. These distinctions may help facilitate quality improvement by creating incentives to address quality concerns; better scores in one dimension (e.g., more patients improve in aggressive behaviour) can be achieved by addressing the second dimension (e.g., fewer patients experience incidence/failure to improve). This ability to balance outcomes provides structure to quality improvement activities. For instance, initiatives to reduce the incidence/failure to improve in behaviours (aggressive, disruptive, or violent) should not be achieved by an increase in restraint use. Alternatively, some QIs may positively enhance each other. For example, improvement in cognitive performance may also improve opportunities to enhance daily functioning such as the management of finances or medications. While these domains are related, it is still important to have specific MHQIs beyond cognition to promote the importance of daily functioning for the person's overall functioning and independence.

Several indicators were retained even though their rates were below 5% in Ontario hospitals because they were considered sufficiently important clinical indicators that they warranted ongoing monitoring. In addition, it is likely to be the case that these indicators are likely to have rates in excess of 5% in at least some new jurisdictions that are implementing the interRAI-MH or CMH. The prevalence MHQIs for control procedures (acute control medication and physical restraint) were included as measures of patient safety at two points during patients' stay. It is reasonable to expect that certain patients may be more likely to experience control procedures at admission, particularly when

untreated symptoms (e.g., due to psychotic symptoms related to violence) may prevent the effectiveness of less restrictive interventions (e.g., talk down interventions). Such characteristics can be accounted through risk adjustment. However, it is important to ensure that control procedures are not used erroneously amongst large numbers of patients. At follow-up it is less likely for control procedures to occur erroneously, particularly given the variety of standard guidelines in place for their use (Allen, Currier, Hughes et al., 2003). Ideally, all inpatient use of control procedures should be avoided. However, thresholds will need to be established by clinical experts to determine an acceptable prevalence of inpatient use of control procedures.

The exclusion of rehospitalization from the MHQIs was warranted because a definition of rehospitalization based on the RAI-MH could only include instances where a patient returned to the same facility. This MHQI would be relatively uninformative, for instance, in urban regions where multiple inpatient MH units/hospitals exist because the likelihood for return to the same hospital may be different than regions where only one inpatient centre is available. Rehospitalization is a commonly used QI for mental health, although some debate exists as to whether it is an indicator of quality or resource utilization (Rosenheck et al., 1999; Humphreys & Weingardt, 2000; Craig et al., 2000). Exclusion of rehospitalization from the recommended MHQIs does not prevent its use as a surveillance tool for overall measurement of system performance. Other systems such as the Discharge Abstract Database (DAD) collected by CIHI would contain relevant information for measuring rehospitalization.

Almost all MHQIs are derived based on scales or summaries of items rather than changes between single items. The expansion of the hallucinations MHQI to include all positive symptoms improves the sensitivity of the MHQIs to detect changes in all symptoms of psychosis rather than only one. This allows the MHQI to capture a greater amount of variability in positive symptoms than would be possible with a single indicator. For instance, using the hallucinations QI, failure to improve occurs if a person did not experience improvement in hallucinations even if the person experienced improvement in other positive symptoms measured on the PSS (e.g., hyperarousal or abnormal movements). Using a QI based on the PSS, the improvement in hyperarousal/movements would be

identified. Interestingly, the EQI for positive symptoms could then further differentiate that experience of improvement. For two persons with similar lengths of time between assessments, the EQI scores could be higher for a person who experienced improvement in hallucinations and hyperarousal compared to a person who improved in hyperarousal but not hallucinations. Thus, the use of scale based MHQIs is an advantage for measuring patterns and the magnitude of change.

Several MHQIs, however, were more appropriately measured with single items. Financial and medication management were considered separately as single items rather than combining them into a scale in order to preserve the connection between specific interventions aimed at improving financial and medication management as distinct clinical problems. Recall from the introduction that the RAI-MH includes care planning applications called MHAPS (or CAPS). Each of these problems has a MHAP outlining a recommended course of intervention. Other MHAPs are available for the interpersonal conflict, ADL functioning, pain, behaviours, and control procedure QIs. Linking these MHAPs to the MHQIs allows for the linkage between care planning at the individual level and quality measurement at the population level. The RAI-MH MHAPs can be used to identify personal strengths and opportunities for improvement that the care team can build on in collaboration with the person to develop a recovery plan. The MHQIs can then be used to track improvement among clusters of persons who trigger the MHAPs. This linkage between care planning and quality will be useful for promoting engagement in quality measurement and improvement among clinical teams providing inpatient services.

The linkage between care planning and quality is also important to promote individual recovery as persons transition from inpatient settings back to the community. A number of MHQIs promote improvements in domains that will be helpful for persons to function independently once discharged from hospital. The creation of a new MHQI domain for interpersonal conflict expands the scope of QIs to include social functioning and personal relationships. Measures of social functioning have been used elsewhere for outcomes related to substance use conditions as well as mental health (Substance Abuse & Mental Health Services Administration, 2006). Including the interpersonal

conflict QIs will encourage services to help persons improve their ability to cope with conflict, hostility, and enhance skills for effective relationships. The ADL, financial management, and medication management MHQIs enhance the promotion of recovery and independence among persons being discharged from inpatient MH. Helping persons build skills to manage their day to day functioning is a key principle of psychosocial rehabilitation as these skills help the person function independently in the community rather than in more restrictive settings such as inpatient units (Anthony, 1993). Effectively building skills to manage daily actives, improve social functioning, and reduce risk of behaviours along with symptom reduction will help the person integrate into and sustain a more independent life. Therefore, it is important that MHQIs measure successes and promote quality improvement.

The variation in rates of improvement between different MHQI domains, and the variability in these rates between hospitals are notable. Domains that are traditionally considered primary areas of focus for psychiatry such as depression, psychosis, and behaviours had consistently high rates of improvement. Less typical domains such as ADL functioning, financial and medication management, and interpersonal conflict were found to have lower rates of improvement and greater variability among hospitals. While preliminary, these results may be related to differences in standards or practices across hospitals for interventions related to less typical conditions. When considering these results, it is important to note that about 80% of the OMHRS sample consisted of persons considered to be acute patients meaning that they typically had stays of 20 days or less. For these individuals, the focus of intervention is typically crisis stabilization, assessment, treatment of acute symptoms with medications, and discharge back to the community. Interventions for daily or social functioning, for instance, may be built into discharge planning as areas of intervention needed when the person returns to the community. It may be that only long stay patients in hospitals with specialized services receive such interventions. This brings into question the scope of responsibility of inpatient services, particularly at the acute level. Should these services be primarily concerned with stabilization and return to the community as fast as possible or should interventions be considered to help persons enhance their ability to manage their daily lives or social connections? If the introduction of medication

treatments is a primary function of acute services, should those same services not be concerned with how the person is able to manage these medications before leaving the acute setting? While these questions will be discussed further in a later section of opportunities for research, they provide an example of how the MHQIs can be used to promote inquiry into the scope of mental health services delivered in different care settings and not just how they perform at a given point in time.

5.2 Risk Adjustment of MHQIs

The rigor of the MHQIs is enhanced by the availability of extensive patient level information from the RAI-MH for risk adjustment. The risk adjusters identified in this study were more specific and diverse than those typically used in research on QIs for mental health. Due to a lack of clinical data in other studies, risk adjusters are usually limited to demographics such as age, gender, or psychiatric diagnoses (Hermann, 2007). The results from the analyses of RAVs are important for demonstrating the relevance of risk adjustment for measuring and comparing quality in mental health. First, the prevalence distributions of potential RAVs were not equally distributed among hospitals. Instead, hospitals had different patient case mixes that could potentially place them at higher or lower risk for certain outcomes. The unequal distribution of RAVs among hospitals provides evidence that selection bias, although not necessarily intentional, is a potential problem for measurement of mental health services.

Second, this study was able to identify patient level risk adjusters that are meaningfully related to each MHQI. The strengths of the multivariate models varied across MHQIs with the strongest models identified for the ADL, financial and medication management, and restraint use indicators. It was interesting that, although it is a common mental health symptom, the depression MHQIs had the fewest number of RAVs. When examining potential RAVs for depression MHQIs, better fitting models were identified when baseline DSI scores were included in the model. Studies of quality in other health sectors have examined the utility of including facility level scores of QI measures (i.e. Facility Admission Profiles) as risk adjusters but their use did not produce more robust models compared to risk adjustment based on patient information (Morris et al., 2003). In the next section on EQIs a

discussion is provided of the law of initial values as a driver of the relationship between baseline DSI scores and the depression QI domains rather than a reflection of increased risk of a positive or adverse outcome. Therefore, adjusting for baseline score may be more appropriate for measuring QIs where the magnitude of change over time in of interest, such as the EQIs.

Among all MHQIs the most common RAV was the CPS. This is not surprising given the global implications that cognitive impairment has on how a person interprets and responds to information and interventions, particularly if communication is affected. Compared to patients who are cognitively intact, it is reasonable to have different expectations for a good or poor outcome among patients who are unable to understand direction from or effectively convey communicate to clinicians. For financial and medication management QIs, the inclusion of capacity to manage property and manage treatment as RAVs could be viewed as collinear with cognitive impairment. However, their inclusion for these MHQIs is appropriate as they explicitly prevent improvement by nature of their definitions. Overall, inclusion of these measures, as well as the CPS, schizophrenia, and age did substantially affect facility scores on these MHQIs; however, there was still a large amount of variation between hospitals' scores indicating that over-adjustment did not occur.

Persons with a mood or substance related disorder were less likely to experience incidence/failure to improve in cognition. It may be that persons with mood or substance conditions experience difficulty with cognitive functioning as a result of their mood condition or substance use, and not because of an inherent cognitive impairment. Treatment of depressive symptoms, for instance, has been shown to improve cognitive functioning in persons with traumatic brain injury (Fann, Uomoto, & Katon, 2001). Similarly, it was not surprising that patients with a mood disorder were also less likely to experience an incidence/ failure to improve in aggressive behaviour given that mood disorders include lack of motivation and depressive symptoms. Inclusion of mood disorders in risk adjustment for aggressive behaviour MHQIs means that hospitals who admit a high number of persons with mood disorders would be expected to have a lower incidence/failure to improve in aggressive behaviour. However, mood disorders also include bipolar disorder which is characterized by periods of mania.

Mania symptoms were included as a RAV for six MHQI domains including aggressive behaviour and cognition. Persons with symptoms of mania may suffer from impaired attention and difficulties with decision making that may hinder their likelihood for improvement in cognition and are more likely to express aggressive behaviour given symptoms such as anger, impulsivity, and frustration (Rossi, Daneluzzo, Arduini, Domenico, Stratta & Petruzzi, 2000)

The third way that the study results demonstrate the importance of risk adjusting MHQIs is the impact risk adjustment had on MHQI scores among hospitals. Similar distributions of MHQI scores existed for both unadjusted and adjusted results; however, the order to which facilities fall within this distribution changed when patient case mix was considered. The percentage of hospitals that changed in absolute rank was greater than 80% for all but one MHQI. Absolute changes in rank are difficult to interpret as small alterations in MHQIs scores could affect a hospital's rank. Quintile rankings are more meaningful for gauging the impact of risk adjustment because they apply a degree of magnitude to the change. Changes in quintile ranks were observed after adjustment of all MHQIs. Even for the addition of one risk adjuster had an impact on how hospitals compared on MHQI scores. The changes in quintile rankings show how conclusions about quality differ when patient case mix is considered.

5.3 Development of Effectiveness Quality Indicators

The EQIs add new dimensions to the evaluation of quality of mental health services. Although research on the effectiveness of psychiatric interventions has existed for some time (e.g., Schinar, Kamis-Gould, Delucia, & Rothbard, 1990), the use of EQIs to compare service providers has not been previously explored.

At a hospital level, EQIs describe the average amount of improvement in an indicator per seven days between assessments. This study evaluated different definitions of inclusion criteria for denominator groups that produced higher or lower scores. While exclusion of those who had baseline measure scores of zero produced higher scores that would appear more interpretable, the scores may not be appropriate representations of actual effectiveness. Among all measures used to measure MHQIs,

1% to 22% of those who scored zero at baseline experienced an incidence in the measure score. Excluding QI measure scores of zero at the initial assessment would have been misleading because those who experienced incidence in the measure at follow-up would be excluded. Therefore, EQIs included all cases in the denominator regardless of the initial score. If higher scores are deemed more meaningful, an alternative approach could have excluded persons who scored 0 at both assessments (i.e., never experienced the QI event) from the denominator. While this would create seemingly more interpretable scores (more pronounced from 0), the interpretation of the denominator and the score itself may be misleading. In public reports, for instance, a transformed score based on a restricted denominator could be generalized to the experience of all persons receiving inpatient services.

The goal of EQIs is to encourage successful and appropriate use of time during inpatient services, and not necessarily shorter lengths of stay. The expectation for each EQI is for steady improvement to occur across a persons' episode of care. It would not be appropriate to penalize a hospital for increasing the length of stay for persons who have not achieved an appropriate degree of improvement for discharge. For certain conditions, it may be expected that treatments take longer to be effective and less improvement or decline may even occur in shorter lengths of stay. For instance, it may be that interpersonal conflict fails to improve or declines over certain periods during an admission among persons who are detoxifying from substance use or become cognizant of different emotions, particularly among persons recovering from post traumatic stress. In these situations, it may be reasonable to expect longer lengths of stay to achieve improvement because interventions for interpersonal conflict may be more intensive involving individual and group therapies. On the other hand, symptoms of depression or psychosis may be resolved sooner in the therapeutic process, often through the introduction of psychotropic medications.

Regardless of the definition of the denominator scores cannot reach 1.0 on a number of EQIs because the ceiling scores for the QI measures are not equal to or greater than 7, the size of the denominator. This applies to the pain, interpersonal conflict, violence, CPS, ADL, financial management, and medication management indicators. Although low, the actual scores for the EQIs are

not as meaningful for understanding quality as the comparison of scores. This will be discussed further in later sections of the discussion.

All EQIs are based on absolute change between EQI measure scores taken at different times. For some measures, such as the CPS or ADL hierarchy, a change from one score to another may represent a more meaningful shift than a change in continuous scales such as the DSI. For instance, a CPS score of 2 indicates mild impairment while a score of 1 indicates the person has borderline intact cognition. For longer scales, such as the DSI, higher scores may also be more likely to improve simply by chance (i.e., law of initial values) or due to the responsiveness of more severe symptoms to interventions (Jin, 1992). The current study attempted to adjust for the likelihood for change among different baseline scores of EQI measures by including the initial assessment scores in regression adjustment. Studies using the FIM indicator to compare rehabilitation services have also included baseline score in risk adjustment for similar purposes (Woo et al. 2008). The hybrid approach to risk adjustment developed for nursing home QIs that includes stratification and covariate regression adjustment may also be applicable for EQI scores (Jones et al., in press). The hybrid method could create quintile strata using baseline measure scores and calculate adjusted EQI scores by using regression adjustment of covariates within each stratum.

Evaluation of EQIs will also need to consider the structure of inpatient mental health services. Several streams of inpatient services may inherently influence EQIs because of the lengths of stay associated with these service types. For instance, some specialized programs are considered long stay programs because their lengths of stay are fixed and longer than 15 days. For some of these programs, patients receive a set menu of programs and services regardless of their level of need. Forensic programs may have longer lengths of stay due to the nature of conditions imposed on the person by the court system. In the current study, clear differences in EQI were found between several patient types, including forensic patients. Therefore, reporting EQI scores, even after adjustment, should consider stratifications by patient type. Such stratifications could drive inquiry into how changes to service

structures could affect the experiences of those receiving services. If new service procedures were implemented for forensic mental health, the stratum specific EQIs may be sensitive to such changes.

5.4 Limitations of MHQI, Risk Adjustment, and EQI Results

There were several limitations with the current stage of development of the MHQIs. First, the Pilot data were drawn from only 7 facilities that were using the RAI-MH prior to the provincial mandate; however, the availability of OMHRS data made it possible to replicate the findings guiding initial MHQI selection. Second, the MHQIs were derived from data that excluded patients with stays of less than 6 days or who had only one assessment available. Establishing outcome MHQIs for these short-stay patients will be challenging and may be limited to prevalence indicators of events such as self harm, harm to others, and control procedures. These prevalence indicators could be useful for identifying the percentages of patients discharged at high risk for self harm, inability to care for self, or harm to others; however, conclusions about changes to their clinical status are not currently possible.

Third, several MHQIs had to be excluded due to unavailable data for medication use. Even though all interRAI instruments include detailed sections on medication use, the OMHRS data requirements do not include mandated submission of medication data. Given the importance of pharmaceutical therapies as part of psychiatric services, the lack of these data is an important limitation of the OMHRS data as it places a constraint on the capacity to make inferences about the underlying causes of apparent quality differences between organizations.

Fourth, the current MHQIs were measured based on changes between admission and discharge assessments. The use of admission to discharge might not be sensitive to information that occurs between admission and discharge. For instance, aggressive or disruptive behaviour or violent events that occur between assessments would be missed. The RAI-MH, however, is intended to be completed at admission, quarterly, discharge and anytime there is a change in a patients' status (significant event or change in care needs). Therefore, ideally, any inpatient behavioural event should trigger a change in

status assessment to review changes that may be relevant to the event thus increasing sensitivity to such events during quality measurement.

Fifth, risk adjusters for all MHQIs were chosen using incidence/failure to improve as the outcome. The rationale for this approach was that variables found to increase the odds of incidence/failure to improve could be assumed to decrease the odds of improvement. However, the impact of risk adjustment on MHQIs seemed to be greater among the incidence/failure to improve MHQIs than improvement MHQIs. It may be that further investigations into risk adjustment for the MHQIs could identify additional variables for inclusion in adjustment of improvement indicators; however, these investigations will be challenged by their ability to provide explanations for why the presence of a risk adjuster for improvement would not be an adjuster for incidence/failure to improve, particularly for public reporting.

Finally, the evaluation of EQIs between hospitals and LHINs did not consider variables that are related to time between assessments but may not be or are concurrently related to the MHQI measures. For instance, involuntary admission status may influence patients' LOS regardless of symptoms, behaviours or other personal characteristics; however, the status may also be under the control of psychiatrists who admit patients involuntarily under application for psychiatric assessments or other involuntary admission practices. Since these practices are not independent from the service setting it is not appropriate to include in risk adjustment. Further research could, however, identify specific patient characteristics that significantly differentiate involuntary and voluntary admission status. The present results established that risk adjustment of EQIs is needed and does influence comparisons; however, further in-depth analysis of risk adjusters for EQIs is clearly needed.

5.5 Opportunities for Future Research on MHQIs, EQIs, and Risk Adjustment

The current research presents a number of opportunities for further research on the MHQIs, effectiveness indicators, risk adjustment, and their use for comparisons at different levels of mental health services. First, further validation of the MHQIs should be evaluated by examining the

relationships between the MHQIs and other indicators of technical quality. These investigations could use surveys of mental health hospitals to determine if the MHQIs are sensitive to changes in practice such as the implementation of new services or policies (e.g., reduction in ACM use following restraint reduction programs). Facility surveys should include considerations for management practices, medication use, staffing, and programming based on best practices identified in the literature as well as clinical experts. In these evaluations, it will also be interesting to determine how outcomes measured by the MHQIs are related to specific processes such as receipt of interventions for social functioning, family support, pain management, and community reintegration. These comparisons could be made using RAI-MH data since several sections on the RAI-MH examine service utilization and receipt of interventions. Understanding how differences in the process of care relate to outcomes will be important for further validating the MHQIs and for understanding how policies and practices influence outcomes of mental health services.

In addition to the provincial implementation of the RAI-MH in Ontario, pilot projects and regional implementations have also occurred in three other Canadian provinces, two US states, Iceland, Finland, Chile, Taiwan, France, Switzerland, and Spain. Therefore, there is opportunity for replication and further validation of the MHQIs in other health regions and systems once sufficient data become available. These activities could lead to comparisons of mental health outcomes regionally and internationally, adding to other activities in place for international benchmarking of mental health quality (Hermann et al., 2006). International evaluation of MHQIs and the possible association between MHQIs and differences in governance, management, and delivery of mental health services will be interesting for informing how mental health services can best be organize to optimize improvement and to prevent adverse outcomes.

Second, further research is needed into the patterns of rates identified between types of MHQIs and the distribution of rates among hospitals and regions. There was substantial variability, even after adjustment, in hospital rates for most MHQIs, particularly among MHQIs measuring changes in cognition, ADL functioning, and behaviours. It is interesting to note that there was less variability for

MHQIs measuring change in depressive symptoms and psychoses. It would be interesting to evaluate this further by comparing how MHQI rates relate to the availability and use of standard guidelines for the treatment of "typical" psychiatric symptoms such as depression or psychosis compared to those for cognition, daily functioning, and behaviours. Perhaps practice patterns for functioning are not uniformly implemented compared to practices to improve depressive or psychotic symptoms. These investigations will further demonstrate the utility of MHQIs to detect differences in specific domains of quality across inpatient facilities highlighting opportunities for quality improvement. These differences also point to the need to examine alternative stratification methodologies to control for population differences that are not fully accounted for by the covariate adjustment strategy used here (e.g., forensic status, geriatric patients).

Third, investigations into the importance of the MHQIs among providers and recipients of mental health services will help determine priority areas for reporting and quality improvement. While the complete set of MHQIs may be important to understanding the quality of mental health services, certain domains may be more important to the recipients of services compared to service providers. Surveys or focus groups could be used to gather preferences for MHQIs by having participants rate QIs based on whether they are reflections of service quality and their importance for sustaining recovery and independence. Organizing the MHQIs based on the preferences of different stakeholders will help public reporting of MHQIs by highlighting issues that are central to those who deliver and receive services. Understanding the divergence and similarities of outcome expectations among the recipients and providers of mental health care will also be important for aligning the delivery of services that will best meet patient needs.

Fourth, further investigations into the EQIs should explore the issue of relative change for defining effectiveness as well as other methods of controlling for likelihood of change using baseline scores. The current study examined variations in absolute change over time as EQIs among different hospitals and regions. Effectiveness indicator scores could also be expressed in terms of relative change over time. Typically, relative change is defined as the ratio of a time 1 score divided by time 2

score. Multiplied by 100, the relative change can be expressed as a percentage change in a QI measure score. Effectiveness indicators expressed in relative terms (e.g., percentage change per week) would standardize the metric used for all EQIs and allow for within group comparisons (e.g., which EQI domain does a hospital perform best?). Use of relative change may also have implications for the interpretation of EQIs. For example, being able to state that a condition improved by 30% per week is, potentially, more meaningful than saying a score improved by 0.12 per week. A prior survey of physicians found treatment decisions by physicians were influenced when the effectiveness of an intervention was expressed in absolute versus relative terms (Forrow, Taylor, & Arnold, 1992). More physicians decided that they would use a treatment when the outcome of that treatment was expressed in relative versus absolute change. Therefore, further investigations into the use of relative change for EQIs may have important implications on decisions about mental health policies and practices.

Fifth, further research is needed to determine the appropriate risk adjustment method for the MHQIs. This study used methods similar to indirect standardization to adjust the MHQIs for comparing hospitals. Indirect standardization is the most common approach for risk adjustment of QIs in all health sectors, including mental health (Hendryx & Teague, 2001). For the MHQIs, an advantage of indirect adjustment using regression is the relative ease with which multiple RAVs can be included. Direct standardization using stratification or stratification weighting may not be possible for MHQIs such as aggressive behaviour where 8 RAVs are included as strata would include very small sample sizes. However, newer methods such as hybrid adjustment or stratification on the propensity score may be appropriate for use with the MHQIs, particularly given the large variability in MHQI scores across OMHRS hospitals. Hybrid adjustment proposed by Jones and colleagues (in press) uses stratification based on a risk variable highly correlated with the QI and then performs regression adjustment within risk strata based on other covariates. Propensity score adjustment combines multiple risk adjusters into a single score allowing for simple stratification of multiple risk adjusters (Huang et al., 2009). The ability to balance risk adjusters is an advantage of using propensity scores, and direct adjustment in general, over indirect adjustment. Regression models provide estimates of the impact of treatment on

outcome even if hospitals being compared have somewhat different distributions of risk (Shahian & Normand, 2008). Therefore, hospitals that treat patients with different risk profiles are still compared. With stratification and propensity score adjustment hospitals are compared within strata consisting of overlapping patient case mix.

Direct adjustment also allows for variation in the effect of different levels of risk adjusters on QI scores among hospitals (Zaslavsky, 2001). For instance, it may be that a larger difference between the depression effectiveness indicator scores exists among patients with lower and higher baseline DSI scores in hospital A while a smaller difference exists in hospital B. Therefore, differences between the two hospitals on low baseline scores might not be the same as the difference between these hospitals on high baseline scores. Regression adjustment alone would not detect these differences without the inclusion of a number of interaction terms. Interaction terms may be more cumbersome and less informative in reporting than stratification. Since the purpose of risk adjustment for this study was to demonstrate the need for and impact of risk adjusting the MHQIs, regression based adjustment was suitable. However, future research should examine the application of direct adjustment using propensity scores or hybrid adjustment for MHQIs and EQIs.

Sixth, further empirical investigations into the comparison of MHQIs and EQIs among hospitals and regions are needed. The current results showed that there were differences in MHQI and EQI scores among hospitals and regions, but didn't evaluate the magnitude of impact hospitals or regions had on differences in quality. Analyses using hierarchical regression models could examine the impact of hospital, region, or both on quality scores while accounting for patient level risk adjusters. These evaluations will be interesting because they can identify how hospitals might differ in MHQI scores after controlling for region and patient characteristics, or vice versa. These analyses can begin to identify, for instance, the impact that region to which a patient resides has on their likelihood for certain outcomes. These analyses may be more useful than simple comparisons of the rankings of hospitals because they begin to identify the magnitude of variance in quality that can be attributed to difference sources

Finally, evaluations will be needed on the impact of implementing the MHQIs and EQIs in public reporting. In the years following implementation it would interesting to examine shifts in practices and policies as well as rates of improvement in quality scores over time that could be related to the implementation of the MHQIs. Possibly due to the relatively recent emergence of quality measurement initiatives in mental health, little research has been done to actually explore the impact that reporting quality has on changes to health service delivery, practices, or governance (Fung, Lim, Mattke, Damberg, & Shekelle). This type of evaluation could identify further opportunities to improve reports of quality; but, more importantly, these evaluations will speak to the value of measuring and reporting quality for changing how mental health services are managed and delivered.

5.6 Implications for Regional Comparisons using the MHQIs and EQIs

Among the 14 LHINs compared in this study, there was significant variation in the number of inpatient mental health hospitals and characteristics of patients admitted to those hospitals. This variation supports the regional management of health services so that issues and needs that may be specific to the region can be addressed. Some LHINs include large geographic regions consisting of mostly rural dwelling populations; others are geographically small and include dense urban populations. These regional characteristics may produce specific challenges to mental health services such as homelessness in urban regions and access to services in rural regions. The LHINs were designed to manage these challenges by setting region specific priorities for the allocation of funds.

The MHQIs are useful at a regional level for examining the impact of region specific initiatives, and differences between these initiatives among LHINs, on inpatient outcomes. The results identified substantial variation between LHINs, even after adjustment for patient risk factors that were not equally distributed among LHINs. Of particular interest was the variation in outcomes for cognition, ADLs, financial and medication management, conflict, and restraint use compared to behaviours, depressive, and positive symptoms. It was encouraging, for instance, to find that the majority of patients with aggressive behaviour improved within each LHIN. However, prevalence for acute control

medication and physical restraint use were also quite high in some LHINs (almost 40%). Further analysis of the MHQIs looking at the prevalence of ACM and restraint use among LHINS with the highest rates of improvement in aggressive behaviour may be interesting to determine if positive behaviour outcomes are being achieved with more restrictive practices. It would also be useful to examine why outcomes related to functioning, cognition, and conflict showed more variation compared to changes in depressive or positive symptoms. For example, it may be that there are regional differences in occupational therapy programming and hiring that relate to the variation in cognitive and functional MHQIs. Recalling from the introduction, LHINs that identified mental health service as a priority did differ in service use compared to non-priority LHINs (Martin & Hirdes, 2008). It will be interesting to examine whether hospitals with low MHQI scores are clustered in LHINs and to determine if this clustering is related to priorities established by LHINs. Stemming from such investigations the LHINs may be able to better implement priority areas for service plans of mental health hospitals using the MHQIs to inform and evaluate these priorities. Such queries exemplify the relevance of the MHQIs for driving inquiry and quality improvement of mental health services at a regional level.

LHINs also showed variation in the rates of EQIs prior to and following patient level risk adjustment. This variability may be attributable to different characteristics of the services offered in LHINs and certain characteristics of LHIN regions. For instance, regions include specialized psychiatric hospitals may have higher EQI scores because most long stay programs are offered in specialized hospitals. Regions with psychiatric units in general hospitals are likely to have shorter lengths of stay and may have higher EQI scores. Therefore, further investigations into LHIN differences in EQIs should stratify results by hospital type to determine if differences are related to the types of services available in each LHIN.

In terms of region characteristics, effectiveness indicators may be useful for identifying regional variations in factors that facilitate discharge beyond patient characteristics. Urban areas may have a multitude of community support services driving shorter LOS's that focus on crisis stabilization,

assessment, and community reintegration. LHINs with large geographic regions may keep patients longer if rural community support services are unavailable. Research could use the EQIs to evaluate innovative health services for rural regions. For instance, do regions who implement tele-psychiatry or video consultation services have better EQI scores than rural regions that do not use such technologies? Also, further research could use the MHQIs and EQIs to examine how urban versus rural issues (e.g., population density, distance from services, availability of ACTT, etc.) affect quality of mental health services.

Finally, the MHQIs present an opportunity for linking quality and accountability of funding and governance of health services. All LHINs are responsible for establishing priorities for improvement in the integration and delivery of health services. LHINs were established with the expectation that they would address accountability for hospitals by improving engagement and transparency of governance based on discussion of regional priorities (Reeleder, Goel, Singer, & Martin, 2008). The quality measurement system based on the RAI-MH can act as a vehicle for sharing common information, identifying priority areas for mental health services, and evaluating the impact of accountability agreements. The diverse array of MHQI domains and effectiveness can drive investigations into how funding of strategic initiatives, and mental health services in general, within and between LHINs is related to the quality of mental health services within each LHIN.

5.7 Applications of MHQIs across Mental Health and other Health Sectors

There are opportunities to expand the use of the RAI-MH MHQIs to assess quality as persons move through different mental health sectors, between inpatient and community mental health and beyond. All interRAI instruments include core items that are consistent across all assessments as well as items that are sector specific. The interRAI Community Mental Health (CMH), for instance, contains 60% of the items used in the RAI-MH. In fact, only the control procedure MHQIs cannot be measured using the interRAI-CMH. Therefore, 23 MHQIs could be used to evaluate community mental health services where the interRAI CMH is available. Indicators can also be developed

specifically for CMH settings. For instance, the interRAI CMH includes more information than the RAI-MH for measuring the incidence of criminal activity and the incidence of substance use. Also, QIs for CMH could focus on enhancements to role functioning (e.g. incidence of supported or full employment), social functioning, meaningful activities (e.g. incidence of activity levels or volunteering) and independence (e.g., improvement in capacity to manage meals, shopping, and transportation). When data become available, procedures used in this study to develop the MHQIs could be applied to evaluate and develop MHQIs based on the interRAI CMH.

Measuring quality in CMH settings will not be without specific challenges. Inpatient settings are, to a certain extent, controlled environments where patients are monitored closely in structured activities. In the community a multitude of exogenous factors may further influence a service provider's likelihood for good or bad quality. The availability of informal supports, income, employment, social networks, housing, and substance use may all influence outcomes regardless of service interventions or supports. Also, the nature in which QIs are defined in the community may be different. It will be important to determine, for instance, when reasonable expectations for change should occur. Compared to inpatient mental health, changes may occur over longer periods of time if persons are not in stages of crisis or experiencing acute symptoms, particularly if QIs are chosen to focus on improvements to functioning, well being, social networks, or participation in activities. The complexity and fragmentation of CMH services will also pose a challenge to quality measurement. Some services may be connected with hospital based services while others are stand-alone agencies, both of which could be managed and governed by different bodies. Services may also vary in the implementation of specialized services such as ACTTs or supported employment. Therefore, gaining consensus on expectations for the quality of mental health services will be difficult due to fragmented purposes and management of CMH services. For instance, it may not be meaningful to compare CMH services that include specialized services such as ACTT services with services that provide general case management. The persons who receive these services may have very different needs and expectations about what these services should achieve. While both are expected to abide by the

standards established in mental health acts and accountability agreements their inherent structures and goals for services may make quality comparisons more complex. In these situations, it may be more meaningful to report MHQIs stratified by service types.

The use of MHQIs between inpatient and community mental health services creates an opportunity to evaluate quality at a system level. Given that a persons' first point of contact with the mental health sector is usually CMH, evaluating quality of mental health services would begin with the initial interRAI CMH assessment. Then, as the person moves through mental health services additional assessments done within CMH or inpatient services will serve as points at which outcomes can be assessed. For instance, for patients discharged from inpatient to community services, completion of the interRAI CMH 30 days after discharge could serve as a third follow-up assessment for inpatient MHQIs and a baseline assessment for community MHQIs. With this approach, conclusions about quality can be made as persons transition between mental health sectors. Analyzing outcomes through these transitions may provide insight into the impact of specific sectors on various mental health outcomes. For instance, it might be expected that changes to functioning and social relationships will be identified as a person transitions through CMH while shifts in specific symptoms occur as a person moves through inpatient services.

While ideal, mental health system level use of MHQIs will be highly complex. Research will be needed, for instance, to determine how to perform risk adjustment when looking at transitions between sectors. The point at which a risk factor could be considered an exogenous factor versus a contextual factor, for instance, is not clear when outcomes involve assessments taken in different settings. In this sense, risk factors such as age or diagnoses may be more appropriate than the level of symptoms at a given time. If one year outcomes of the mental health system were of interest (e.g., improvement in depressive symptoms over one year of MH service use) the first initial assessment (either the inpatient or CMH) could potentially be used for risk adjustment. A second problem would be the right censoring of persons as they move out of the mental health system and into other health or social services sectors (e.g., LTC) or the judicial system (e.g., prison). In these instances, results can

include the person's final assessment as an endpoint to capture their system level outcome. Under these circumstances, however, it will be difficult to determine the ongoing outcomes of a person's mental health condition once they leave the mental health sector. For long term care, chronic care, and home care as well as the prison system other interRAI tools are available that could address certain MHQIs (e.g., depressive symptoms). To achieve such integration for MHQI evaluations, however, further linkages of data and health systems would be needed so that persons' outcomes can be tracked as they move through the health system.

The use of MHQIs within MH services and at a MH system level presents an opportunity to introduce a common mechanism for linking service delivery and accountability within and between inpatient and community mental health. Fragmentation of community mental health services is often due to the non-uniform implementation of evidence-based practices such as ACTT for persons with severe mental illness, particularly when regionally managed MH services lack the technical expertise to properly administer, manage, and evaluate such services (Latimer, 2005). The MHQIs including EQIs and risk adjustment represent a common approach that could be included in evaluating accountability among and between inpatient and CMH settings between regions such as LHINs. Since the MHQIs are linked to individual characteristics and care planning, consistent methods of service planning and evaluation can be implemented across sectors. This common approach to quality and accountability could help prevent and repair fragmentation by using common information from the point of care in both inpatient and CMH settings to the point of evaluation and policy development within and between service providers, LHINS, and Provincially.

5.8 Implications for Recipients of Mental Health Services

At the heart of quality measurement should be the inherent benefit of evaluation for the recipients and users of mental health services. In mental health, a variety of structure and process based indicators already in place allow evaluators to identify whether or not services are in place, acceissible, and appropriately used. The MHQIs extend the utility of quality measurement for MH service

recipients by providing information on the impact of these services in improving the lives of those who receive them. From a safety perspective, the balance between MHQIs that measure control procedures and MHQIs for behaviours and violence may promote alternative interventions for preventing violence or aggression rather than simply controlling it. The use of the Harm to Others and Control Procedures CAPs in services that are identified to have high rates of aggressive or violent behaviours may help inform these interventions.

While the MHQIs include important domains for changes in symptoms or behaviours that cause distress, they also include other important aspects of functioning. The inclusion of measures for ADL functioning, management of finances and medications, interpersonal conflict, and pain identifies the importance of these issues for promoting independence and community reintegration. Rather than hospitals focusing simply on psychiatric symptom reduction, the MHQIs promote enhancements to the capacity of individuals to manage their daily lives. Improvements in social functioning and behaviours could help persons build better social connections and prevent adverse experiences such as social isolation and police interventions.

The MHQIs are also important to service recipients because they promote improvement, rather than simply monitoring adverse outcomes. The concept of promoting improvement is important at an individual level because it focuses on the development of strengths rather than aspects of decline. Being able to identify how mental health services help improve the lives of service recipients is important for instilling hope in those who need services. It might also be important for removing stigma from the public's view of mental illness. The MHQIs can help Identify to the public that mental illnesses are conditions that can improve and that persons can experience recovery in domains beyond symptoms and behaviours. This could have benefits for community reintegration, social services, and employment. The concept of improvement is new for quality measurement and is essential in promoting management and clinical practices that reinforce recovery.

The application of risk adjustment is also important for promoting fair access to services among persons with severe and chronic conditions. Risk adjustment prevents intentional selection bias

of potentially "hard-to-treat" patients such as those with personality disorders. Without risk adjustment, services could select patients who would have the best opportunity for good outcomes or are not at risk for adverse events. Instead, risk adjustment changes the outcome expectation for services who admit difficult to treat or high risk patients allowing for more equitable comparisons between services. While indirect, the use of risk adjustment is important for ensuring that quality measurement does not result in poor quality practices such as risk selection.

5.9 Implications for Public Reporting and Accountability

The MHQIs will be valuable for a number of quality reporting initiatives within Ontario. Use of the MHQIs in reporting systems will create opportunities to share best practices among providers and the public. The implementation of the MHQIs into CIHI's mental health reporting system will provide a mechanism for sharing MHQI results among all Ontario hospitals with inpatient mental health beds. In addition to reports, the OMHRS team at CIHI holds quarterly teleconferences available to all hospitals that submit RAI-MH data. Therefore, the teleconferences could serve as a forum for education about the MHQIs and their interpretation as well as identifying and sharing best practices based on MHQI results.

The MHQIs will also be relevant for the Ontario Health Quality Council (OHQC). The OHQC is the primary organization for reporting health care quality in Ontario. To date, the OHQC report cards and special investigations have examined patient safety, acute care, long term care, and home care. The OHQC reports typically include indicators of access, system integration, appropriateness, and outcomes. No reports have been produced, to date, that specifically examine mental health services. For mental health, the OHQC could draw on a variety of structure and process information and indicators that are available from prior reporting frameworks such as the OHA Hospital Reports. The MHQIs, EQIs, risk adjustment, and their applications at hospital and regional levels will be useful for the OHQC to implement reporting that includes clinical outcomes. These outcomes can be linked to

structure and process indicators at regional levels to determine how regional differences might affect outcomes.

Public reports will need to consider whether QIs should be reported separately or combined into single indicators of quality. For instance, in the U.S. the Centers for Medicare and Medicaid Services (CMS) combine QIs into a 5-star rating system for nursing home care (CMS, 2009). While a composite quality score is easier to interpret for public reporting, it may be inaccurate for describing the actual quality of health care. Like health care itself, quality of health care is multidimensional encompassing different domains of patient care and patient outcome. As such, some facilities may perform better on certain domains of quality but not others. Combining QIs into composite measures of quality may result in neutral descriptions of quality where hospital effects could be masked (Mor, 2005). Instead, multiple QIs are helpful for identifying domains where services excel and others where improvement is needed.

Reporting MHQIs and other QIs for mental health will need to be sensitive to the stakeholders of reports. To maximize the potential impact of quality reports for quality improvement and accountability risks to misinterpretation of quality results should be minimized. Implementing MHQIs into public reports will, inevitably, trigger discussion about processes or structures that may influence MHQI outcomes. Therefore, reports on quality need to include structure, process, and outcome indicators making linkages between these indicators whenever possible. In presenting these results, careful consideration will be needed to present MHQIs in ways that are meaningful to more stakeholders than only experts in quality measurement. While technical appendices are essential to ensure transparency in how MHQIs are scored, plain language explanations of QI calculation, risk adjustment, and comparisons will be needed. Such explanations should state that MHQIs represent markers for different patient experiences with mental health services, not necessarily final conclusions about overall quality of care. In this sense, combining MHQI scores into a single indicator may actually produce less meaningful indications about quality of care. For instance, combining indicators into a single score based on the results presented in the present studies would have masked findings

suggesting that traditional psychiatric outcomes such as improvement in depressive symptoms are higher and less diverse than less traditional outcomes such as improvement in medication management. The use of multiple indicators, instead, allows reports on quality to identify specific areas when services excel and opportunities for further inquiry or improvement.

Implications for publicly reporting risk adjusted scores will also need to find ways to provide simple explanations for what is an extremely complex concept. There is concern that misunderstanding of indicators, their definition, and interpretation may be unfair to hospitals being compared and the general public interested in results (Wallace, Teare, Verrall, & Chan, 2007). For less technical stakeholders it will be important to ensure that key aspects of risk adjustment are conveyed. Essentially, these reports should explain that MHQI scores are adjusted for the individual characteristics that persons bring with them into care because these characteristics have an influence on care outcomes regardless of the interventions the person receives. Since different care settings being compared accept different numbers of patients who have these characteristics we need to adjust MHQI scores to provide fairer comparisons of quality. Thus, an adjusted MHQI score represents the service providers' score if the organization accepted a patient with average characteristics, or risks. These rather simple explanations will provide some insight into risk adjustment for non-technical stakeholders of MHQI reports; however, it is recommended that further education with detailed examples be provided to ensure that the concept and results from risk adjusted MHQI reports are not misunderstood.

Finally, public reports on MHQIs and EQIs will need to emphasize that the QIs are meaningful for quality comparisons as well as quality improvement. For quality comparisons, the MHQI or EQI scores, themselves, are not meaningful unless compared with a reference group or standard (Jones et al., in press). Thresholds of acceptable scores such as the number of standard deviations from the mean, percentile ranks, or quintile ranks are typically used. In rehabilitation medicine in the U.S., a facility's raw Functional Independence Measure score is compared to the confidence interval around the covariate adjusted FIM score (Uniform System for Medical Rehabilitation, 2008). This provides information to the facility about whether it is performing better (raw score > upper CL) or worse (raw

score < CL) than expected based on patient case mix. Public reports could then identify the number of hospitals that did better or worse than expected. These options represent methods of setting standards or benchmarks to which MHQI scores are expected to compare. Each reporting agency should set these benchmarks so that groups are fairly compared (i.e., benchmarks are attainable) but drive improvement in quality as well (i.e. it will be difficult for all groups to attain).

From a quality improvement standpoint, quality reporting should emphasize the integration of MHQIs into program evaluation or quality monitoring at the hospital or agency level (Bowen & Kreidler, 2008). Since each hospital with mental health beds in Ontario is responsible for submitting RAI-MH data to CIHI, the data can also be used internally. Hospitals can use the information at a program specific level to identify opportunities for improvement on an ongoing basis, linking the MHQIs to aspects of program staffing, delivery, and changes. These efforts can be useful for improving quality delivered at the point of care and for demonstrating improvement in public reporting of MHQI scores. They may also use MHQI results as a track record of evidence for efforts to improve quality as part of the accreditation process.

The use of MHQIs in pubic reporting, or for any purpose, is inherently grounded in the need for accountability of mental health services. The benefits of the MHQIs for service recipients have already been discussed. For hospitals, the MHQIs provide a mechanism to demonstrate to managing or governing bodies that they are providing effective services to those who need them. The MHQIs also provide a mechanism for those who administer service funding and those who receive funding that funding is being applied appropriately to services that are effective. While MHQI results in public reporting will identify hospitals that don't perform as optimally as others, the use of the MHQIs can be reassuring to the public that hospitals who don't achieve good outcomes can be identified so that quality improvement initiatives can be implemented.

The MHQIs also help provide balance against the inappropriate use of assessment information for enhancing funding. Since the RAI-MH SCIPP CMI is included in the psychiatric hospital funding formula for Ontario (JPPC, 2009) there may have been incentives for hospitals to maximize funding by

selecting patients who meet criteria for higher funding. For instance, patients with schizophrenia who are new admissions, have psychotic symptoms, and aggressive behaviour qualify for the highest SCIPP CMI group. A hospital that wanted to maximize funding could choose to admit a high number of these patients. Without the MHQIs, this hospital would not be accountable for ensuring that appropriate outcomes for these patients are achieved prior to discharge (e.g., reduction in positive symptoms and behaviours). With the MHQIs, that hospital would receive poorer scores on improvement in positive symptoms and aggressive behaviours. Therefore, the MHQIs provide a balancing mechanism so that recipients of funds cannot take advantage of assessments to maximize funding.

Prior to the development of MHQIs, no standard mechanism was in place for providers or administrators of health services to demonstrate that the services provided actually affect outcomes of those who received services. Subjective evaluations of satisfaction and perception of care are important and have been available for some time. However, the utility of the MHQIs as a set of clinically meaningful outcomes that reflect patterns and magnitudes of change and can be connected to care planning, funding, and public reporting will enhance the understanding and improvement of mental health services.

6.0 Conclusion

This dissertation has demonstrated that the RAI-MH contains valuable information for performing fair evaluations of quality of inpatient mental health services. The MHQIs are meaningful in that they are representative among inpatients of mental health hospitals in Ontario, feasible because they are based on provincially available data connected to public reporting infrastructures, and actionable since their content is related to clinical guidelines for interventions and information used to fund and manage inpatient mental health services. Without this research, stakeholders' understanding of the quality of inpatient mental health services would be constrained to conclusions about the types and appropriateness of services offered and the numbers of persons who were able to attain those services. While important, these conclusions lack information about the impact and effectiveness of

services. With the MHQIs, EQIs, and risk adjustment, there is opportunity to begin to investigate how mental health services actually affect the lives of those who receive them.

Encouragingly, the results from the MHQIs show that most persons do experience improvement in a variety of domains during the course of inpatient treatment. However, differences in these experiences among hospitals and regions with inpatient mental health beds in Ontario, particularly among non-traditional psychiatric domains, indicate that there are great opportunities to improve the impact of inpatient mental health services. The research completed in this dissertation represents a starting point for further research into the applications of MHQIs for understanding and comparing quality of mental health services. The application of this research into quality monitoring and reporting for mental health services will be essential for better understanding services, policies, and management practices that contribute to the overall well-being of individuals recovering from mental health conditions.

REFERENCES

Agesti, A. (2002). *An Introduction to Categorical Data Analysis*, 2nd ed. New York, NY: Wiley., p. 281.

Allen, M.H., Currier, G.W., Hughes, D.H., et al., (2003) Treatment of behavioral emergencies: a summary of the Expert Consensus Guidelines. *Journal of Psychiatric Practice, 9*,16-38.

American Psychiatric Association Task Force on Quality Indicators (1999). *Report of the American Psychiatric Association Task Force on Quality Indicators*. American Psychiatric Publishing, Inc.

Anthony, W. A. (1993). Recovery from mental illness: The guiding vision of the mental health service system in the 1990's. *Psychosocial Rehabilitation Journal, 16*(4), 11-23

Antonuccio, D.O., Danton, W.G., & DeNelsky, G.Y. (1995). Psychotherapy Versus Medication for Depression: Challenging the Conventional Wisdom With Data *Professional Psychology: Research and Practice, 26* (6), 574–585

Arling, G., Karon, S. L., Sainfort, F., Zimmerman, D. R., & Ross, R., (1997). Risk adjustment of nursing home quality indicators. *The Gerontologist, 37*, 757-766.

Arling, G., Kane, R. L., Lewis, T., & Mueller, C. (2005). Future development of nursing home quality indicators. *The Gerontologist, 45*(2), 147-156.

Ash, A.S., Shwartz, M., Pekoz, E.A. (2003). *Comparing outcomes across providers. In Iezzoni, LI (Ed) Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, Il: Health Administration Press, pp 297-355.

Atherly, A., Fink, A. S., Campbell, D. C., Mentzer, R. M., Henderson, W., Khuri, S., et al. (2004). Evaluating alternative risk-adjustment strategies for surgery. *American Journal of Surgery, 188*(5), 566-570.

Austin, P.C & Lee, D.S. (2009). The concept of the marginally matched subject in propensity-score matched analyses. *Pharmacoepidemiology and drug safety,18*(6):469-82.

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods, 7*(2), 127-150.

Banks, S.M., Pandiani, J.A., & Bramley, J. (2001). Approaches to risk-adjusting outcome measures applied to criminal justice involvement after community service. *The Journal of Behavioral Health, Services & Research, 28*(3), 235-246.

Banks, S. M., Pandiani, J. A., Schacht, L., & Gauvin, L. (1999). A risk-adjusted measure of hospitalization rates for evaluating community mental health program performance. *Administration and Policy in Mental Health., 26*(4), 269-279.

Bates, B., & Stineman, M. (2000). Outcome indicators for stroke: Application of an algorithm treatment across the continuum of postacute rehabilitation services. *Archives of Physical Medicine and Rehabilitation, 81*, 1468-1478.

Bhasin, K. & Williams, P. (2007). *Understanding LHINS: A review of the health system integration act and the integrated health services*. Report for the Canadian Research Network for Care in the Community and the Ontario Community Support Association. Downloaded from http://www.changefoundation.ca/docs/Bhasin.williams.2007.pdf

Berg, K., Fries, B., Jones, R., Mattke, S., Moore, T., Mor, V., et al. (2001). *Identification and Evaluation of Existing Quality Indicators that are Appropriate for Use in Long-Term Care Settings*. CMS Contract No: 500-95-0062/T.O. #4 .

Björkgren, M.A., Häkkinen ,U., Finne-Soveri, U.H., & Fries B.E. (1999). Validity and reliability of Resource Utilization Groups (RUG-III) in Finnish long-term care facilities, Scandinavian J*ournal of Public Health, 27*, 228 - 234.

Bowen, S., & Kreindler, S.A. (2008). Indicator madness: A cautionary reflection on the use of indicators in healthcare. *Healthcare Policy, 13* (4), 41-48.

Brook, R.H., McGlynn, E.A., & Shekelle, P.G. (2000). Defining and measuring quality of care: a perspective from US researchers. *International Journal for Quality in Health Care, 12* (4), 281-295.

Burgess, J. F., Christiansen, C. L., Michalak, S. E., & Morris, C. N. (2000). Medical profiling: Improving standards and risk adjustments using hierarchical models. *Journal of Health Economics, 19*(3), 291-309.

Burrows, A. B., Morris, J. N., Simon, S. E., Hirdes, J. P., & Phillips, C. (2000). Development of a minimum data set-based depression rating scale for use in nursing homes. *Age and Ageing, 29*(2), 165-172.

Canadian Council for Health Services Accreditation (2003). *AIM: Achieving Improved Measurement Accreditation Program*, 2nd Edition.

Canadian Institute for Health Information (1999). *National Consensus Conference on Population Health Indicators Final Report*, Ontario.

Canadian Institute for Health Information (2001a). *Development of National Indicators and Reports for Mental Health and Addiction Service: External Field Review Results*, Ottawa, Canada.

Canadian Institute for Health Information (2001b). *Development of Indicators for Mental Health & Addiction Service: Phase 1 Project Report July 2001*, Ottawa, Canada.

Canadian Institute for Health Information (2005) *The Health Indicators Project: The Next Five Years. Report from the second consensus conference on population health indicators*. Retrieved from: http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=indicators_e

Canadian Institute for Health Information (2008). *Hospital Mental Health Services in Canada, 2005–2006*. Retrieved from: http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=AR_364_E

Carrillo E, Garcia-Altes A, Peiro S, Portella E, et al. System for the classification of patients in mid and long-term care facilities: Resource Utilization Groups, version III. Validation in Spain. (in Spanish). *Revista de Gerontologia, 6*(4): 276-284

CMS (2009). *Design for Nursing Home Compare Five-Star Quality Rating System: Technical Users' Guide*. Downloaded from: http://www.cms.hhs.gov/CertificationandComplianc/Downloads/usersguide.pdf

Cohen-Mansfield J. (1986). Agitated behaviors in the elderly: II. Preliminary results in the cognitively deteriorated. *Journal of the American Geriatrics Society,34*, 722–727.

Commission on the Future of Health Care in Canada. (2002) *Building on Values: The Future of Health Care in Canada*. Ottawa: ON, National Library of Canada. Downloaded from http://dsp-psd.pwgsc.gc.ca/Collection/CP32-85-2002E.pdf

Cook, N. (2007). Use and misuse of the receiver operating characteric curve in risk prediction. *Circulation, 115*, 928-935.

Craig, T.J., Fennig, S., Tanenberg-Karant, M., & Bromet, E., (2000). Rapid versus delayed readmission in first-admission psychosis: Quality indicators for managed care? *Annals of Clinical Psychiatry, 12*, 233-238.

Cuffel, B.J. (2004). Use of Hierarchical Linear Models in the Case Mix Adjustment of Mental Health Provider Profile Scores. In Hendryx, M.S. *Conducting Case-Mix Adjustment for Mental Health Performance Indicators*. Cambridge, MA: Human Services Research Institute. Downloaded from: http://www.tecathsri.org/products_list.asp.

D'Agostino, R. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*, 2265-2281.

Dalby, D. M., Hirdes, J. P., & Fries, B. E. (2005). Risk adjustment methods for home care quality indicators (HCQIs) based on the minimum data set for home care. *BMC Health Services Research, 5(*1), 7.

Daley, J., Iezzoni, L., & Shwartz, M. (2003). Conceptual and practical issues in developing risk adjustment methods. In Iezzoni, L.I. (ed.). *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, IL: Health Administration Press, pp.179-207.

Day, N., Byar, D., & Green, S. (1980). Overadjustment in case-control studies. American *Journal of Epidemiology, 112* (5), 696-706.

Dickey, B., Hermann, R. C., & Eisen, S. V. (1998). Assessing the quality of psychiatric care: Research methods and application in clinical practice. *Harvard Review of Psychiatry, 6(*2), 88-96.

Dickey, B., Normand, S. L., Eisen, S., Hermann, R., Cleary, P., Cortes, D., et al. (2006). Associations between adherence to guidelines for antipsychotic dose and health status, side effects, and patient care experiences. *Medical Care, 44*(9), 827-834.

Dickey, B., Normand, S. L., Hermann, R. C., Eisen, S. V., Cortes, D. E., Cleary, P. D., et al. (2003). Guideline recommendations for treatment of schizophrenia: The impact of managed care. *Archives of General Psychiatry, 60*(4), 340-348.

Donabedian A. (1980). *Explorations in quality assessment and monitoring, vol 1: The definition of quality and approaches to its assessment*. Ann Arbor: Health Administration Press.

Donabedian A. (1982). *Explorations in quality assessment and monitoring, vol 2: The criteria and standards of quality*. Ann Arbor: Health Administration Press.

Dow, M. G., Boaz, T. L., & Thornton, D. (2001). Risk adjustment of florida mental health outcomes data: Concepts, methods, and results. *The Journal of Behavioral Health.Services & Research, 28*(3), 258-272.

Druss, B. G., Rosenheck, R. A., & Stolar, M. (1999). Patient satisfaction and administrative measures as indicators of the quality of mental health care. *Psychiatric Services, 50*(8), 1053-1058.

Edlund, ,M.J., Young, ,A.S., Kung, ,F.Y., Sherbourne, ,C.D., & Wells, ,K.B. (2003). Does satisfaction reflect the technical quality of mental health care? *Health Services Research, 38*(2), 631-645.

Eisen, S.V., Wilcox, M., Leff, H.S., Schaefer, E., Culhane, M.A. (1999) Assessing behavioral health outcomes in outpatient programs: Reliability and validity of the BASIS-32. The *Journal of Behavioral Health Services & Research, 26*, 5-17.

Evans, S. & Hosmer, D. (2004). Goodness of fit tests for logistic GEE models: Simulation results. *Communications in Statistics, 33*, 247-258.

Fauman, M. (1989). Quality assurance monitoring in psychiatry. *American Journal of Psychiatry, 146*, 1121-1130.

Fann, J.R., Uomoto, J.M. and Katon, W.J., (2001). Cognitive improvement with treatment of depression following mild traumatic brain injury. *Psychosomatics, **42***, 48–54

Fisher ES, Wennberg DE, Stukel TA, et al., (2003a). The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. *Annals of Internal Medicine. 138*(4),273-87.

Fisher ES, Wennber, Stukel T., (2003b) The implications of regional variations in Medicare spending, part 2: Health outcomes and satisfaction with Care. *Annals of Internal Medicine, 138*(4),288–298.

Fisher, E., Goodman, D., Skinner, J., & Bronner, K. (2009). *Health care spending, quality, and outcomes: More isn't always better* [A Dartmouth Atlas Project topic brief]. Retrieved from http://www.dartmouthadas.org/adases/Spending_ Brief_022709.pdf

Folstein MF, Folstein SE, & McHugh PR (1975). Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J*ournal of Psychiatric Research, 12* (3),189–98.

Fowler FJ Jr, Gallagher, Anthony et al. (2008). Relationship between regional per capita medicare expenditures and patient perceptions of quality of care. J*ournal of the American Medical Association, 299*(20):2406–2412.

Fries BE, Schneider DP, Foley JW, Gavazzi M, Burke R, Cornelius E. (1994) Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). *Medical Care 32*(7): 668-685.

Fries, B. E., Simon, S. E., Morris, J. N., Flodstrom, C., & Bookstein, F. L. (2001). Pain in U.S. nursing homes: Validating a pain scale for the minimum data set. *The Gerontologist, 41*(2), 173-179.

Fries, B. E., Morris, J. N., Aliaga, P., & Jones, R. (2003). Risk adjusting outcome measures for post-acute care. *American Journal of Medical Quality : The Official Journal of the American College of Medical Quality, 18*(2), 66-72.

Forrow, L., Taylor, B., Arnold, R.M. (1992) Absolutely relative: How research results are summarized can affect treatment decisions. *American Journal of Medicine, 22*,121-5

Fung, C., Lim, Y.W., Mattke, S., et al. (2008). Systematic review: The evidence that publishing patient care performance data improves quality of care. *Annals of Internal Medicine, 148*, 111-123.

Ganju, V., Smith, M.E., Adams, N., Allen, J., Bible, J., Danforth, M. et al (2005). *The MHSIP Quality Report: The Next Generation of Mental Health Performance Measures*. Rockville, MD: Center for Mental Health Services, Mental Health Statistics Improvement Program.

George L, Durbin J, and Koegl CJ (2008). System-wide implementation of ACT in Ontario: An ongoing improvement effort. *The Journal of Behavioral Health Services & Research,13*.

Gerolamo, A. (2004). State of the science: Outcomes of acute inpatient psychiatric care. *Archives of Psychiatric Nursing, 18* (6), 203-214.

Goering, P., Wasylenki, D., & Durbin, J. (2000). Canada's mental health system. International *Journal of Law and Psychiatry.Special Issue: Mental Health Policy 2000: An International Review, 23*(3-4), 345-359.

Goldberg, R.W., Green-Paden, L.D., Lehman, A.F., & Gold, J.M. (2001). Correlates of insight in serious mental illness**.** *The Journal of nervous and mental disease, 189*(3)*,*137-45.

Goldney, R.D., Fisher, L.J., Walmsley, S., Kent, P. & Cooper, A.W. (1996) A pilot study of the Health of the Nation Outcome Scale as a measurement of outcome in a private psychiatric hospital. *Australasian Psychiatry, 4* (6): 319-321.

Granger, C.V., Cotter, A.C., Hamilton, B.B., & Fiedler, R.C. (1993) Functional assessment scales: a study of persons after stroke. *Archives of Physical Medicine and Rehabililtation, 74*,133-138.

Gray, L. C. Bernabei, R. Berg, K. et al. (2008) Standardizing assessment of elderly people in acute care: The interRAI acute care instrument. *Journal of the American Geriatrics Society. 56*, 536-541.

Greenberg, G. A., & Rosenheck, R. A. (2004). Consumer satisfaction with inpatient mental health treatment in the department of veterans affairs. *Administration and Policy in Mental Health, 31*(6), 465-481.

Greenberg, G. A., & Rosenheck, R. A. (2005). Using the GAF as a national mental health outcome measure in the department of veterans affairs. *Psychiatric Services (Washington, D.C.), 56*(4), 420-426.

Greenberg, G. A., & Rosenheck, R. A. (2006). Use of nationwide outcomes monitoring data to compare clinical outcomes in specialized mental health programs and general psychiatric clinics in the veterans health administration. *The Psychiatric Quarterly, 77*(2), 151-172.

Gruneir A, Miller SC, Feng Z, et al. (2008). Relationship between state medicaid policies, nursing home racial composition, and the risk of hospitalization for black and white residents. *Health Services Research. 43*(3), 869-81.

Hawes, C., Morris, J., Phillips, C.D., et al. (1997). Development of the Nursing Home Resident Assessment Instrument in the USA. *Age and Ageing, 26*(2) , 19-25.

Hall, D., & Siegel, B. (2005). Taking the measure of measures. quality data initiatives and the challenge of effective and efficient data. *Journal of AHIMA / American Health.Information Management Association, 76*(2), 42-4, 46.

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. *143*,29 –36.

Hartwell, S, Fisher, W., & Deng, X. (2009) The impact of regionalization on reentry service outcomes for individuals with severe mental illness. *Psychiatric Services, 60*(3), 394-7

Hawes, C., Mor, V., Phillips, C.D., et al., (1997). The OBRA-87 nursing home regulations and implementation of the Resident Assessment Instrument: effects on process quality. *Journal of the American Geriatrics Society. 45*(8), 977–985.

Health Canada (1994). *Quality care and outcome indicators in the mental health field: A guide to the literature*. Ottawa: Minister of Supply and Services Canada.

Health Canada (2002). *Mental health, mental illness and addiction: overview policies and programs in Canada. A report on mental illness in Canada. Ottawa*, ON. Retrieved on December 5, 2008 from: http://www.phac-aspc.gc.ca/publicat/miic-mmac/index-eng.php

Hendryx, M. S., Dyck, D. G., & Srebnik, D. (1999). Risk-adjusted outcome models for public mental health outpatient programs. *Health Services Research, 34*(1 Pt 1), 171-195.

Hendryx, M. S., & Teague, G. B. (2001). Comparing alternative risk-adjustment models. The *Journal of Behavioral Health Services & Research, 28*(3), 247-257.

Hendryx, M. S., Beigel, A., & Doucette, A. (2001). Introduction: Risk-adjustment issues in mental health services. *Journal of Behavioral Health Services & Research, 28*(3), 225-234.

Hendryx, M., Moore, R., Leepe, T., et al. (2001) An examination of methods for risk-adjustment of rehospitalization rates. *Mental Health Services Research, 3*(1):15-24.

Hendryx, M., Russo, J.E., Stegner,B., et al., (2003) Predicting rehospitalization and outpatient services from administrative and clinical databases. *Journal of Behavioral Health Services and Research, 30*, 346-355.

Hendryx, M.S. (2004). *Conducting Case-Mix Adjustment for Mental Health Performance Indicators.* Cambridge, MA: Human Services Research Institute. Downloaded from: http://www.tecathsri.org/products_list.asp.

Herman, R.C. & Palmer, R.H. (2002). Common ground: A framework for selecting core quality measures for mental health and substance abuse care. *Psychiatric Services, 53* (3), 281-287.

Hermann, R.C, Provost, S. (2003). Interpreting Measurement Data for Quality Improvement: Means, Norms, Benchmarks, and Standards. *Psychiatric Services, 54*(5), 655-57.

Hermann, ,R.C., Palmer, ,H., Leff, ,S., Shwartz, ,M., Provost, ,S., Chan, ,J., et al. (2004). Achieving consensus across diverse stakeholders on quality measures for mental healthcare. *Medical Care., 42*(12), 1246-1253.

Hermann, R. C., Leff, H. S., Palmer, R. H., Yang, D., Teller, T., Provost, S., et al. (2000). Quality measures for mental health care: Results from a national inventory. *Medical Care Research and Review : MCRR, 57 Suppl 2*, 136-154.

Hermann, R. C., Chan, J. A., Provost, S. E., & Chiu, W. T. (2006). Statistical benchmarks for process measures of quality of care for mental and substance use disorders. *Psychiatric Services (Washington, D.C.), 57*(10), 1461-1467.

Hermann, R. C., Finnerty, M., Provost, S., Palmer, R. H., Chan, J., Lagodmos, G., et al. (2002). Process measures for the assessment and improvement of quality of care for schizophrenia. *Schizophrenia Bulletin, 28*(1), 95-104.

Hermann, R. C., Mattke, S., Somekh, D., Silfverhielm, H., Goldner, E., Glover, G., et al. (2006). Quality indicators for international benchmarking of mental health care. *International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care / ISQua, 18* Suppl 1, 31-38.

Hermann, R. C., & Palmer, R. H. (2002). Common ground: A framework for selecting core quality measures for mental health and substance abuse care. *Psychiatric Services (Washington, D.C.), 53*(3), 281-287.

Hermann, R. C., Rollins, C. K., & Chan, J. A. (2007). Risk-adjusting outcomes of mental health and substance-related care: A review of the literature. *Harvard Review of Psychiatry, 15*(2), 52-69.

Hermann, R.C. (2005). *Improving Mental Healthcare: A Guide to Measurement-Based Quality Improvement.* Washington DC: American Psychiatric Press, Inc.

Hirdes JP, Botz CA, Kozak J, & Lepp V. (1996) Identifying an appropriate case-mix measure for chronic care: Evidence from an Ontario pilot study. *Healthcare Management Forum 9*(1):40-46.

Hirdes, J.P., Fries, B.E., Morris, J.N et al. (1999). Integrated health information systems based on the RAI/MDS series of instruments. *Healthcare Management Forun, 12*, 30-40.

Hirdes JP, Fries BE, Botz C, et al. (2003). T*he System for Classification of In-Patient Psychiatry (SCIPP): A New Case-Mix Methodology for Mental Health*. Unpublished manuscript.

Hirdes, J. P., Fries, B. E., Morris, J. N., Ikegami, N., Zimmerman, D., Dalby, D. M., et al. (2004). Home care quality indicators (HCQIs) based on the MDS-HC. *The Gerontologist, 44*(5), 665-679.

Hirdes, J. P., Marhaba, M., Smith, T. F., Clyburn, L., Mitchell, L., Lemick, R. A., et al. (2000). Development of the resident assessment instrument--mental health (RAI-MH). *Hospital Quarterly, 4*(2), 44-51.

Hirdes, J. P., Smith, T. F., Rabinowitz, T., Yamauchi, K., Perez, E., Telegdi, N. C., et al. (2002). The resident assessment instrument-mental health (RAI-MH): Inter-rater reliability and convergent validity. *The journal of behavioral health services & research, 29*(4), 419-432.

Hirdes, J. P., Frijters, D. H., & Teare, G. F. (2003). The MDS-CHESS scale: A new measure to predict mortality in institutionalized older people. *Journal of the American Geriatrics Society, 51*(1), 96-100.

Hirdes, J.P., Ljunggren, G., Morris, J., et al., (2008). Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. *BMC Health Services Research, 8*, 277.

Hodgson, ,R.E., Lewis, ,M., & Boardman, A.P. (2001). Prediction of readmission to acute psychiatric units. *Social Psychiatry and Psychiatric Epidemiology, 36*(6), 304-309.

Hoffman DL, Dukes EM, Wittchen HU (2008) Human and economic burden of generalized anxiety disorder. *Depress Anxiety 25*:72–90.

Hosmer, D.W. & Lemeshow, S. (2000). *Applied Logistic Regression,* 2nd edn. New York, NY: Wiley.

Hu, F., Goldberg, J., Hedeker, D., et al., (1998). Comparison of population-averaged and subject specific approaches for analyzing repeated binary outcomes. *American Journal of epidemiology, 147*, 694-703.

Huang, I., Frangakis, C., Dominici, F., Diette, G. B., & Wu, A. W. (2005). Application of a propensity score approach for risk adjustment in profiling multiple physician groups on asthma care. *Health Services Research, 40*(1), 253-278.

Humphreys, K. & Weingardt, K. (2000). Assessing readmission to substance abuse treatment as an indicator of outcome and program performance. *Psychiatric Services, 51*, 1568-1569.

Hussey, P.S., Mattke, S., Morse, L.,, & Ridgely, M.S. (2007). *Evaluation of the use of AHRQ and other quality indicators. Report for the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services*. Downloaded from: http://www.ahrq.gov/about/evaluations/qualityindicators/

Iezzoni, L. I. (1997). The risks of risk adjustment. *JAMA : The Journal of the American Medical Association, 278*(19), 1600-1607.

Iezzoni, L.I. (2003a). *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, IL: Health Administration Press.

Iezzoni, L.I. (2003b). Coded data from administrative sources. In Iezzoni, LI (Ed) *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, Il: Health Administration Press.

Ikegami N, Fries BE, Takagi U, et al. (1994) Applying RUG-III in Japanese long-term care facilities. *The Gerontologist 34* (5): 628 - 639.

Imbens, G.W. (2000). The Role of the Propensity Score in Estimating Dose-Response Functions. *Biometrika 87* (3): 706–10.

Jacobs P, Dewa C, Bland R et al. (2007) *Mental Health Economic Statistics : In your pocket. Alberta Mental Health Board and Institute for Health Economics*. Downloaded from http://www.phs.ualberta.ca//pdfs/IHE_mental_health_stats.pdf

Jin, P. (1992). Towards a reconceptualization of the law of initial value. *Psychological Bulletin, 111*, 176-184.

Joint Commission on Accreditation of Healthcare Organizations (1999). *ORYX: The next evolution in accreditation*. Retrieved from: http://www.jcaho.org/perfmeas/oryx/oryx_qa.htm

Joint Commission on Accreditation of Healthcare Organizations (2008) *The Specifications Manual for National Hospital Inpatient Quality Measures - Hospital-Based Inpatient Psychiatric Services (HBIPS) Core Measure Set, Version 2.0.*

Joint Policy and Planning Committee (1999). *Resident Assessment Instrument-Mental Health (RAI-MH): Quality Indicators for Mental Health (QIMHs)*.

JPPC Mental Health Technical Working Group. (2008). *Grouper and Weighting Methodology for Adult Inpatient Mental Health Care in Ontario*. Toronto, ON: Ontario Joint Policy and Planning Committee. Retrieved from www.jppc.org

Jones, R.N., Hirdes, J.P., Poss, J.W. et al. (in press). Adjustment of Nursing Home Quality Indicators. *BMC Health Services*.

Kay, S.R., Fiszbein, A., & Opler, L.A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin, 13*, 261-276.

Kirby, M. (2004a). *Mental Health, Mental Illness, and Addiction: Overview of Policies and Practices in Canada*. Interim Report of the Standing Senate Committee on Social Affairs, Science, and Technology. Retrieved on November 5th, 2008 from: http://www.parl.gc.ca

Kirby, M. (2004b). *Mental Health Policies and Programs in Selected Countries*. Interim Report of the Standing Senate Committee on Social Affairs, Science, and Technology. Retrieved on November 5th, 2008 from: http://www.parl.gc.ca

Kirby, M. (2004c). *Mental Health, Mental Illness, and Addiction: Issues and Options for Canada*. Interim Report of the Standing Senate Committee on Social Affairs, Science, and Technology. Retrieved on November 5th, 2008 from: http://www.parl.gc.ca

Lambert, E.W., Doucette, A., & Bickman, L. (2001). Measuring mental health outcomes with pre-post designs. *The Journal of Behavioral Health Services & Research, 28* (3), 273-286.

Landis, R. & Koch, G. (1977), The measurement of observer agreement for categorical data. *Biometrics, 22*, 159-174.

Latimer, E.A. (1999). Economic impacts of assertive community treatment: a review of the literature. *Canadian Journal of Psychiatry, 44*, 443-454.

Latimer, E. (2005). Community-based care for people with severe mental illness in Canada. *International Journal of Law & Psychiatry, 28*(5), 561-573.

Lehman A, Goldman H, Dixon L, & Churchill R (2004). *Evidence-based mental health treatments and services: Examples to inform public policy*. Report for the Milbank Memorial Fund.

Leon, B. & Hedeker, D. (2007). A comparison of mixed-effects quantile stratification propensity adjustment strategies for longitudinal treatment effectiveness analyses of continuous outcomes. *Statistics in Medicine; 26*, 2650–2665

Liang, K.Y. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,73*:13-22.

Lieberman, P., Wiitala, S., Elliott, B., et al., (1998). Decreasing length of stay: Are there effects on outcomes of psychiatric hospitalization? *American Journal of Psychiatry, 155*, 905-909.

Lied, T. R., Kazandjian, V. A., & Hohman, S. F. (1999). Impact of risk adjusted clinical outcomes methodology--quality measures on hospital mortality data: A statistical and case study approach. *American Journal of Medical Quality : The Official Journal of the American College of Medical Quality, 14*(6), 255-261.

Lifford, R., Mohammed, M.A., Spiegelhalter, D, & Thomson, R. (2004). Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *The Lancet, 363*, 1147-1154.

Leigh, J.P. (1988). Assessing the importance of the independent variable in multiple regression: Is stepwise unwise?. *J Clin Epidemiol, 41* (7), 669-677.

Lin E., Degendorger N., Durbin, J., Prendergast P., & Goering P. (2001). *Mental Health. In Hospital Report Research Collaborative. Hospital Report 2001. Preliminary Studies – Volume One Exploring: Rehabilitation and Mental Health*. Joint Initiative of the Ontario Hospital Association and the Government of Ontario. Toronto: Hospital Report Research Collaborative, University of Toronto.Retrieved from www.hospitalreport.ca .

Lin E., Durbin J., Koegl C., Murray M., Tucker T., Daniel I., et al. (2004). *Hospital Report 2004:Mental Health.*. Joint Initiative of the Ontario Hospital Association and the Government of Ontario. Toronto: Hospital Report Research Collaborative, University of Toronto.Retrieved from www.hospitalreport.ca .

Lin, E., Durbin, J., Koegl, C., et al. (2008). *Hospital Report 2007: Mental Health*. Joint Initiative of the Ontario Hospital Association and the Government of Ontario. Toronto: Hospital Report Research Collaborative, University of Toronto.

Lorenz, S., & Sederer, L. I. (. (2001). Risk adjustment of clinical outcome data. In B. Dickey, & L. I. Sederer (Eds.)*, Improving mental health care: Commitment to quality*. (pp. 89-99). Washington, DC, US: American Psychiatric Publishing, Inc.

Love, T. (2008). R*educing the Impact of Selection Bias with Propensity Scores.* Presentation at the 7th International Conference on Health Policy Statistics, Philadelphia, January 18, 2008.

Love, T.E., Cebul, R.D., Thomas, C.L ., & Dawson, N.V. (2003) Effect of matching for propensity on the balance of "unmeasured" covariates. *Journal of Clinical Epidemiology,* 56, 920.

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. *The American Journal of Psychiatry., 154*(3), 337-340.

Mant, J. (2001). Process versus outcome indicators in the assessment of quality of health care. *International Journal for Quality in Health Care, 13* (6), 475-480.

Martin, L., Hirdes, JP, Fries, BE, Smith, TF. (2007) Development and Psychometric Properties of an Assessment for Persons With Intellectual Disability: The interRAI ID. *Journal of Policy and Practice in Intellectual Disabilities, 4*: 1 23.

Martin, L., & Hirdes, J.P. (2008). Mental health needs and service use in Ontario. *Healthcare Management Forum, Spring 2009*, 40-46.

Martin, L., Hirdes, J.P., Morris J., et al. (2009). Validating the Mental Health Assessment Protocols (MHAPs) in the Resident Assessment Instrument Mental Health (RAI-MH). *Journal of Psychiatric and Mental Health Nursing*, online publication.

McAuliffe, W.E. (1979). Measuring the quality of medical care: Process versus outcome. *Millbank Memorial Fund Quarterly/Health and Society, 57*, 118-151.

McEwan, K. L., & Goldner, E. M. (2002). Keeping mental health reform on course: Selecting indicators of mental health system performance. *Canadian Journal of Community Mental Health = Revue Canadienne De Sante Mentale Communautaire, 21*(1), 5-16.

McGrath, B.M. & Tempier, R. (2003). Implementing quality management in psychiatry: From theory to practice-Shifiting focus from process to outcome. *Can J Psychiatry, 48*, 467-474.

Meehan, T. J., Stedman, T. J., Neuendorf, K. E., Francisco, I. D., & Neilson, M. G. (2007). Benchmarking australia's mental health services: Is it possible and useful? *Australian Health Review : A Publication of the Australian Hospital Association, 31*(4), 623-627.

Mental Health Statistics Improvement Program (1996). *The Final Report of the Mental Health Statistics Improvement Program (MHSIP) Task Force on a Consumer Oriented Mental Health Report Card*. Centre for Mental Health Services, Department of Health and Human Services, Washington DC.

Mor, V., Angelelli, J., Jones, R., Roy, J., Moore, T., Morris, J. (2003). Inter-rater Reliability of Nursing Home Quality Indicators in the U.S. *BMC Health Serv Res. 3*(1): 20.

Mor V, Branco K, Fleishman J, Hawes C, Phillips C, Morris JN, Fries BE. (1995) The Structure Of Social Engagement Among Nursing Home Residents. *Journals of Gerontology: Psychological Sciences. 50*(1),P1-P8.

Mor, V. (2005). Improving the quality of long-term care with better information. *The Milbank Quarterly, 83*.

Morris, J., Fries, B. E., Steel, K., Ikegami, N., Bernabei, R., & Carpenter, I., et al (1997). Comprehensive clinical assessment in community setting: Applicability of the MDS-HC. *Journal of the American Geriatrics Society, 45* 1017-1024

Morris, J. N., Fries, B. E., & Morris, S. A. (1999). Scaling ADLs within the MDS. *The Journals of Gerontology.Series A, Biological Sciences and Medical Sciences, 54(*11), M546-53.

Morris, J. N., Fries, B. E., Mehr, D. R., & Hawes, C. (1994). MDS cognitive performance scale©. *Journals of Gerontology, 49*(4), M174-M182.

Morris, J. N., Nonemaker, S., Murphy, K., & Hawes, C. (1997). A commitment to change: Revision of HCFA's RAI. *Journal of the American Geriatrics Society, 45*(8), 1011-1016.

Morris J, Murphy K, Mor V, Berg K, Moore T. (2002) *Validation of Long Term and Post Acute Care Quality Indicators*. Final Draft Report, August 2, 2002. CMS Contract 500-95-0062.

Morton, V. & Torgerson, D. (2009). Effect of regression to the mean on decision making in health care. *British Medical Journal, 326*, 1083-1084.

National Association of State Mental Health Program Directors (NASMHPD; 1998). *Performance Measures for Mental Health Systems: Standardized Framework*. NASMHPD Research Institute.

New Freedom Commission on Mental Health, (2003). *Achieving the Promise: Transforming Mental Health Care in America*. Final Report. DHHS Pub. No. SMA-03-3832. Rockville, MD:. Downloaded from: http://www.mentalhealthcommission.gov/reports/FinalReport/downloads/downloads.html

Nicholl, J. (2007). Case mix adjustment in non-randomised observational evaluations: The constant risk fallacy. *Journal of Epidemiology and Community Health, 61*, 1010-1013.

NMHWG Information Strategy Committee Performance Indicator Drafting Group (2005). *Key Performance Indicators for Australian Public Mental Health Services*. ISC Discussion Paper No. 6 Australian Government Department of Health and Aging, Canberra.

Oken, D. & Heath, H. (1963). The law of initial values: Some further considerations. *Psychosomatic Medicine, 25*, 3-12.

Ontario Health Quality Council (2009). *Q-Monitor: 2009 Report on Ontario's Health System.* Downloaded from http://www.ohqc.ca/pdfs/ohqc_2009_report_-_english.pdf

Ontario Health Quality Council & Joint Policy and Planning Committee (2008). *Accountability Aggreements in Ontario's Health Sustem: How can they Accelerate Quality Improvement and enhance Public Reporting*. OHQC & JPPC White Paper. Downloaded from http://www.ohqc.ca/pdfs/accountability_agreements_in_ontario-july_31_2008.pdf

Ontario Ministry of Health and Long Term Care (1999). *Making it Happen: Implementation Plan for Mental Health Reform*. Toronto: MOHLTC.

Ontario Ministry of Health and Long Term Care (2003). *Mental Health Accountability Framework*. Toronto: MOHLTC.

Perlman, C.M. & Hirdes, J.P. (2008). The aggressive behaviour scale (ABS): A new scale to measure aggression based on the MDS. *Journal of the American Geriatrics Society, 56*, 2298-2303.

Phillips, C.D., Zimmerman, D., Bernabei, R., Jonsson, P.V. (1997) Using the Resident Assessment Instrument for Quality Enhancement in Nursing Homes. *Age and Ageing, 26*: 77-81.

Phillips, S. D., Hargis, M. B., Kramer, T. L., Lensing, S. Y., Taylor, J. L., Burns, B. J., et al. (2000). Toward a level playing field: Predictive factors for the outcomes of mental health treatment for adolescents. J*ournal of the American Academy of Child and Adolescent Psychiatry, 39*(12), 1485-1495.

Phillips, S. D., Kramer, T. L., Compton, S. N., Burns, B. J., & Robbins, J. M. (2003). Case-mix adjustment of adolescent mental health treatment outcomes. *Journal of Behavioral Health Services & Research, 30*(1), 125-136.

Raudenbush, S.A. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods Advanced Quantitative Techniques in the Social Sciences, 2nd Edition*. Thousand Oaks: CA, Sage Publications, Inc.

Rantz, M.J., Popejoy, L., Mehr, D.R., Zwygart-Stauffacher, M., Hicks, L.L. and Grando, V. (1997). Verifying Nursing Home Care Quality Using Minimum Data Set Quality Indicators and Other Quality Measures. *Journal of Nursing Care Quality 12*: 54-62.

Reeleder, D., Goel, V., Singer, P., & Martin, D. (2006). Accountability agreements in Ontario hospitals: Are they fair? *Journal of Public Administration Research and theory, 18*, 161-175.

Richardson, D., Tarnow-Mordi, W. O., & Lee, S. K. (1999). Risk adjustment for quality improvement. *Pediatrics, 103*(1 Suppl E), 255-265.

Rosenbaum, P. R., and Rubin, D. B., (1983), The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

Rosenheck, R., Fontana, A., & Stolar, M. (1999). Assessing quality of care: Administrative indicators and clinical outcomes in posttraumatic stress disorder. *Medical Care, 37*, 180-188.

Rossi, A., Daneluzzo, E., Arduini, L. et al., (2000). Cognitive symptoms of mania in pure and mixed episodes evaluated with the Positive and Negative Syndrome Scale. *European Archives of Psychiatry and Clinical Neuroscience, 250*, 254-256.

Schacht, L. M., & Hines, H. (2003). Recent applications of risk adjustment for performance measures used by state inpatient psychiatric facilities. *Harvard Review of Psychiatry., 11*(4), 220-224.

Schinnar, A., Kamis-Gould, E., Delucia, N., & Rothbard, A. (1990). Organiational determinants of efficiency and effectiveness in mental health partial care programs. *Health Services Research, 25,* 387-420.

Schisterman, E., Cole, S., & Platt, R. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology, 20* (4), 488-495.

Schneider, E.C., Zaslavsky, A.M., & Epstein, A.M. (2002). Racial Disparities in the Quality of Care for Enrollees in Medicare Managed Care. *JAMA. 287*:1288-1294.

Secker, B., Goldenberg, M., Gibson, B., et al. (2006). Just regionalisation: rehabilitating care for people with disabilities and chronic illnesses. *BMC Medical Ethics, 7* (9).

Shahian, D. M., & Normand, S. L. T. (2008). Comparison of "risk-adjusted" hospital outcomes. *Circulation, 117*, 1955-1963.

Shwartz, M., & Ash, A. (2003). Evaluating risk adjustment models empirically. In Iezzoni, L.I. (ed.), *Risk Adjustment for Measuring Health Care Outcomes*, 3rd ed. Chicago, IL: Health Administration Press., pp. 231-275.

Sirovich,B., Gottlieb, D., Welch, G., & Fisher, E. (2006) Regional variations in health care Intensity and physician perceptions ofquality of care. *Annals of Internal Medicine, 144*: 641-649.

Srebnik, D., Hendryx, M., Stevenson, J., Caverly, S., Dyck, D. G., & Cauce, A. M. (1997). Development of outcome indicators for monitoring the quality of public mental health care. *Psychiatric Services, 48*(7), 903-909.

Steel, K., Ljunggren, G., Topinkova, E., Morris, J.N., Vitale, C., Parzuchowski, P., Fries, B.E., on behalf of interRAI. (2003). RAI-PC: An Assessment Instrument for Palliative Care in all Settings. *Am J Hospice & Palliative Care, 20*(3):211-219.

Steel, Z., McDonald, R, Silove, D, Bauman, A., Sandford, P., Herron, J., Minas H., (2006). Pathways to the first contact with specialist mental health care. *Australian and New Zealand Journal of Psychiatry, 40* (4), 347-354.

Stephens, T., & Joubert, N. (2001). "The Economic Burden of Mental Health Problems in Canada." *Chronic Diseases in Canada* 22(1), 18-23.

Substance Abuse & Mental Health Services Administration (2006). *National Outcome Measures (NOMs) for Co-Occurring Disorders*. Downloaded from: http://www.oas.samhsa.gov/NOMsCoOccur2k6.pdf

Topinková, E., Neuwirth, J., Mellanová, A., Stanková, M., & Haas, T. (2000). Case-mix classification for Postacute and Long-Term Care. Validation of Resource Utilization Groups III (RUG-III) in the Czech Republic. In Czech. *Casopis Lékaru Ceský. 139*(2), 42-48.

Uniform Data Systems for Medical Rehabilitation (2008). *The FIM System CMG Report Explanation Guide, Version 1.0*. Amherst, NY.

VanderWeele, T. (2006). The use of propensity score methods in psychiatric research. *International Journal of Methods in Psychiatric Research, 15*(2), 95-103.

Veldhuizen, S, Urbanoski, K, & Cairney, J. (2007). Geographical variation in the prevalence of problematic substance use in canada. *Can J Psychiatry, 52*, 426–433

Wallace, J., Teare, G., Verrall, T., & Chan, B. (2007). *Public Reporting on the Quality of Healthcare: Emerging Evidence on Promising Practices for Effective Reporting*. Ottawa: ON, Canadian Health Services Research Foundation.

Wang, J., Donnan, P., Steinke, D & MacDonald, T. (2001). The multiple propensity score for analysis of dose-response relationships in drug safety studies. *Pharmacoepidemiolgy and Drug Safety, 10,* 105-111.

Weissman, E. M., Rosenheck, R. A., & Essock, S. M. (2002). Impact of modifying risk adjustment models on rankings of access to care in the VA mental health report card. *Psychiatric Services, 53*(9), 1153-1158.

Weitzen, S., Lapane, K., Toledano, A et al., (2004). Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiology and Drug Safety, 13,* 841-853.

Wennberg JE, Fisher ES, Skinner JS. (2002). Geography and the debate over medicare reform. *Health Affairs, 21*, w96–w114.

Wilcosky, T. C., & Chambless, L. E. (1985). A comparison of direct adjustment and regression adjustment of epidemiologic measures. *Journal of Chronic Diseases, 38*(10), 849-856.

Woo, J. & Shiu, Y. (2008). Inpatient stroke rehabilitation efficiency: Influence of organization or service delivery and staff numbers. *BMC Health Services Research, 8*, 86.

World Health Organization World Mental Health Survey Consortium (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA, 291* (21), 2581-2590.

Zaslavsky, A. (2001). Statistical issues in reporting quality data: small samples and casemix variation. *International Journal for Quality in Health Care, 13,* 481-488.

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121-130.

Zimmerman, D. R. (2003). Improving nursing home quality of care through outcomes data: The MDS quality indicators. *International Journal of Geriatric Psychiatry, 18*(3), 250-257.

Zimmerman, D., Karon, S., Arling, G., Clark, B. et al (1995). Development and testing of nursing home quality indicators. *Health Care Financing Review, 16*, 107-115.

APPENDICES

Appendix A: Geographic, Contextual, and Health Indicator Information about LHINs.

List of LHINs in Ontario:

1. Erie St. Clair
2. South West
3. Waterloo Wellington
4. Hamilton Niagara Haldimand Brant
5. Central West
6. Mississauga Halton
7. Toronto Central
8. Central
9. Central East
10. South East
11. Champlain
12. North Simcoe Muskoka
13. North East
14. North West

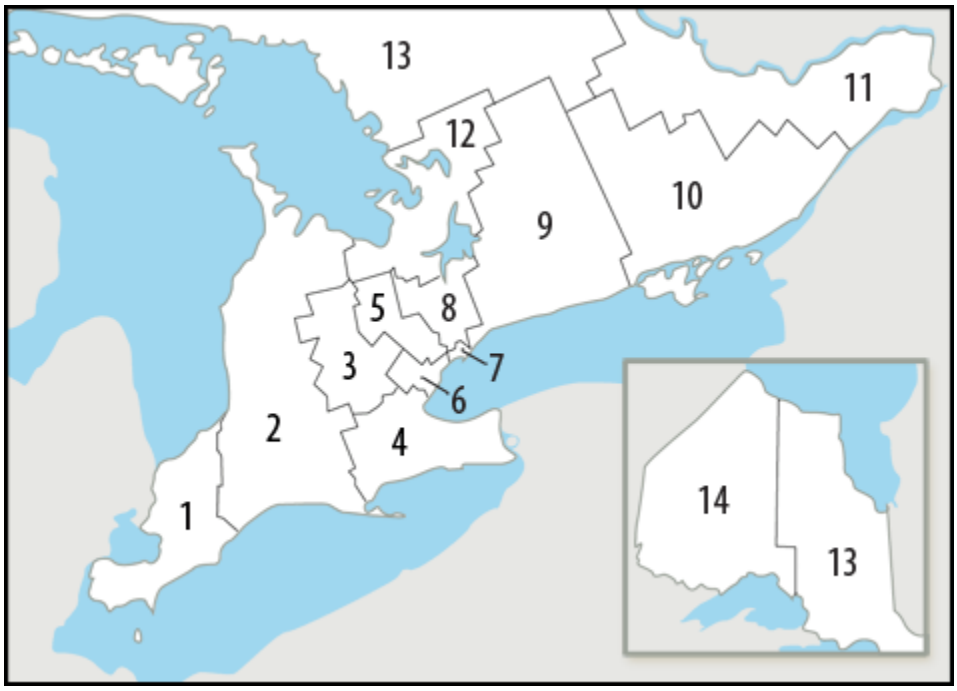


Figure 1. Map of LHINs in Ontario.

Table 1. Regional contextual information for Ontario LHINs in 2005.

Health Region	Regional Population ('000)	Percent Female Population	Percent 19 and Younger	Percent 65 and Older	Number of General Hospitals with Mental Illness/Addiction Separations	Number of Psychiatric Hospitals
Central	1,577	50.68	24.94	11.31	8	0
Central East	1,464	50.88	24.59	13.14	14	1
Central West	741	50.38	26.84	9.47	3	0
Champlain	1,180	50.75	24.33	12.61	17	3
Erie St. Clair	646	50.31	25.31	13.69	8	0
Hamilton Niagara Haldimand Brant	1,365	50.70	24.60	14.72	19	1
Mississauga Halton	1,076	50.44	26.77	9.76	5	0
North East	568	50.58	23.57	15.55	25	3
North Simcoe Muskoka	423	50.19	25.52	14.13	6	2
North West	241	49.89	26.11	13.23	14	1
South East	483	50.51	23.27	15.90	11	1
South West	929	50.56	25.08	14.47	27	2
Toronto Central	1,149	51.30	22.32	13.26	7	2
Waterloo Wellington	699	50.16	26.04	11.60	8	1

Source: CIHI Quick Statistics on Mental Health. Retrieved on December 8, 2008 http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=statistics_results_topic_mentalhealth_e&cw_topic=Health%20Services&cw_subtopic=Hospital%20Mental%20Health%20Services

Table 2. Regional health indicators for Ontario hospitals by health regions in 2005.

	Health Region	Separations	Regional Crude Separation Rate /100K	Regional Standardized Separation Rate/100K	Days Stay	Regional Days Stay Rate/100K	Regional Standardized Days Stay Rate/100K	30-Day Acute Care Hospital Readmission Rate/100 Inpatients with Mental Illness as Primary Diagnosis	1-Year Acute Care Hospital Readmission Rate/100 Inpatients with Mental Illness as Primary Diagnosis
1	Waterloo Wellington	2,677	383.22	378.04	26,727	3,826.00	3,733.31	10.66	24.45
2	Toronto Central	4,745	413.06	385.74	75,079	6,535.70	5,947.78	11.56	25.70
3	South West	4,781	514.37	498.60	52,042	5,599.00	5,224.44	9.58	23.81
4	South East	2,228	461.36	454.13	29,365	6,080.70	5,604.20	8.59	21.12
5	North West	1,978	820.64	849.82	23,600	9,791.30	9,701.33	9.30	25.27
6	North Simcoe Muskoka	2,122	501.65	482.67	18,044	4,265.70	3,901.81	8.03	20.72
7	North East	5,007	881.93	872.61	47,674	8,397.30	7,678.18	11.84	27.73
8	Mississauga Halton	3,072	285.49	282.16	39,744	3,693.50	3,674.87	11.47	25.68
9	Hamilton Niagara Haldimand Brant	6,441	471.97	461.05	75,837	5,557.00	5,140.38	6.92	20.36
10	Erie St. Clair	3,552	549.54	536.46	48,080	7,438.60	7,133.72	6.16	18.18
11	Champlain	5,192	439.84	422.00	80,681	6,834.80	6,429.00	8.73	23.32
12	Central West	2,423	326.90	326.36	27,437	3,701.70	3,690.32	9.22	25.13
13	Central East	5,941	405.87	388.28	70,063	4,786.50	4,454.98	9.58	24.78
14	Central	4,523	286.74	278.27	62,898	3,987.40	3,864.19	8.00	20.18
15	All Regions	54,682	436.01	420.97	677,271	5,400.30	5,078.84	9.17	23.11

Source: CIHI Quick Statistics on Mental Health. Retrieved on December 8, 2008 http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=statistics_results_topic_mentalhealth_e&cw_topic=Health%20Services&cw_subtopic=Hospital%20Mental%20Health%20Services

Appendix B: Selected quality indicators for mental health from the U.S. and internationally.

Table 1. Consensus based quality indicators chosen for mental health quality in the U.S

Indicator	Setting	Conditions
≥ 1 visit with adult caregiver of child ≤ 13 treated for a psychiatric or substance-related disorder in 3-month period24	Outpatient	Mental disorders, substance-use
Clinician contact with family member of consenting individuals with schizophrenia at initial evaluation	Inpatient, outpatient	Schizophrenia
Cumulative daily antipsychotic dosage between 300-1000 CPZ equivalents at hospital discharge for schizophrenia	Inpatient	Schizophrenia
Prescription of atypical drug for individuals with ≥ 1 clinical service for schizophrenia in 6-month period.	All	Schizophrenia
Length of treatment ≥ 90 days after initiation for substance-related disorder	All	Substance Use
≥ 3 medication visits or ≥ 8 psychotherapy visits in a 12-week period after new diagnosis of major depression	Outpatient	Depression
Clinician contact with family member of consenting individuals with schizophrenia at initial evaluation	Inpatient, outpatient	Schizophrenia
Cumulative daily antipsychotic dosage between 300–1000 CPZ equivalents at hospital discharge for schizophrenia	Inpatient	Schizophrenia
Prescription of atypical antipsychotic drug for individuals with >1 clinical service for schizophrenia in 6-month period	All	Schizophrenia
Length of treatment > 90 days after initiation for substance-related disorder	All	Substance-use disorders
> 3 medication visits or > 8 Adults psychotherapy visits in 12-week period after new diagnosis of major depression	Outpatient	Depression
> 12-week continuation after initiation of antidepressant drug for major depression	Outpatient	Depression
Daily antipsychotic dosage between 0.5–9.0 CPZ equivalents per kg body weigh at discharge for individual <18 hospitalized for psychotic disorder	Inpatient, residential	Psychotic disorders
Daily antipsychotic dosage < 200 CPZ equivalents for nursing home resident with dementia without psychotic symptoms in 3-month period	Nursing home	Dementia
> 1 serum drug level taken for individuals with bipolar disorder treated with mood stabilizers in 12 month period	Outpatient	Bipolar disorder
Avoidance of an anticholinergic antidepressant drug for individuals > 65 prescribed antidepressants	Inpatient, outpatient	Depression
> 1 psychotherapy visit for individuals within 6 months of hospitalization or ER visit for borderline personality disorder	Outpatient	Borderline personality disorder

Adapted from Herman et al., 2004

Table 2. Quality indicators selected for international benchmarking (Herman et al., 2006).

Domain	Quality Indicator	Description
Treatment	Visits during acute phase treatment of depression	% of persons with a new diagnosis of major depression who receive at least three medication visits or at least eight psychotherapy visits in a 12-week period.
	Hospital readmissions for psychiatric patients	% of discharges from psychiatric in-patient care during a 12-month reporting period readmitted to psychiatric in-patient care that occurred within 7 and 30 days.
	Length of treatment for substance-related disorders	% of persons initiating treatment for a substance-related disorder with treatment lasting at least 90 days.
	Use of anticholinergic antidepressant drugs among elderly patients	% of persons age 65+ years prescribed antidepressants using an anticholinergic anti-depressant drug.
	Continuous antidepressant medication treatment in acute phase	% of persons age ≥18 years who are diagnosed with a new episode of depression and treated with antidepressant medication, with an 84-day (12-week acute treatment phase) treatment with antidepressant medication.
	Continuous antidepressant medication treatment in continuation phase	% of persons age ≥18 years who are diagnosed with a new episode of depression and treated with antidepressant medication, with a 180-day treatment of antidepressant medication.
Continuity	Timely ambulatory follow-up after mental health hospitalization	% of persons hospitalized for primary mental health diagnoses with an ambulatory mental health encounter with a mental health practitioner within 7 and 30 days of discharge.
	Continuity of visits after hospitalization for dual psychiatric/ substance-related conditions	% of persons discharged with a dual diagnosis of psychiatric disorder and substance abuse with at least four psychiatric and at least four substance abuse visits within the 12 months after discharge.
	Racial/ethnic disparities in mental health follow-up rates	% of persons with a mental health-related visit receiving at least one visit in 12 months after initial visit stratified by race/ethnicity.
	Continuity of visits after mental health-related hospitalization	% of persons hospitalized for psychiatric or substance-related disorder with at least one visit per month for 6 months after hospitalization.
Coordination	Case management for severe psychiatric disorders	% of persons with a specified severe psychiatric disorder in contact with the health care system who receive case management (all types).
Outcome	Mortality for persons with severe psychiatric disorders	Standardized mortality rate for % of persons in total population with specified severe psychiatric disorders.

Appendix C: Ontario Mental Health Accountability Framework quality domains and indicators.

Domain	Quality Indicator
Acceptability	Consumer satisfaction
	Formal complaints
	Charter of rights
	Consumer/family involvement in treatment
	Consumer/family involvement in planning/delivery
	Cultural Sensitivity
	Consumer/family choice of services
Accessibility	Service reach to persons with severe mental illness
	Service reach to homeless
	Access to psychiatrists
	Access to primary care
	Wait-time for needed services
	Availability of afterhours care & transportation
	Denial of service
	Early intervention
	Consumer perception of access
	Identify human resource gaps
	Access to continuum of mental health service
	Criminal justice involvement
Appropriateness	Existence of best practice programs
	Fidelity: adherence to best practices*
	Best practice programs provided to persons with SMI*
	Treatment protocol for co-morbidity*
	Hospital readmission rate*
	Involuntary committal rate*
	Average length of stay in Acute Care*
	Time in community programs
	Use of seclusions/restraints*
	Level of service appropriate to needs of individual*
	Needs based funding and spending*
	Consumer perception of appropriateness*
	Availability of community services
	Criminal justice system involvement
	Community/institutional balance
Continuity	Continuity mechanisms*
	Emergency room use*
	Community follow-up after hospitalization*
	Documented discharge plans*
	Cases lost to follow-up*
	Clear, visible points of accountability*
Effectiveness	Community tenure*
	Mortality*
	Criminal justice system involvement*
	Clinical status*
	Functional status*
	Involvement in meaningful daytime activity
	Housing status*
	Quality of life*
	Physical health status

Domain	Quality Indicator
Competence	Resources available to train staff to meet required competencies
	Resources available for on the job development and learning
	Meets professional certification/professional standards
Efficiency	Mental health spending per capita*
	Proportion of staff funding spent on administration and support*
	Needs-based resource allocation strategy*
	Community/institutional spending balance*
	Resource intensity tool*
	Utility costs/costs per client*
	Budget for performance monitoring*
Safety	Complications associated with ECT*
	Medication errors/side effects*
	Critical incidents*
	Suicides*
	Homicides*
	Involuntary committal rate
	Risk management practiced
	Identify research/practices to reduce adverse events and errors

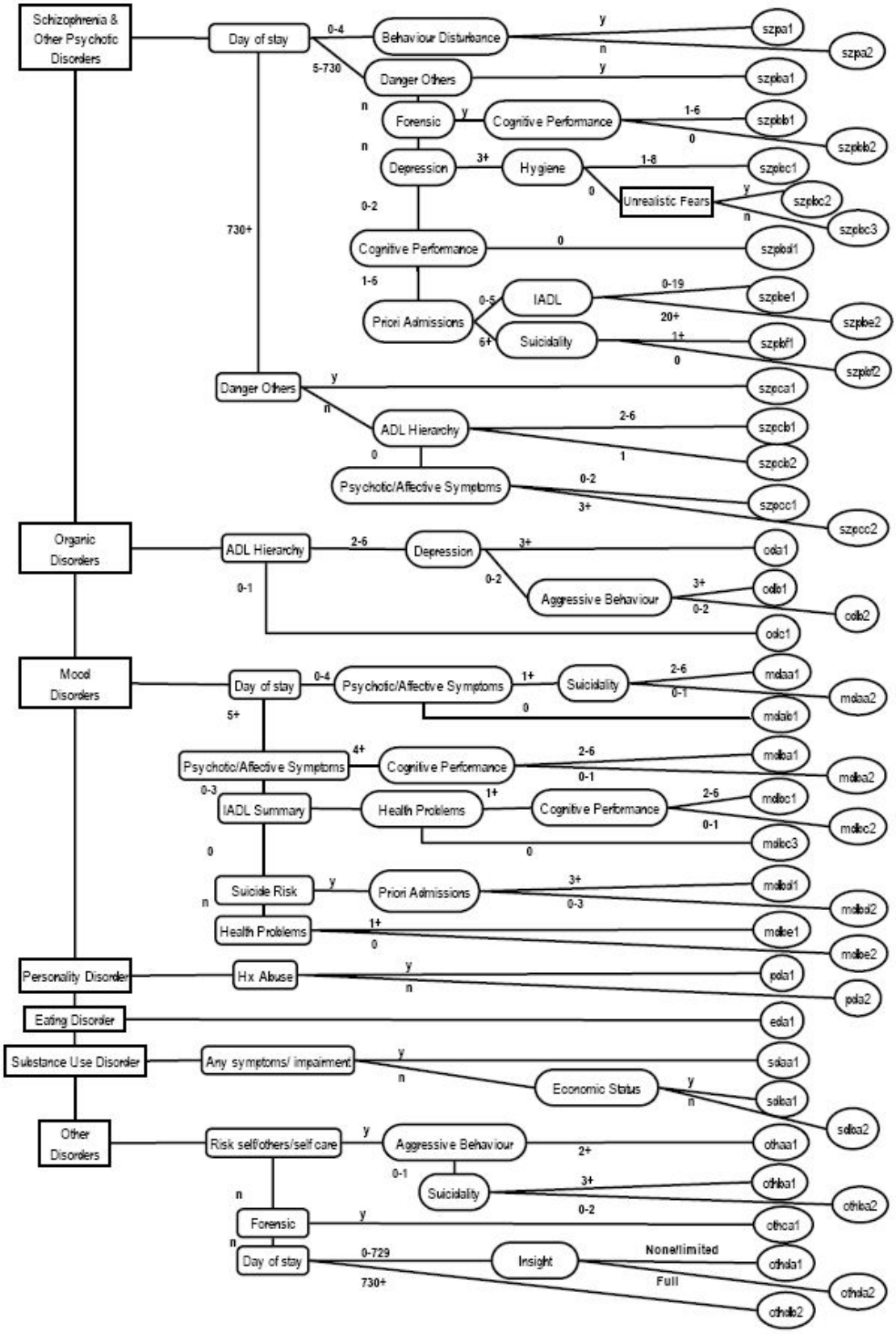
Appendix D: The Resident Assessment Instrument-Mental Health

Please visit http://catalog.interrai.org/catalog for information on how to obtain a copy of the RAI-MH

Appendix E: Mental Health Assessment Protocols and Clinical Assessment Protocols for the RAI-MH

Mental Health Assessment Protocols	RAI-MH Clinical Assessment Protocols
Violence	Interpersonal Conflict
Self Harm	Harm to Others
Abuse by Others	Social Relationships
Criminal Activity	Pain
Self-Care	Control Interventions
Social Functioning	Substance Abuse
Interpersonal Conflict	Traumatic Life Events
Vocational Rehabilitation	Sleep
Support Systems	Vocational Rehabilitation
Economic Status	Financial Issues
Adherence	Smoking
Psychotropic Drug Review	Self-harm
Physical Restraint and Seclusion	Physical Wellness
Acute Control Medications	Criminal Activity
Revolving Door	Medication Management
Discharge Resources	Falls
Addictive Behaviours	Nutrition
Nutrition	Rehospitalization
Dehydration	Self-Care/Decision Integrity
Polydipsia	Support Systems
Skin and Foot Conditions	
Oral Health	
Pain	
Bladder/Bowel Functioning	
Cognition	
Communication Disorders	
Behaviour Disturbance	
Decision Integrity	

Appendix F: System for Classification of Inpatient Psychiatry (SCIPP) Classification Algorithm



Appendix G: JPPC Domains, Indicators, and Potential Risk Adjusters for Measuring Quality of Mental Health based on the RAI-MH.

Domain	Indicator	Potential Risk Adjusters
Behavioural/ Emotional Patterns	1. Remission rate of symptoms of depression	Any Physical or Medical Comorbidity
	2. Incidence of symptoms of depression	Any Physical or Medical Comorbidity
	3. Remission rate of aggressive behaviour disturbance	Cognitive Impairment or Command Hallucinations
	4. Incidence of aggressive behaviour disturbance	Cognitive Impairment or Command Hallucinations
	5. Remission rate of disruptive behaviour disturbance	Cognitive Impairment
	6. Incidence of disruptive behaviour disturbance	Cognitive Impairment
	7. Prevalence of violent behaviour	None
Cognitive Patterns	8. Remission rate of hallucinations	Schizophrenia Diagnosis
	9. Incidence of hallucinations	Schizophrenia Diagnosis
	10. Improvement in cognitive impairment	Stroke, Brain Injury, Dementia
	11. Incidence of cognitive impairment	Stroke, Brain Injury, Dementia
Nutrition/Eating	12. Incidence of weight loss	Eating Disorder
	13. Incidence of weight gain	Eating Disorder
	14. Prevalence of dehydration	None
Physical Functioning	15. Improvement of ADL functioning	Cognitive Impairment, Physical or Neurological disorders, Chronic physical problem
	16. Incidence of ADL functioning	As above
	17. Improvement in financial management IADL	Cognitive Impairment, incapable of managing finances, legal guardian or substitute decision maker
	18. Incidence in financial management IADL	As above
	19. Improvement in medication management IADL	As above
	20. Incidence in medication management IADL	As above
Clinical Management	21. Prevalence of extrapyramidal symptoms	Parkinson's Disease or Stroke
	22. Prevalence of rehospitalization	None
	23. Prevalence of unauthorized leaves of absence	None
	24. Prevalence of inpatient suicide attempts	None

Domain	Indicator	Potential Risk Adjusters
	25. Prevalence of self-injury (non-suicidal)	None
	26. Prevalence of pain without analgesic use or pain management	None
	27. Prevalence of smoking/tobacco addiction without an offer of therapy	None
	28. Prevalence of signs of substance abuse without therapy	None
	29. Prevalence of psychotropic medication underuse	Patient has stopped taking meds due to side effects
	30. Prevalence of fire setting	None
Restraint Use	31. Prevalence of chemical restraint use	Violence in last 7 days
	32. Prevalence of physical restraint use	Violence in last 7 days
	33. Prevalence of seclusion room use	Violence in last 7 days
Sexual Violence	34. Prevalence of sexual violence (perpetrator)	None
Accidents	35. Prevalence of falls	None

Adapted from:

Joint Policy and Planning Committee. *Resident Assessment Instrument-Mental Health (RAI-MH): Quality Indicators for Mental Health (QIMHs).* 1999.

Appendix H: Definitions and Interpretations of Scales Embedded in the RAI-MH.

ACRONYM	SCALE	ITEMS IN THE SCALE	INTERPRETATION
ABS	**Aggressive Behaviour Scale (ABS)** Measure of aggressive behaviour.	E1b - Verbal abuse E1c - Physical abuse E1c - Socially Inappropriate/ disruptive E1d - Resists care	Scores range from 0 to 12. 0 = no signs of aggression 1-4 = mild to moderate aggression 5+ = more severe aggression
ADLH	**Activities of Daily Living (ADL) Hierarchy Scale** Measure of functional performance, reflecting a person's ability to carry out activities of everyday living	G1a - Personal hygiene G1b/c - Locomotion G1d - Toilet Use G1e - Eating	Scores range from 0 to 6: 0 = independent 1 = supervision required 2 = limited impairment 3 = extensive assistance required; level 1 4 = extensive assistance required; level 2 5 = dependent 6 = total dependence
CAGE	**CAGE – Substance use screener**	C4a - Felt need to **C**ut down on substance use C4b - **A**ngered by criticisms from others C4c - **G**uilt about substance use C4d - **E**ye-opener (drinking/using substances in the morning)	Scores range from 0 to 4. A score of 2 or more is considered to indicate a potential problem with substance addiction
CHESS	**Changes in Health, End stage disease, and Symptoms Scale** Reflects a person's health instability.	F4 - Cognitive decline in last 90 days G4 - ADL decline in last 90 days I1c - Shortness of breath N2c - Dehydration/insufficient fluid N2a - Weight loss N2d - Decrease in amount of fluid/food I1q - Edema I1j - Vomiting	Scores range between 0 and 4, where: 0=Not at all unstable 4=Highly unstable
CPS	**Cognitive Performance Scale (CPS)** Describes the person's cognitive status	F2 - Daily decision-making F1a - Short-term memory H3 - Expression (i.e., making self understood) G1e - Self-performance in eating	Scores range from 0 to 6: 0 = intact 1 = borderline intact 2 = mild impairment 3 = moderate impairment 4 = moderate to severe impairment 5 = severe impairment 6 = very severe impairment

ACRONYM	SCALE	ITEMS IN THE SCALE	INTERPRETATION
DRS	**Depression Rating Scale (DRS)** Describes the mood status of an individual.	B1d - Negative statements B1dd - Persistent anger B1p - Unrealistic fears B1cc - Repetitive health complaints B1o - Repetitive anxious complaints B1a - Sad, worried facial expression B1b - Crying or tearfulness	Scores may vary between 0 and 14. 3+ indicative of possible depression 6+ indicative of more severe depression.
DSI	**Depressive Severity Index (DSI)**	B1a - Sad, pained facial expressions B1d - Negative statements B1e - Self-deprecation B1f - Guilt/shame B1g - Hopelessness	Scores range 0-15. Higher scores indicate more depressive symptoms.
IADL	**Instrumental Activities of Daily Living (IADL) Capacity Scale** Estimate of higher-level function, reflecting others' perception of a person's ability to carry out IADLs	G2a - Meal preparation G2d - Managing finances G2b - Managing medications G2e - Phone use G2c - Transportation G3 - Stamina	Scores range from 0 to 33. Higher scores indicate less capacity.
PAIN	**Pain Scale** Summarizes the presence and intensity of pain.	I8a - Pain frequency I8b - Pain intensity	Scores may range between 0 and 4: 0 = No pain 1 = Less than daily pain 2 = Daily pain but not severe 3 = Daily severe pain 4 = Daily excruciating pain
PSS	**Positive Symptoms Scale – Long Form (PSS)**	B1u - Hallucinations B1v - Command hallucinations B1w - Delusions B1x - Abnormal thought process B1h - Inflated self-worth B1i - Hyper-Arousal B1l - Pressured speech B1ee - Abnormal/unusual movements	Scores range from 0-24 Higher scores indicate greater number and frequency of psychotic symptoms.

ACRONYM	SCALE	ITEMS IN THE SCALE	INTERPRETATION
RHO	**Risk of Harm to Others (RHO)** Reflects risk of harm to others.	D2c - Violent thoughts D1a - Violent actions D2b - Violent ideation B1w - Delusions B2 - Insight into mental health B1gg – Sleep Problems PSS_Long Aggressive Behaviour Scale (ABS)	Scores range from 0-6. Higher scores indicate increased risk of harm to others.
SCI	**Self Care Index (SCI)** Reflects risk of inability to care for self due to psychiatric symptoms.	F2 - Daily decision making PSS_Short (1st 4 items of PSS_long) B2 - Insight into mental health B2c - Decreased energy B1x - Abnormal thought process H3 - Difficulty making self understood	Scores range from 0-6. Higher scores indicate decreased ability to care for self due to psychiatric symptoms.
SOS	**Severity of Self-harm (SOS)** Reflects risk of harm to self.	D1a - Considered self-injurious act D1b - Intent of any self-injurious act was to kill self Depressive Severity Index (DSI) D1d - Family/others concerned about person's risk for self-injury PSS_Short Cognitive Performance Scale (CPS)	Scores range from 0-6. Higher scores indicate increased risk of self-harm.

Appendix I: Mental Health and Quality Expert Consultation Participants

Name	Occupation	Affiliation
Brant Fries, PhD	Professor, Health Management an Policy, Senior Research Scientist	University of Michigan, Ann Arbor, MI
	Chief of Health Systems Research	Geriatric Research, Eduation, and Clinical Center at Ann Arbor VA Medical Centre, Ann Arbor, MI
	President	InterRAI
John Hirdes, PhD	Professor, Department of Health Studies and Gerontology	University of Waterloo, Waterloo, ON
	Scientific Director	Homewood Research Institute, Guelph, ON
	Member	InterRAI
John Morris, PhD	Co-director of Research and Training	Hebrew Rehabilitation Centre for the Aged, Boston, MASS
	Member	InterRAI
Terry Rabinowitz, MD, FAPM	Associate Professor of Psychiatry and Family Practice	University of Vermont College of Medicine, Burlington, VT
	Director	Psychiatric Consultation Service, Fletcher Allen Health Care, Burlington, VT.
Trevor Smith, PhD	Assistant Professor, Department of Sociology	Nippising University, North Bay, ON
	Member	InterRAI

Appendix J: Results from the randomization of the OMHRS data into test and replication data.

Characteristic	Test Data		Validation Data	
	%	n	%	n
# of hospitals		70		70
Age				
< 25	11.6	1753	11.9	1797
25 to 44	40.4	6093	40.7	6146
45 to 64	34.7	5238	33.5	5056
65 or more	13.3	2006	13.9	2097
Male	49.7	7502	49.8	7521
Designated Patient Type				
Acute	79.4	11982	78.8	11899
Long Stay	13.2	1992	13.0	1970
Geriatric	4.3	655	4.8	727
Forensic	3.1	469	3.3	502
Diagnoses				
Mood	51.0	7705	51.0	7701
Dementia	6.8	1023	7.1	1073
Schizophrenia/psychosis	36.8	5560	36.5	5513
Substance Use Related	25.7	3887	26.5	4006