NOTE TO USERS

uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Adrian Dalloo
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Epidemiology)
GRADE / DEGREE

Department of Epidemiology and Community Medicine
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Can the Canadian Case Mix Age Group (0-17 Yrs) Be Further Refined To Better Represent The Pediatric Population?

TITRE DE LA THÈSE / TITLE OF THESIS

D. Coyle
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

R. Spasoff
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

J. Grimshaw

D. Moher

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Can The Canadian Case Mix Age Group (0-17 Yrs) Be Further Refined To Better Represent The Pediatric Population?

Adrian R. Dalloo

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the M. Sc. Degree in Epidemiology

Department of Epidemiology and Community Medicine
University of Ottawa

Your file Votre référence
ISBN: 0-494-11248-4
Our file Notre référence
ISBN: 0-494-11248-4

Abstract

The Canadian Case Mix Groups or CMG™ methodology groups acute care patients into clinically similar and resource homogenous groups, and further stratifies patients into three age categories (years): 0-17, 18-69 and 70+. Some healthcare practitioners believe that the 0-17 years CMG age group is too broad, and does not reflect age differences associated with hospital resources consumption. A methodological study was conducted, using linked Canadian inpatient activity and case cost data from 1997/98 to 2000/01, to examine whether or not the 0-17 years CMG age group can be further refined to better represent age differences in hospital resources consumption. The study utilized the Classification and Regression Tree (CART) algorithm and regression analyses to develop new age groups. Of the 123 CMG included in the study, CART recommended further age splits for at least 48% of the CMG. The study recommends that, subject to clinical validation, fixed age splits at 0.5 and 1.5 years of age could be applied across all CMG.

Acknowledgements

Doug Coyle provided excellent thesis supervision. His support, patience and encouragement kept me motivated throughout every aspect of this thesis. Bob Spasoff provided exceptional guidance with helping me to develop and clearly articulate my ideas. Keith O'Rouke provided much statistical expertise and advice, especially with CART. My colleague, Craig Homan, advised me of the problem with the pediatric age group in the current ICD-9-based CMG grouper and inspired me to make exploring this issue my thesis topic. Without Craig's clinical insight and knowledge, this study would not have happened. The Canadian Institute for Health Information (CIHI) provided access to the hospital activity and cost data needed for this thesis at no charge, through their Graduate Student Data Access Program. Andre Lalonde (CIHI) facilitated access to computer resources at CIHI required to complete the analyses, and was very supportive; he even proofread several drafts. CIHI's Case Mix Team answered many questions about the CMG methodology. To Grace Carter (RAND Corporation) and Jim Georgoulakis (APS Consulting), your expertise in the area of patient grouping methodologies was invaluable. I learned a lot from you all.

To my parents (Joseph, Hilda and Carmine), thank you for giving me the support I needed to see this thesis through to the end. I thank all my friends and family for their support and words of encouragement. Even "Sasha" kept me company while I worked many late nights. To my wife Laura, thank you for being my inspiration. You helped me to stay positive, focused and keep everything in perspective. I couldn't have succeeded without you.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1: Development of Patient Grouping Methodologies

1.1 Introduction

Case Mix is the practice of describing hospital patients with similar diagnoses and/or treatment requirements or patient mix within a specific health care setting with respect to length of stay and cost. To accomplish this task, patient grouping methodologies were developed and have since been implemented around the world to assist clinicians, hospital administrators and decision-makers in understanding and planning for the distinct mix of patients being seen at their hospital. In Canada, the Case Mix Groups or CMG™ represent the acute care inpatient grouping methodology used throughout most of the country. CMG are used to group inpatient records into clinically similar groups (based on the most responsible diagnosis and/or principal procedure) and/or groups that are homogenous with respect to resource consumption.[1-4]

Once patient records are grouped to the appropriate CMG, a Complexity Overlay (or PIx™) is applied to account for any secondary diagnoses that may significantly increase a patient's length of stay or may result in more costly treatment.[1-4] After complexity is applied, these records may then be partitioned into one of three age categories: 0-17 years, 18-69 years, and 70+ years. For each CMG, complexity level and age category (or APIx) cell, an Expected Length of Stay (ELOS) statistic and Resource Intensity Weight (or RIW™) are calculated to describe the patient's consumption of hospital resources for their given medical condition. The ELOS statistic represents the amount of time a patient can expect to stay in hospital for a given disease, condition or treatment, and the RIW describes the expected amount of

---

™Registered Trade-mark of the Canadian Institute for Health Information.

resources consumed by the "average" patient within a CMG, complexity level and age group.[1,2]

However, some healthcare providers who work closely with children believe that the 0-17 years CMG age group is too broad.[5,6] Critics point out that the 0-17 year age range does not adequately represent the pediatric population, especially with respect to age differences associated with the length of stay in hospital and the cost of treatment and care for certain medical conditions, such as respiratory or gastrointestinal diseases. This thesis will, therefore, examine whether or not the 0-17 years CMG age group can be further refined to better represent the level of resources consumed by pediatric patients of different ages. The importance of this question will be described in this chapter in further detail. Specifically, this chapter will highlight the:

i) Development of patient grouping methodologies;

ii) Development of the CMG grouping methodology;

iii) Age splits used by the different inpatient grouping methodologies; and

iv) Objectives proposed for this thesis.

1.2 Background

Much of the following overview on patient grouping methodologies, including the development of the CMG grouping methodology, has already been published in a background paper I authored for the Canadian Institute for Health Information entitled, *"Acute Care Grouping Methodologies: From Diagnosis Related Groups to Case Mix Groups Redevelopment: Background Paper for the Redevelopment of the CMG/PIx Grouping Methodology Using ICD-10-CA/CCI Classification Systems."*

A patient grouping methodology is a system that describes discrete clusters of patient types. More specifically, it is a way of relating the type of patients a hospital treats to the costs incurred by that hospital.[7,8] Acute Inpatient cases are categorized based on:

i) Clinical criteria (e.g., primary diagnosis, comorbid diagnoses, procedures);

ii) Demographic information (e.g., age, sex); and

iii) Resource consumption (e.g., costs, length of stay).

Depending on the philosophy of the grouping methodology, the end result is groups of cases that are clinically similar and/or homogenous with respect to resource use.

Grouping methodologies were initially developed as a cost management tool to help clinicians and hospitals monitor quality of care and utilization of services.[9] Most recently, groupers are utilized for a variety of purposes including epidemiological monitoring, clinical management, standardized comparison of hospital activity, hospital budgeting and program planning, hospital funding and reimbursement, and as a prospective payment system. However, no one grouper can be used to do all of these things well. In fact, most groupers have been designed for one purpose (i.e., to measure hospital performance), but then have been used by those who work in health care management to meet other needs (i.e., as a payment/reimbursement/funding tool). Multi-purpose groupers are difficult to develop and maintain, and may not provide desired results.[10]

In the last 25 years, groupers have been developed using two main approaches: clinical input and statistical analyses. Groupers built on clinical input from the medical community only used medical criteria to split cases. Medical criteria sometimes included

data elements not routinely collected, and often resulted in too many terminal cells.[7] On the other hand, groupers based solely on statistical analyses, such as clustering, factor analysis, regression, or decision trees, often resulted in terminal groups that did not make sense clinically since they only used measures of resource consumption as the principle splitting criteria.[7]

Those that work in the area of grouper development have since recognized that the development of a practical grouper requires combining these two main approaches.[11] As a result, several basic criteria have been identified as essential for grouper development. Grouping methodologies must limit data elements to routinely collected data, generate a manageable number of possible categories, demonstrate some degree of clinical coherence, demonstrate statistical homogeneity with respect to either length of stay or total resource use, and be dynamic.[7]

The most common approach to grouper development consists in adapting an existing system to country-specific requirements.[11] This is most evident with the Diagnosis Related Groups (DRG) system, which has been adopted and/or modified for use by many countries worldwide, or used as the foundation for building a country-specific patient classification system.

The development of the DRG first began in the United States in 1967.[9] With the introduction of Medicare, all hospitals were required to implement a utilization review and quality assurance program to monitor utilization of services and quality of care in order to receive Medicare funding.[9] A group of physicians from a local hospital in Connecticut approached Dr. Robert B. Fetter and his colleagues at Yale University for help in measuring and evaluating their hospital's performance.[9]

In developing an appropriate hospital management tool, Fetter and his team were faced with several major challenges. The final product had to include all hospital services, incorporate thousands of diagnoses and procedures, account for multiple diseases and treatments of individual patients, differentiate between high- and low-cost care, and develop clinically meaningful categories all within a reasonable number of groups. In the years to follow, the DRG system emerged, and several versions of this patient classification system were developed using the International Classification of Diseases, Eighth Revision-Adapted (ICDA-8), the Hospital Adaptation of the International Classification of Diseases-Adapted, Second Edition (H-ICDA-2) and Commission on Professional and Hospital Activities (CPHA) classification systems.[12] Between 1980 and 1982, the International Statistical Classification of Diseases, Injuries, and Causes of Death, Ninth Revision - Clinical Modification (ICD-9-CM) version of the DRG system was created. New Jersey was the first to adopt and use the DRG system as a prospective payment system (PPS).[7,8]

The DRG methodology developed at Yale University followed the ICD-9 system's organ-system approach and divided cases into 23 mutually exclusive groups called Major Diagnostic Categories (MDC).[9,13] Within each MDC, cases were then subdivided into discrete patient clusters using recursive partitioning. Fetter et al. used secondary diagnosis, principal procedure, sex, age, discharge status, complications and comorbidities (as per a standard list), in addition to principal diagnosis, to classify cases into clinically cohesive groups with similar length of stay patterns and/or hospital resources consumption.[9,13] Subsequent DRG systems used all operating room procedures, high cost procedures normally done outside the operating room, and birth weight for neonates as part of the grouping algorithm.

The newly created DRG system, however, was fraught with problems.[14] To begin with, many felt the DRG themselves were not clinically meaningful since they included regional or organ-specific procedures, or were defined based on medical problem, signs and symptoms, and/or treatments.[14] In addition, the DRG system could not accurately capture the severity of illness of patients, relative weights were based on unreliable data, and the system was not viewed as being dynamic to keep up with changes in medical treatment and technology.[14] Despite these limitations, Canada adapted the ICD-9-CM-based DRG system in 1983 to accommodate their ICD-9 and Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures (CCP) classification systems.[15]

1.2.1 Canadian Case Mix Groups (CMG)

The CMG methodology is the Canadian equivalent of the DRG system and "was designed to aggregate acute care inpatients with similar clinical and resource utilization characteristics."[1] Acute care inpatients are grouped based on clinical and administrative data collected in the Discharge Abstract Database (DAD) held at the Canadian Institute for Health Information. The creation of a Canadian grouper stemmed from the fact that those in health care management wanted:

i) To improve the comparability of national health care data;

ii) To enhance the relationship between diagnoses and length of stay, especially secondary diagnoses that contribute to longer lengths of stay; and

iii) To provide a tool for utilization management based on Canadian health care data.[15]

Since the CMG system was a direct adaptation of the DRG system, it shared the same body system approach as its first step in dividing cases. In fact, the Major Clinical Categories (or MCC) in the CMG system are the same as the Major Diagnostic Categories (or MDC) in the DRG system (see Appendix B). However, the similarities stopped there, as different criteria were used to further subdivide cases. To begin with, DRG assignment is driven by principal diagnosis (PDx), whereas CMG assignment is driven by most responsible diagnosis (MRDx). This represents the most significant difference between the two systems as the MRDx attempts to identify the diagnosis that can account for the greatest proportion of a patient's length of stay versus PDx or the admitting diagnosis.

The next major difference between these two systems is with respect to how comorbidities and complications are treated. CMG uses diagnosis type (a flag to denote whether a condition was diagnosed pre- or post-admission) and a diagnosis grade list (a statistically defined list of comorbidities that may increase a patient's length of stay by at least one day) to identify other secondary diagnoses impacting length of stay and/or where more costly treatment might be reasonably expected. This interaction led to the development of PIx, and reflects how complicated a given case is to treat. In contrast, DRG uses pre-defined complications and comorbidities (CC) tables that have distinct severity levels (i.e., minor, moderate, major) assigned to a selected group of secondary diagnoses. This measure, however, may not acknowledge significant post admission comorbidities and only uses the secondary diagnosis with the highest severity level.

Finally, patient length of stay was used to create the CMG, whereas charge data were used in the creation of the DRG. In the end, the CMG and DRG systems differ with respect to the number of terminal cells (See Appendix C). The CMG methodology has a

total of 478 base-CMG groups, whereas the DRG system has between 321 and 367 base-DRG groups.

The CMG system has evolved over time. Many of its developments are highlighted in Table 1 below.

Table 1: Case Mix Groups Evolution

Year	Description
1983	Adapted the DRG system to accommodate ICD-9/CCP
1987	Mapped CMG structure back to ICD-9-CM
1991	Expert team established to ensure CMG reflected Canadian requirements and hospital practice patterns
1992-1997	Began modifications to selected MCC: 2-8, 11-15, 19, 24-25
1997	Removal of CC and age splits Introduction of PIx and Age Adjustment
2000-2001	Backward conversion of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision – The Canadian Enhancement (ICD-10-CA) to ICD-9
2003	Revised the diagnosis grade list to address variations in coding practices Initiated CMG Grouper Redevelopment using the ICD-10-CA and Canadian Classification of Health Interventions (CCI) classification systems

The CMG methodology utilizes the patient's MRDx to assign the case to one of 25 MCC, with the exception of patients less than 29 days old who are grouped based on birth weight.[2-4] A MCC identifies either a body system (e.g. Digestive System) or other specific types of clinical problems (e.g. Neonates, HIV Infections or Burns).[2-4] Within the MCC, the case is then assigned to the most appropriate CMG within the medical or surgical branch based on the presence or absence of an operative procedure.[2-4]

Because the resource requirements may vary significantly from one patient to the next within a given CMG, a Complexity Overlay and Age Adjustment are added to the CMG methodology. The PIx overlay uses the presence of co-morbid conditions to account for variations in length of stay and resource use, whereas the age adjustment accounts for differences due to the patient's age. As a result, these improved "estimates are sensitive to patient characteristics such as: the type and number of co-morbid conditions, the time of onset of co-morbid conditions, conditions affecting multiple body systems, and the patient's age group."[1]

The CMG age adjustment attempts to draw attention to the fact that there are differing health care needs and costs associated with caring for children compared to adults and the elderly. By convention adopted from the DRG system, the CMG methodology groups acute care inpatients into one of three age categories: 0-17 years, 18-69 years and 70+ years. Despite these distinct groupings, some healthcare providers who work closely with children feel that the 0-17 yrs age group does not adequately represent the pediatric population, especially with respect to their medical condition, the severity of that condition, length of stay in hospital and the cost of care.[6]

1.2.2 International Inpatient Grouping Methodologies and Age Splits

Over the years, several variations, modifications and improvements were made to the initial DRG system.

1.2.2.1 Health Care Financing Administration-DRG (HCFA-DRG)

Despite the shortcomings of the DRG system, the Center for Medicare and Medicaid Services (CMS) formerly the Health Care Financing Administration (HCFA) at the Department of Health and Human Services in the United States adopted the DRG system in 1983 as a Medicare PPS for hospitals.[8] This unprecedented move was "the start of a new method of payment intended as a national price for a hospital stay based on the reason for the hospital stay."[14] CMS (formerly HCFA) assumed responsibility for annual updates to the DRG system, but modifications focused only on problems relating to the elderly and disabled populations. As a result, many DRG were split at age 65.

1.2.2.2 Refined-DRG (R-DRG)

Several years after the implementation of the HCFA-DRG system, HCFA recognized that the presence or absence of complications and comorbidities (CC) resulted in certain types of patients being assigned to a different DRG.[8] The HCFA-DRG system defined a CC as a secondary diagnosis that significantly increases hospital resource use. Wanting to change the use of CC, HCFA funded a project at Yale University during the mid-1980's to help address this issue and refine the DRG methodology.

The project mapped all CC-related diagnoses into 136 secondary diagnosis groups, where each was assigned a CC complexity level that was disease and procedure specific.[8,16] For surgical patients, each secondary diagnosis group was assigned to one of four levels: non-CC, moderate-CC, major-CC and catastrophic-CC. For medical patients, each secondary diagnosis group was assigned to one of three levels: non-CC, major-CC and catastrophic-CC. If several CC were listed, the refined-DRG (R-DRG) grouper took the highest-level secondary diagnosis. The presence of multiple CC at one level did not result in grouping to a higher-level subgroup. All age and CC splits from the original DRG system were removed and replaced with these medical/surgical subgroups.

1.2.2.3 All Patient-DRG (AP-DRG)

Building on the success of HCFA using the DRG system as PPS, New York State passed legislation to use the DRG system as a PPS for all non-Medicare patients in 1987.[8] As a result, the New York Health Department (NYHD) had to review the applicability of the HCFA-DRG system for a non-Medicare population, and evaluate it for neonates and those infected with HIV. The NYHD concluded that the HCFA-DRG system was not adequate for the non-Medicare population nor were there any provisions for the neonate or HIV-infected populations.[8]

The NYHD contracted 3M Health Information Systems (3M HIS) to modify the HCFA-DRG system for the non-Medicare population.[8] 3M researched and developed all necessary modifications, and included the Pediatric Modified Diagnosis Related Groups (PM-DRG) developed by the National Association of Children's Hospitals and Related Institutions (NACHRI), and introduced MDC 24 for HIV infection patients. The CC list

was further revised, and MDC 25 was added to capture multiple traumas. In addition, modifications were added for transplants, long-term mechanical ventilation, cystic fibrosis, nutritional disorders, high-risk obstetric care, acute leukemia, hemophilia and sickle cell anemia.[8] DRG were split on age only if the DRG demonstrated that resource consumption varied by age. As a result, some DRG that warranted a pediatric age split were limited to one age group; 0-17 years to differentiate all children (i.e., less than 18 years of age) from adults (i.e., 18 years of age and older).[8,17]

1.2.2.4 All Patient Refined-DRG (APR-DRG)

The All Patient Refined Diagnosis Related Groups (APR-DRG) are widely used throughout the United States, Europe and selected parts of Asia. Using the base structure of the AP-DRG system, 3M HIS added four subgroups to each group in an attempt to better describe a patient's severity of illness.[8] This refinement resulted in a significant change to the grouping logic. All age and CC distinctions were removed and replaced with two groups: one to describe severity of illness (SOI), and the other to describe the risk of hospital mortality (ROM).[8]

Both the severity and mortality groups contained four subgroups: minor, moderate, major and extreme. With these additions, a case was now assigned three distinct descriptors: i) the base-DRG; ii) the SOI subgroup; and iii) the ROM subgroup.[8] Subgroup assignment is based on interaction between secondary diagnoses, age, principal diagnosis, and the presence of certain non-operative procedures. Some non-CC in previous DRG systems were now moderate-, major- or extreme-CC, and vice-versa, and multiple CC were now recognized.[8] In addition, a completely new set of DRG was developed for the neonatal MDC (i.e., those less than 29 days).

1.2.2.5 International Refined-DRG (IR-DRG)

The International Refined Diagnosis Related Groups (IR-DRG) were created in response to the international community not being able to develop their own country-specific grouper.[18] To fill the international void, 3M HIS built the IR-DRG system using the same logic, structure and age splits as the AP-DRG and APR-DRG systems. It incorporates the same severity of illness adjustment using secondary diagnoses, but uses only three subgroups: without CC, with CC and with major-CC. The IR-DRG does not recognize multiple CC since 3M HIS discovered that most international clinical datasets do not contain more than two secondary diagnoses.[18] In addition, several DRG eliminated from U.S. versions of the DRG system were re-introduced to capture those outpatient procedures in the U.S. that are still being performed in the inpatient setting in other countries.[18]

The most unique aspect of the IR-DRG is the underlying coding classification system. The base-DRG were intended to be compatible with both ICD-9-CM and ICD-10 without any mapping between coding systems.[18] Therefore, at least theoretically, cases could be grouped to the same IR-DRG regardless of the coding classification system used. As a result, the IR-DRG system could accommodate country-specific coding modifications and procedure coding systems. IR-DRG Version 2.0 is currently under development, and will be procedure driven in order to group all types of inpatients and outpatients.[18] Information on age group refinement is not available.

1.2.2.6 Australian Refined Diagnosis Related Groups (AR-DRG)

Australia was one of the first countries to create a country specific version of the DRG system. After using the AP-DRG for several years, Australia realized that in order for the DRG system to remain valid and meaningful for its case mix purposes the grouper needed to be customized.[19] As a result, modifications were made to the DRG system to keep pace with changes in medical, surgical and coding practices. In 1992, the Australian National Diagnosis Related Groups (AN-DRG) were released, and annual updates were made until 1996.[20] The AN-DRG system only applies age splits to categories where a single diagnosis may differ by age.[20] In this instance, the patient is grouped into one of three age groups: i) 0-17 years; ii) 18-64 years; and iii) 65+ years.

In 1997, the Australian Refined Diagnosis Related Groups (AR-DRG) were introduced to better reflect hospital conditions, advances in medical technology, the introduction of new diagnosis and procedure codes, and to address problems with clinical and resource homogeneity of some DRG.[21] A DRG in the new system may be split on a specific age only when the two age groups denoted by the split are shown to be significantly different with respect to resource consumption.

1.2.2.7 British Health Resource Groups (HRG)

In Great Britain, the Health Resource Groups (HRG) were developed primarily to reflect a fundamental change from the DRG grouping logic.[22] With HRG, cases are assigned to an HRG based on the procedure codes recorded.[23] If more than one procedure is recorded, then a procedure hierarchy is used to choose the grouping procedure. For minor procedures and non-resource intensive procedures, or when no procedure is recorded, the hierarchy will assign an HRG on the basis of the ICD-10

diagnosis codes recorded. In the majority of cases where grouping is based on diagnoses, the primary diagnosis is used to assign the HRG. Once a case is assigned to an HRG, the group may be split on age if resource consumption varies by age. The most common age splits within the HRG system occur at either age 17, 50, or 70 years.

1.2.2.8 French Effeuillage Progressif (EfP)

The French grouper, Groupes Homogènes de Malades (GHM), evolved in 1986 from the HCFA-DRG and was designed to be comparable to the AP-DRG.[24] In 1999, a modified version of the GHM system called Effeuillage Progressif (EfP) was introduced to improve its capacity to explain cost variations and medical usefulness. The modification to the initial GHM system involved removing CC splits and allowing one or more secondary GHM to be added to the base GHM.[25] These secondary GHM are determined using all secondary diagnoses classified as CC. Determining secondary GHM is an iterative process that relies on specific exclusions between diagnoses and/or GHM. The best GHM is chosen based on medical and/or economic criteria. This information is then used to add costs over and above the base cost of each primary GHM. No age splits are used in the GHM system.

1.2.3 Age Split Research In Inpatient Grouping Methodologies

After reviewing the major inpatient grouping methodologies used around the world, it is evident that age splits are applied inconsistently across different grouping methodologies and vary by clinical group within the grouping methodology (see Table 2). Within any major grouping methodology, the 0-17 year age group usually defines the pediatric age split. Furthermore, most grouping methodologies are proprietary products, and so very little or no research has been conducted and/or published in the area of age splits with inpatient grouping methodologies.

Table 2: Inpatient Grouping Methodologies and Age Splits for Selected Countries

Country	Grouping Methodology[11,26,27]	Age Groups or Age Splits Used (Years)
Australia	AR-DRG	Varies by DRG
Belgium	APR-DRG	Varies by DRG
Bulgaria	IR-DRG	Varies by DRG
Canada	CMG/Plx	0-17, 18-69, & 70+
Costa Rica	HCFA-DRG	Varies by DRG
Czech Republic	AP-DRG, IR-DRG	Varies by DRG
Denmark	Nord-DRG, Dk-DRG*	Varies by DRG
Estonia	Nord-DRG	Varies by DRG
Finland	Nord-DRG	Varies by DRG
France	GHM, EfP	None
Germany	G-DRG (AR-DRG)	Varies by DRG
Greece	HCFA-DRG	0-64, 65+
Hungary	HBC (HFCA-DRG)	Varies by DRG
Italy	HCFA-DRG, APR-DRG	Varies by DRG
Netherlands	DBC	Varies by DRG
Norway	Nord-DRG	Varies by DRG
Portugal	HCFA-DRG	Varies by DRG
Romania	HCFA-DRG, AP-DRG, IR-DRG	Varies by DRG
Spain	HCFA-DRG	Varies by DRG
Sweden	Nord-DRG	Varies by DRG
Switzerland	AP-DRG	Varies by DRG
United Kingdom	HRG, AP-DRG (Wales)	Varies b/w 17, 50 & 70
United States	HCFA-DRG, R-DRG, AP-DRG, APR-DRG	Varies by DRG

( ) grouper the country specific system is based on * grouper now in use

A recent Australian study in 1996 used pediatric nursing service weights to show that children under 3 years of age and those with congenital abnormalities or chronic illness required significantly more nursing care than older children.[28,29] In the United States, five age groups are used for the purpose of measuring payment redistribution.[8] These age groups are: i) neonates 0-30 days; ii) 1 month-17 years: with and without a chronic diagnosis split; iii) 18-64 years; iv) 65-79 years; and v) 80+ years. In Canada, the Canadian Pediatric Decision Support Network (CPDSN), a network of 8 pediatric hospitals across Canada, identified that the 0-17 year age group used by the CMG methodology is too broad a category for grouping pediatric patients.[5] CPDSN recognizes that there are significant differences in treatment and resource utilization between a very young child and a teenager with similar medical problems and/or procedures. As a result, CPDSN has suggested that the 0-17 year age group be sub-divided into: 0-28 days, 29 days-11 months, 12 months-2 years, 3-4 years, 5-9 years, 10-13 years and 14+ years, to reflect these differences.[5] To date, there is no evidence to validate these suggested age groups, and very little research is available on how pediatric patients should be grouped.

1.2.4 Thesis Objectives

The purpose of this thesis is to identify any meaningful age splits through statistical analyses, and to determine whether CIHI's 0-17 year CMG age group can be refined to better represent the level of resources consumed by pediatric patients. Specifically:

1. Identify meaningful age groupings between 0 and 17 years by applying the classification and regression tree (CART) algorithm to inpatient cost data; and
2. Validate the new age groups by applying the new age splits to an independent or 'holdout' set of inpatient cost data.

The importance of this study is three-fold:

a) To explore whether refining the 0-17 year CMG age group is practical and if so, to create new age groups that meaningfully represent the pediatric community in Canada for the purpose of calibrating national ELOS and RIW values;

b) To determine in terms of healthcare management and resource utilization that the pediatric population is different from other populations but not homogenous, and therefore, should not be considered as such; and

c) To provide information that can be used when existing case mix groupers are redeveloped to accommodate the ICD-10-CA and CCI classification systems.

The next chapter will provide an overview of the methodology used for the study, including a description of the dataset, the CART algorithm and a step-by-step approach of the analysis. Chapter 3 will present the results of the CART analysis, and Chapter 4 will discuss those results in more detail, including the limitations of the study, and make recommendations for additional research.

Chapter 2: Methodology

2.1 Introduction

In chapter 1, the 'case mix' concept was introduced, and inpatient grouping methodologies used both in Canada and around the world were reviewed. Although inpatient grouping methodologies have come a long way since their initial development, it was evident that limited information was available on how existing age group definitions were created. Some systems simply adopted age group definitions from other systems, and little research was done to improve or refine existing age categories or create new ones. Moreover, those who work in pediatric health care believe that the 0-17 year age group, used with the Canadian Case Mix Groups (or CMG) methodology, cannot be used to adequately describe the hospital resource consumption for all children within this age range. Hence, the objective of this thesis is to determine whether or not the 0-17 year CMG age group can be further refined. In this chapter, the methodology developed to answer the thesis question will be described in detail. Specifically, this chapter will outline the:

i) Study design;

ii) Data sources, including a description of the case costing data and the Discharge Abstract Database (DAD);

iii) Classification and regression tree (CART) algorithm;

iv) Advantages and disadvantages of CART versus other statistical techniques; and

v) The analysis developed to determine whether or not the 0-17 year CMG age group can be further refined.

2.1.1 Study Design

The thesis proposes to conduct a methodological study, using secondary data, to identify new age groups between 0 and 17 years that better represent and/or reflect the Canadian pediatric population in the calculation of national Expected Length of Stay (ELOS) and Resource Intensity Weights (or RIW) values. The study will utilize the CART algorithm, a form of binary recursive partitioning, to develop new age groups using Canadian inpatient cost data for the 0-17 year age group from fiscal year 1997/98 to 2000/01 (approximately 70,000 records). These cost data will be randomly divided (40/30/30 percent) into three groups: i) a 'learn' dataset for growing the largest tree possible; ii) a 'validate' dataset to prune the largest tree and select the optimal sub-tree based on the reduction in variance in the dependent variable; and iii) a 'test' dataset to be used as a holdout sample to assess the optimal sub-tree. Using the 'learn' and 'validate' datasets, between 2 to 8 age partitions will be modelled based on knowledge of the dataset and practicality of implementing new age splits. The new age splits will be validated using the 'test' dataset.

2.1.2 Data Sources

Canadian inpatient case cost and activity data required for this project were available from the Canadian Institute for Health Information (CIHI), and were obtained at no cost through CIHI's Graduate Student Data Access Program (GSDAP).[30] A brief description of the cost and activity data is outlined below.

2.1.2.1 Acute Inpatient Case Costing Data

For the purpose of calculating annual RIW, CIHI receives limited case costing data from the province of Ontario, Alberta and British Columbia.[1,2] Within these provinces (except British Columbia), institutions that participate in case costing initiatives submit cost data to their respective Ministries of Health. Case costing initiatives in Canada were established in an attempt to get a better sense of what it costs to treat patients who are admitted to and treated within acute care facilities.[31] However, these systems are extremely expensive to implement and maintain, and so case costing initiatives have been limited to few (i.e., less than 30) acute care facilities across Canada.

Within a specific acute care facility, case costing data can capture direct, indirect and total costs associated with each patient visit. Direct costs include workload measurement derived costs, such as costs related to nursing services, operating and recovery room, and allied professional services (e.g. Physiotherapy, Occupational Therapy, Respiratory Therapy, Social Work); laboratory tests and diagnostic imaging costs (based on the average cost per procedure); actual costs of traceable drugs and supplies; and other costs, such as equipment used, referred-out expenses and compensation (not including physician billings).[2,32,33] Indirect costs, on the other hand, include costs related to administration, human resources, systems support, communications, material management, volunteer services, plant operation, security and maintenance, housekeeping, laundry, food services, biomedical engineering, registration, case management, patient transport and health records.[2,32,33] These costs are usually calculated as a percentage of the departmental operating budget in proportion to the direct costs. It is important to note that case costing methodologies

vary between Ontario, Alberta and British Columbia, and even between facilities within these provinces with respect to what information is actually captured as direct and indirect costs. For example, one facility may capture food services as a direct cost, while another facility may capture it as an indirect cost. However, such differences should not significantly affect the overall or total cost associated with the patient visit.

Many Canadian health care facilities use a functional accounting system to capture case costing data. A functional accounting system allows for direct expenses (budgeted and actual), workload statistics and revenues for specific activities or services to be recorded in specific departmental sub-accounts or functional centres.[32] Therefore, functional centres capture costs related to treatment and care for each day a patient remains in hospital. All costs associated with the patient's stay in hospital are then rolled up or summed by functional centre. These functional centre-level costs are then used to determine direct, indirect and total costs for each patient.

Case costing data are submitted to CIHI separately from their associated activity data, and are linked using several variables as a unique key. Once linked, the record provides a complete picture of the clinical, demographic and administrative information associated with the case costing data. At time of the data request, case costing data were available for fiscal years 1997/98 through to 2000/01.

2.1.2.2 The Discharge Abstract Database

In 1963, the Hospital Medical Records Institute (HMRI), a founding member of CIHI, developed the Discharge Abstract Database (DAD) to collect data on hospital discharges within the province of Ontario.[34] Since then, the DAD has expanded to

provide national coverage of hospital discharges in Canada. In addition to data collection and processing of hospital discharges, the DAD provides submitting facilities with hospital-specific reports, national comparative reporting based on peer groups, and support to facility-specific clinical research.[34] These data are also used to evaluate patient expected length of stay and resource consumption, as defined using the CMG/Complexity Overlay (or PIx) and RIW methodologies.

The DAD contains clinical, demographic and administrative information on hospital patient discharges including among others; most responsible physician/diagnosis, principal procedure, patient gender, date of birth, postal code, institution/hospital number, and admission/discharge information (e.g. admission category, length of stay), from fiscal year 1979/80 to 2001/2002.[34] Currently, the discharge data includes inpatient acute, chronic and rehabilitation care, as well as day surgery cases. Most recently, enhancements have been made for the collection of additional data on mental health inpatients and therapeutic abortions. CIHI receives data directly from participating hospitals or their provincial/territorial ministries of health (about 85% of all hospital inpatient discharges in Canada) representing approximately 4.3 million records annually.

CIHI adheres to strict policies to protect the privacy and confidentiality of health information it receives through the DAD.[35] Any identifying information, such as patient name, postal code or service providers, was removed before the case costing data and activity data were linked. The Case Mix department at CIHI conducted the record linkage using the DAD unique key comprising of the Fiscal Year, Province, Institution Number, Batch Period, Batch Number and Abstract Number. Once the record linkage

was complete, the DAD unique key was removed from the dataset. The key data elements for this study were: *age, Case Mix Groups or CMG, and total costs.*

2.1.3 The Classification And Regression Tree (CART) Algorithm

Many DRG-based patient classification systems have been developed using recursive partitioning.[9] Recursive partitioning has also been used successfully in other applications of health care to create clinical decision rules.[36-39] Recursive partitioning allows for the modelling of associations between variables like the patient's diagnosis and resource use. Therefore, the same methodology will be appropriate to determine whether any associations exist between a pediatric patient's age and resource use. One of the better-known recursive partitioning techniques is the classification and regression tree (CART) algorithm developed by Leo Breiman, Jerome H. Friedman, Richard A. Olshen and Charles J. Stone in the late 1970's.[40] Their work on CART was published in a monograph in 1984.

CART is a classification technique used to organize or partition data into two or more groups using binary recursive partitioning.[40-42] The process is binary because parent nodes are always split into exactly two child nodes, such that the dataset is partitioned into mutually exclusive subsets.[43] The process is also recursive because each child node can then be treated as a parent node, and thus be spilt into two child nodes.[43] The process continues until further splitting is impossible (e.g., each terminal node has only 1 case) or stopped by some pre-determined stopping rule (e.g., a terminal cell cannot have less than 200 cases). The right thresholds for stopping are not known in advance; therefore CART grows the largest tree possible and then prunes nodes off where necessary to create a tree of a given complexity that produces the largest

reduction in variance of the dependent variable.[43] Often used in data exploration, CART may reveal associations and structure in data, not previously evident but sensible and useful once found. The algorithm can produce two types of trees: i) classification trees for categorical response variables; or ii) regression trees for continuous response variables. For this study, the focus will be on regression trees since the dependent variable (i.e., total inpatient cost) is a continuous variable.

Regression trees using CART can be described in several key phases: 1) Dataset Partitioning 2) Tree Growing; 3) Stopping Tree Growing 4) Tree Pruning; and 5) Optimal Tree Selection.

2.1.3.1 Dataset Partitioning

Dataset partitioning is usually employed for large datasets (i.e., ≥ 1000 cases) and involves randomly dividing the data into several (equal or unequal) groups.[40] In this study, the data are partitioned 40/30/30 into a 'learn', 'validate' and 'test' datasets. The 'learn' dataset is used in the initial model building exercise to grow the largest tree possible, the 'validate' dataset is then used to prune the largest tree to find the best sub-tree, and the 'test' dataset is used as a holdout sample to assess the best-computed tree derived from the 'learn' and 'validate' datasets.

2.1.3.2 Tree Growing

The objective of the tree-growing phase is to grow the largest tree possible. At the beginning of the tree building process, all records contained in the 'learn' dataset are placed at the parent node. The parent node is then split into two child nodes, and each

child node is recursively partitioned until further splitting is impossible or stopped. A step-by-step of the tree growing process is explained below:

1. Divide the dataset into a 'learn' sample (to grow the tree), a 'validate' sample (to prune the tree and select the best sub-tree), and a 'test' sample (to assess the best sub-tree);
2. Using the 'learn' sample, search all possible splits for each predictor from the parent node. In general, for *n* cases with *k* continuous variables, there are *nk* possible splits.[40] The CART algorithm examines all possible splits for every value of a predictor variable being considered for the model. As a result, the splitting threshold for each predictor variable is increased by each value to find the "best split".
3. For a continuous response variable, splitting is based on the reduction in variance;
4. The "best split" is chosen based on the minimum sum of squares, $\sum_{i=1}^{N}(y_i - d(x_i))^2$, among all possible predictors[40]
5. The goodness of the split is assessed based on: A) Accuracy; and B) Node Impurity;
   A. Accuracy = mean squared error of the predictor is computed using the same data *N* (i.e., the learn sample) as used in constructing the predictor *d*;[40]

$$R(d) = \frac{1}{N}\sum_{i=1}^{N}(y_i - d(x_i))^2$$

   B. Node Impurity is the least squared deviation (for a continuous response variable) to decrease the impurity from parent node to child node;[40]

$$R(t) = \frac{1}{N_w(t)} \sum_{i=t} w_i f_i \left(y_i - \bar{y}(t)\right)^2$$

where:

$N_w(t)$ = weighted number of cases in node t
w_i = value of weighting variable for case i
f_i = value of frequency variable
y_i = value of the response variable
$\bar{y}(t)$ = mean for node t;

2.1.3.3 Stopping Tree Growing

The tree growing process is stopped when each child node is represented as a terminal node, otherwise the process continues until this is achieved. In other words, the largest tree is reached and further splitting is impossible when each terminal node contains only one case. The tree growing process can also be stopped if some predetermined splitting threshold is reached. Splitting thresholds may include criteria such as minimum node size, or maximum tree width or depth. With a potential split at each node in this study, splitting was stopped when a node contained less than 10% of the total sample, the terminal node reached the minimum node size of 1 record and/or the tree depth of six levels was reached.

2.1.3.4 Tree Pruning

Since the purpose of the CART algorithm is to find the best possible way to split these data, the number of terminal nodes is not usually predetermined. The objective of the tree-pruning phase is to prune nodes off of an overgrown tree to create a sequence of smaller and simpler trees, and thus find a sub-tree of appropriate complexity to allow adequate fit to the data without overfitting.[40] Pruning relies on a complexity parameter (α), which attempts to measure how much fitting to the data the tree allows, which in turn

is a function of the number of terminal nodes.[40] The actual value of α used is determined using the 'validate' dataset.[40] As α increases, more nodes are pruned away. Pruning is accomplished in several steps:

1. Using the 'validate' dataset, derive sub-trees from the largest tree grown;
2. Calculate the misclassification error rate for each sub-tree;[40]

$$R_\alpha(T) = R(T) + \alpha\left|\tilde{T}\right|$$

where:

$R(T)$ = misclassification cost on the 'validate' dataset for sub-tree T;

$\left|\tilde{T}\right|$ = complexity or size of tree (i.e., number of leaves);

α = complexity parameter (a real number ≥ 0), which is a measure of how much additional accuracy a split must add to the tree to warrant additional complexity; it controls the trade-off between the fit to the 'learn' dataset and tree complexity;

3. For each value of α, CART finds the sub-tree $T(\alpha) \leq T_{max}$ which minimizes $R_\alpha(T)$;[40]
4. The optimal tree is the sub-tree that minimizes $R(T)$; the optimal tree can also be pruned to fit the investigator's needs;[43,44]

2.1.3.5 Selection Of The Optimal Tree

For continuous response variables, the "best split" is based upon maximizing the homogeneity of cases in each child node using least-square deviation.[40,42-44] Least-square deviation is a measure of node impurity and equals zero when only one class is present at each node.[40] The maximal tree will always fit the 'learn' dataset with higher accuracy than any other tree. The performance of the maximal tree on the 'learn' dataset is called the resubstitution cost, and usually overestimates the performance on

an independent dataset from a similar population. Therefore, the 'test' dataset is used as a holdout sample to estimate the proportion of cases misclassified, or variance, for each tree derived from the 'learn' and 'validate' datasets.

A regression function is fitted for each cell of the partition associated with sub-tree *T* to minimize the estimation error. The optimal tree is then selected at the correct complexity parameter α to fit the information in the 'learn' dataset.[40,42-44] In other words, the optimal tree is selected (with respect to its performance on the 'test' dataset) at α with the lowest misclassification rate or variance.

2.1.4 CART Versus Other Partitioning Techniques

Although more traditional statistical techniques, such as multiple linear regression and discriminant function analysis, could be used to explore whether or not the 0-17 year CMG age group can be refined, these methods are not recursive and hierarchical in nature like CART. Therefore, these more traditional statistical techniques are not well suited for creating distinct groups within patient grouping methodologies like CART. Multiple linear regression does not easily allow for modelling of non-linear associations and is unable to create distinct patient groups.[45] On the other hand, discriminant function analysis can be used to classify observations into distinct groups based on certain predictor variables; however, this technique usually relies on prior knowledge of the groups.[46] As a result, discriminant function analysis identifies predictor variables that can be used simultaneously to assign observations to pre-defined, naturally occurring groups.[46]

Other classification techniques, such as cluster analysis, could also be used to explore whether or not the 0-17 year CMG age group can be refined, but these methods rely on the same assumptions as regression analysis. In hierarchical cluster analysis, observed data are organized into meaningful subgroups of cases, in the absence of any *a priori* hypotheses, such that the resultant classification has an increasing number of nested groups.[47-49] The number of groups or clusters is usually not predetermined as the purpose of this procedure is to find the best possible way to group these data. Classification is based upon minimizing within-group variation and maximizing between-group variation.[47-49] However, the data must be interval or have true dichotomies, and have linear associations.

With CART, there are several advantages over more traditional statistical techniques. CART is non-parametric in nature as no distributional assumptions are made; therefore, the algorithm can model non-linear associations, is not affected by outliers, collinearity or heteroscedasticity, and can handle missing data using surrogate variables.[50,51] Some critics point out that the CART methodology is not probabilistic in nature, and therefore provides no probability level or confidence interval associated with predictions derived from the data.[50,51] Although bootstrapping may provide a way of assessing the accuracy of the prediction, and even generate confidence intervals, the accuracy of the results is more robust when the model is applied to other data.

2.1.5 Identifying Possible Age Splits

All models explored in used age in years (the independent variable) to create new age splits based on Canadian acute care inpatient costs for the 0-17 year CMG age group from fiscal year 1997/98 to 2000/01 (the dependent variable). As described

earlier, the data used for the study was obtained at no cost through CIHI's Graduate Student Data Access Program.[30]

Before the analyses, the dataset received from CIHI was manipulated. Specific records not to be included in the analysis were removed, key data elements (i.e., age, CMG and total cost) were formatted, and descriptive statistics of these key data elements were produced.

To begin with, a single variable was created to capture the age of all patients in years. Age information submitted to the DAD is captured in several fields (i.e., age_units (a number) and age_code (to denote whether the number in age_units represents year, month, day or newborn)). These fields were then used to translate the age of each patient into years. For example, for patients less than two years of age, their age was actually recorded in months, and those less than one month old were recorded in days. As a result, the age in years for newborns (less than 29 days), patients 29 days to 11 months and patients 12-23 months were assigned as 0, 0 and 1, respectively.

The next step in the process was to exclude records not to be included in the analysis. All newborns (less than 29 days) were excluded because the CMG/Plx methodology already assigns these patients to their own age-specific CMG (i.e., CMG 625 to CMG 648). As such, patients 0 years of age now represents those 29 days to 11 months of age. Once newborns were removed, two datasets were created:

1) a 'Special' dataset to include as many valid costs records as possible; and
2) a 'Typical' dataset to include only those cost records that passed the same cost edits as applied by the Canadian Institute for Health Information.

2.1.5.1 Special Dataset

The 'Special' dataset contained all remaining records (those 29 days – 17 years) except those cases designated as long-stay outliers, transfers, diagnosis not usually hospitalized and ungroupable. Long-stay outliers (or cases that fall outside of the CMG/Plx-specific length of stay trim point and flagged as such by CIHI), were removed so as to not highly skew the cost data. Long-stay outliers are usually cases where the patient's course of treatment is most likely complete and he/she is waiting to be transferred to home or to a non-acute care facility. Any day(s) spent in hospitals beyond those required for treatment are considered to be alternate level of care (ALC) days. ALC days add costs above and beyond the normal course of treatment, but many facilities in Canada do not consistently identify ALC days in the patient's visit. As a result, cases were removed where the patient's length of stay (LOS) was greater than the trim point. For each analytic cell representing CMG, complexity level and age category (the Aplx cell), the trim point is the LOS cut-off value calculated using the TRIM formula below.[2]

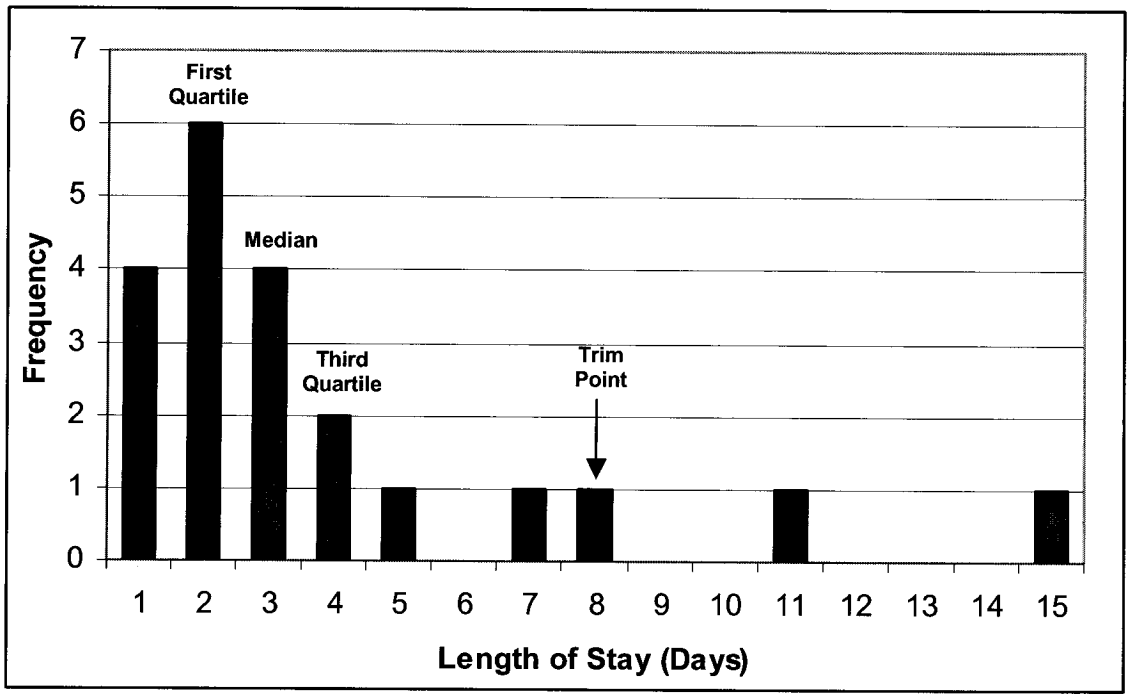
$$Trim_{ijk} = Q3_{ijk} + 2(Q3_{ijk}-Q1_{ijk})$$

where:

$Q1_{ijk}$ = first quartile of LOS for a given Aplx cell
$Q3_{ijk}$ = third quartile of LOS for a given Aplx cell
i = CMG
j = Complexity Level
k = Age Category;

For example, Figure 1 illustrates that cases with a LOS beyond 8 days would be excluded or "trimmed out" from the dataset for the CMG below.

Figure 1: Trim Point Example



Transfer cases were removed because it was difficult to determine whether or not all costs associated with the transfer (i.e., costs associated with the transfer-in facility and costs associated with the transfer-out facility) were captured for the entire hospital visit. Cases where the diagnosis was not usually hospitalized were removed, as these represented special cases that did not/could not follow current clinical practice within the acute care setting. Lastly, all ungroupable cases that could not be assigned to a CMG were removed as these cases contained some type of coding error (i.e., invalid MRDx).

2.1.5.2 Typical Dataset

The 'Typical' dataset is different than the 'Special' dataset. The 'Typical' dataset contained all records (those 29 days to 17 years) after applying CIHI's cost record edits used in the calculation of RIW, and removing all atypical records (defined by CIHI). Costs edits included removing records that had no associated cost, costs of less than $200 with a length of stay greater than 1 day, and a cost per diem of less than $50. Atypical records included those flagged as long-stay outliers, transfers, sign-outs, deaths, diagnosis not usually hospitalized and ungroupable. Many of these patient's either did not complete the normal course of treatment or incurred costs not usually associated with the normal course of treatment

These two datasets were created to determine if there was a difference in the resulting age splits using cases classified as 'typical', according to CIHI definitions, compared to using all cases with reliable cost information. To ensure that each CMG had sufficient volume for the CART algorithm, only CMG with 100 or more records were kept in the 'Special' dataset. For comparison purposes, CMG kept in the 'Special' dataset were also kept in the 'Typical' dataset. A total of 123 out of 472 CMG were included in the both the 'Special" and 'Typical' datasets. Table 3 highlights the number of records removed from the original dataset to create the 'Special' and 'Typical' datasets. Both datasets retained approximately 87% of the available cost records. Appendix D shows the number of costs records for each CMG included in the 'Special' and 'Typical' datasets.

Table 3: Summary of Cost Records Included the 'Special' and 'Typical' Datasets

Description	Special Dataset	Typical Dataset
All ICD-9-CM/ICD-9/CCP-based cost records (1997-2000) received from CIHI (Sept. 2003)	**196,978**	**196,978**
Remove Newborns	**117,387**	**117,387**
Apply Cost Record Edits (remove if)		
Total Cost = $0	---	5
Total Cost < $200 & TLOS >1 day	---	5,288
Cost per diem < $50	---	2,025
Clean Cost Records	**79,591**	**72,273**
Atypical Records Removed		
Long-Stay Outliers	2,252	1,904
Transfers	8,592	7,744
Sign-outs	---	259
Deaths	---	467
Diagnosis Not Usually Hospitalized	105	91
Ungroupable	35	28
Total Cost Records Available	**68,607**	**61,780**
Special Dataset: Remove CMG < 100 cost records	**59,580**	**---**
Typical Dataset: Include Same CMG as Special Dataset	**---**	**54,022**

Using these two datasets, descriptive statistics on total cost and age were computed using the SAS System for Windows, Release 8.02. Results for the total cost and age variables are found in Chapter 3.

2.1.6 Model Building

Using total costs as the target variable, four different tree models were built using age (in years) and/or CMG as the input variables. Each model is described in detail below.

2.1.6.1 Model 1: Global Age Splits

The first model used the CART algorithm to create global age splits based on total costs regardless of CMG, and examined the overall mean squared error for the resulting splits. This model would readily highlight those age groups that were homogenous with respect to total costs associated with the hospital stay, regardless of the reason for admission. The CART algorithm was used to generate global age splits two different ways; i) using the 'Special' or 'Typical' dataset with no partitioning; and ii) using the 'Special' or 'Typical' dataset partitioned 40/30/30 into a 'learn', 'validate' and 'test' datasets. This comparison was employed to test whether or not partitioning the dataset made a difference with respect to the age splits produced. See the Modelling Using CART section below for details.

2.1.6.2 Model 2: CMG-Based Splits

The second model created CMG splits and examined the overall mean squared error for these resulting splits. In this model, the CMG represented the clinical basis against which further age splits could be explored and compared. In other words, cases within a CMG have been diagnosed with the same disease/condition or received similar treatment for their condition, and thus, the CMG represent the clinical unit of analysis. Any further age splits would be disease or treatment specific. Therefore, it makes more

sense to use CMG as the starting point for the analysis. Using the entire 'Special' or 'Typical' dataset, all 123 CMG were included in the regression model below. See the Regression Models section below for details.

$$total = \sum_{i=1}^{123} \beta_i cmg_i$$

where:

total = total cost of the acute inpatient visit;
cmg = clinical unit of analysis, representing a specific disease or treatment (123 unique values);
i = CMG 1 to CMG 123;
β = the regression coefficient for a given CMG;

2.1.6.3 Model 3: CMG-Specific Age Splits

Starting at the CMG-level, the third model used the CART algorithm to create CMG-specific age splits based on total costs, and examined the overall mean squared error for these resulting splits. Any further age splits would be disease or treatment specific, and thus, more representative of how health care resources are consumed in the pediatric population. Both the 'Special' and 'Typical' datasets were partitioned before the CART algorithm was applied. See the Modelling Using CART section below for details.

2.1.6.4 Model 4: Fixed Age Splits Applied Across CMG

The fourth model created fixed age splits applied across all CMG, and examined the overall mean squared error for the resulting splits. This model was explored to determine whether or not there were two age splits between 0 and 17 years that could be applied universally across all CMG, since applying CMG-specific age splits may be impractical and unnecessary. Two age splits would create a maximum of three age groups, and thus may be more feasible to implement without creating volume issues.

Using the entire 'Special' or 'Typical' dataset, each of the 136 two age split combinations between 0 and 17 years were included in the regression model below. See the Regression Models section below for details.

$$total_{i=1}^{136} = \beta_i \, fixed\_agesplit_i$$

where:

total = total cost of the acute inpatient visit;
fixed_agesplit = dummy variable representing one of the 136 two age split combinations between 0 and 17 years;
β = the regression coefficient for a given two age split combination;

2.1.7 Modelling Using CART

For the Global Age Splits (Model 1) and CMG-Specific Age Splits (Model 3), the 'Special' and 'Typical' datasets were partitioned 40/30/30 into a 'learn', 'validate' and 'test' dataset. Each 'learn' dataset was used for the tree-building process. Using the options specific to the CART algorithm, the Tree Node in SAS Enterprise Miner for Windows NT, Release 4.1 was used to carry out the tree-building process.[52] The tree-building process was limited to two branches per node and a maximum depth of six levels using variance reduction as the splitting criterion.

The Tree Node was setup to search for split rules that maximized the reduction in variance.[52] The maximal tree was grown using the 'learn' dataset, and then pruned to select for minimum tree complexity using the 'validate' dataset. Once the optimal tree was selected based on reduction in variance, it was then assessed using the 'test' dataset.[52] Since the optimal tree had a large number of nested splits, the number of splits was also assessed based upon knowledge of the dataset and practicality of implementation (i.e., 2-3 age groups may be more acceptable and relevant to the pediatric community than 18 different groups).

2.1.8 Regression Models

To assess the CMG-Based Splits (Model 2) and Fixed Age Splits (Model 4), the PROC REG procedure in the SAS System for Windows, Release 8.02 was used to calculate the mean squared error for all 123 CMG and 136 fixed age splits, respectively. Using PROC REG, a simple linear regression of total cost (the dependent variable) was performed on a CMG or fixed age split (the independent variable) using the following SAS statements:

Model 2 – CMG-Based Splits:

```
proc reg;
        model total=cmg/selection=mse;
run;
```

where:

total = total cost of the acute inpatient visit;
cmg = clinical unit of analysis, representing a specific disease or treatment (123 unique values);
mse = mean squared error;

Model 4 – Fixed Age Splits:

```
proc reg;
        model total=fixed_agesplit/selection=mse;
        by cmg;
run;
```

where:

total = total cost of the acute inpatient visit;
fixed_agesplit = dummy variable representing one of the 136 two age split combinations between 0 and 17 years;
mse = mean squared error;

Options used in PROC REG are specified after the slash (/) at the end of the model statement. Since the study is interested in the mean squared error (MSE) of the

specified model, the 'mse' option was used in the regression model statement to output the MSE value for the regression model.

2.1.9 Calculating The Overall Mean Squared Error For Each Model

The mean squared error (MSE) is the sums of squares (SSE) divided by the respective degrees of freedom.[53,54]

$$MSE = \frac{SSE}{n}$$

where:

SSE = sum of squared errors
n = sample size

The Tree Node in SAS Enterprise Miner calculated a separate MSE for the 'learn', 'validate' and 'test' datasets. An estimate of the overall or weighted MSE for age splits modelled using CART was calculated as follows:

Model 1 - Global Age Splits:

$$MSE_{Overall} = \frac{(MSE_{Learn} * n_{Learn}) + (MSE_{Validate} * n_{Validate}) + (MSE_{Test} * n_{Test})}{n_{Total}}$$

Model 3 – CMG-Specific Age Splits:

$$MSE_{Overall} = \frac{\sum_{i=1}^{123}\left[(MSE_{i,Learn} * n_{i,Learn}) + (MSE_{i,Validate} * n_{i,Validate}) + (MSE_{i,Test} * n_{i,Test})\right]}{n_{Total}}$$

where:

i = CMG.

Using PROC REG in SAS, however, the MSE is calculated as follows:

$$MSE_{SAS} = \frac{SSE}{n-p} = \frac{\Sigma(y-\hat{y})^2}{n-p}$$

where:

SSE = sum of squared errors
y = observed value
$\hat{y}$ = the regression line
n = number of observations in the sample or sample size
p = number of parameters in the regression model

Using p in the denominator corrects for the fact that the sample SSE usually decreases as new predictors are added to a regression model, even if those predictors may not be useful in the population. Therefore, the correction allows MSE values of multiple linear regression models to be comparable to MSE values of simple linear regression models. All MSE values obtained from the PROC REG procedure in SAS were adjusted by removing the correction as follows:

Model 2 – CMG-Based Splits:

$$MSE_{Overall} = \frac{MSE_{SAS} * (n-p)}{n}$$

Model 4 – Fixed Age Splits:

Because the MSE was determined for each CMG, the overall MSE value for the regression model was calculated as follows:

$$MSE_{Overall} = \frac{\sum_{i=1}^{123}\left(MSE_{i,SAS} * (n_i - p_i)\right)}{n_{Total}}$$

where:

i = CMG.

The overall MSE calculated for each model using the 'Special' or 'Typical' datasets were then compared to see which model produced the smallest MSE, and thus, provided the most ideal age splits.

In sum, the CART algorithm is most ideal for exploring the creation of new age splits within the existing 0-17 year CMG age category. The next chapter will highlight the results from the age splitting exercise using the CART algorithm and other regression models.

Chapter 3: Results

3.1 Introduction

Chapter 2 provided an overview of the data sources used for the study, including a description of the Discharge Abstract Database (DAD) and Canadian inpatient case cost data, as well as a description of the classification and regression tree (CART) methodology developed by Breiman et al. The chapter also described in detail, four models developed and employed to answer the thesis question, "Can the Canadian case mix age group (0-17 yrs) be further refined to better represent the pediatric population?" These four models were not only created to test whether or not the 0-17 year CMG age group could be further refined, but also to determine what method was most practical and/or efficient for achieving this objective.

In this chapter, the results of the study will be presented in detail. Specifically, this chapter will present:

i.) Descriptive statistics for the Total Cost and Age variables for the 'Special' and 'Typical' datasets used in the study; and

ii.) Results for the four models:

1) Model 1: Global Age Splits;
2) Model 2: CMG-Based Age Splits;
3) Model 3: CMG-Specific Age Splits; and
4) Model 4: Fixed Age Splits Applied Across All CMG.

3.2 Descriptive Statistics For Total Cost And Age

As outlined in the Methodology section, two datasets were created; one to include as many valid costs records as possible (i.e., the 'Special' dataset), and the other to include cost records that passed the cost edits as applied by the Canadian Institute for Health Information (i.e., the 'Typical' dataset). Descriptive statistics for the Total Cost and Age variables in each dataset are highlighted in the tables and figures below.

Table 4: Descriptive Statistics for Total Cost, 'Special' and 'Typical' Datasets

Dataset	Minimum	25th Percentile	Mean	75th Percentile	Maximum
Special	$0.00	$736.32	$2,795.91	$2,879.41	$378,339.13
Typical	$50.04	$933.63	$2,986.89	$3,070.45	$206,466.24

Figure 2: Frequency Distribution of Total Cost for Pediatric Patients Aged 0-17 years, 'Special' and 'Typical' Datasets

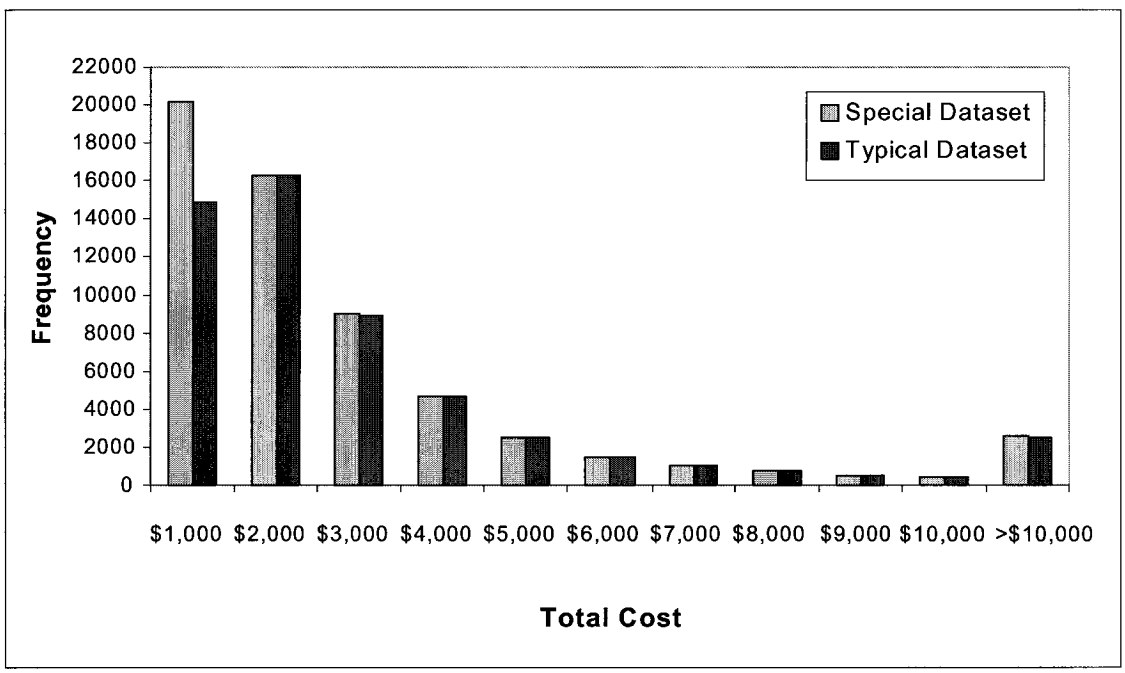
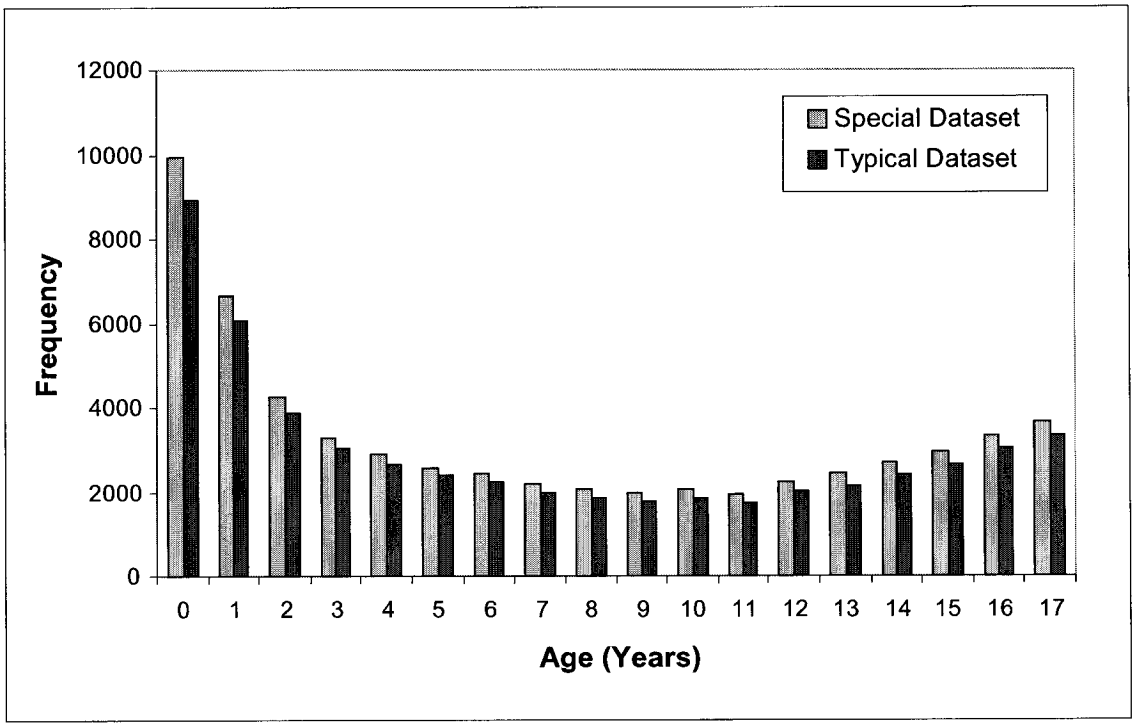


Table 5: Descriptive Statistics for Age (years), 'Special' and 'Typical' Datasets

Dataset	Minimum	25th Percentile	Mean	75th Percentile	Maximum
Special	0	1	6.90	13	17
Typical	0	1	6.89	13	17

Figure 3: Age Distribution for Pediatric Patients Aged 0-17 years, 'Special' and 'Typical' Datasets



3.3 Model Results

For each of the four models developed, age split results are presented below. In general, the lower the mean squared error (MSE), the better the model with respect to minimizing variation in total costs between age groups and/or CMG.

3.3.1 Model 1: Global Age Splits

For Model 1, the results for all records (in the 'Special' and 'Typical' datasets) split by age using CART are shown in Table 6 and Table 7. These tables illustrate that the difference in the MSE between the partitioned and non-partitioned dataset was minimal. However, the number of leaves (or terminal groups) was always greater in the non-partitioned datasets (perhaps due to overfitting of these data). The MSE for the 'Typical' dataset is 16% lower than for the 'Special' dataset given the removal of outliers. For both datasets, the MSE is slightly higher with partitioning than without. Tree diagrams are also shown in Figure 4 to Figure 7. In these figures, the shading represents the cost homogeneity of the cases. In general, the darker the node, the more homogenous are the total costs of the cases in that node.

Table 6: Optimal Number of Age Splits and Mean Squared Error (MSE) for Global Age Splits, 'Special' Dataset (CMG with Count ≥ 100 Records)

Tree	Number of Leaves	MSE Overall
All Records: No Partition	11	34847570.96
All Records: Partitioned 40/30/30	6	34852804.00

Table 7: Optimal Number of Age Splits and Mean Squared Error (MSE) for Global Age Splits, 'Typical' Dataset (Same CMG as 'Special' Dataset)

Tree	Number of Leaves	MSE Overall
All Records: No Partition	10	29118396.69
All Records: Partitioned 40/30/30	7	29127870.56

Figure 4: Tree Diagram for Global Age Splits, 'Special' Dataset - No Partition

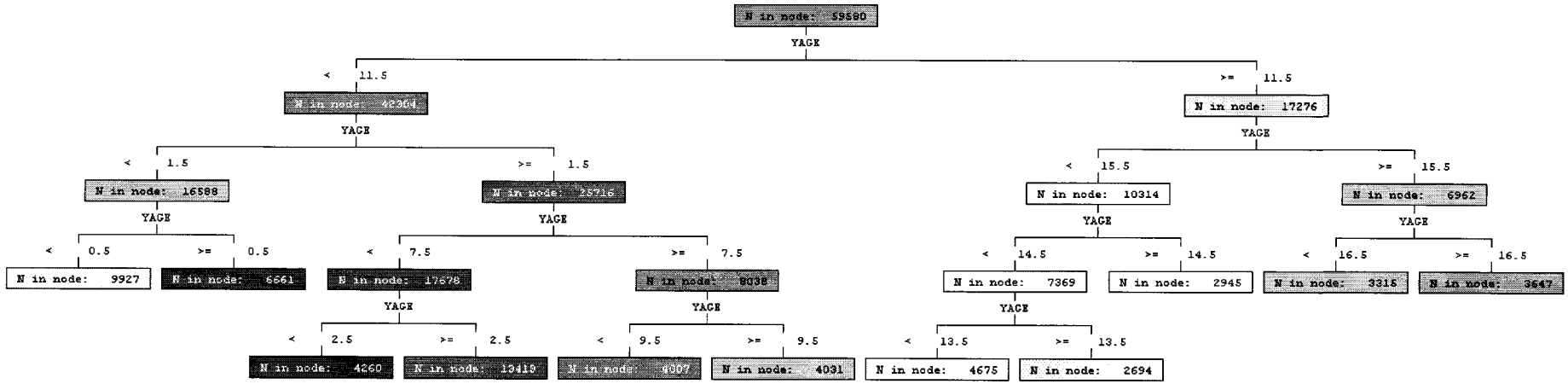


Figure 5: Tree Diagram for Global Age Splits, 'Special' Dataset - Partition 40/30/30 (Validate dataset shown)

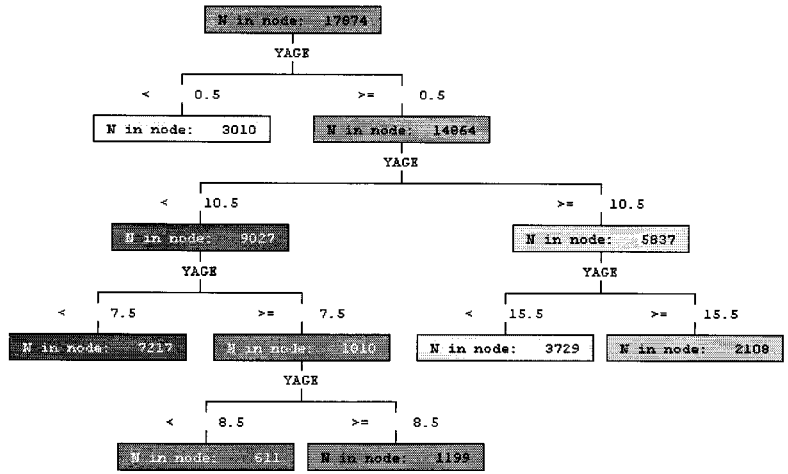


Figure 6: Tree Diagram for Global Age Splits, 'Typical' Dataset - No Partition

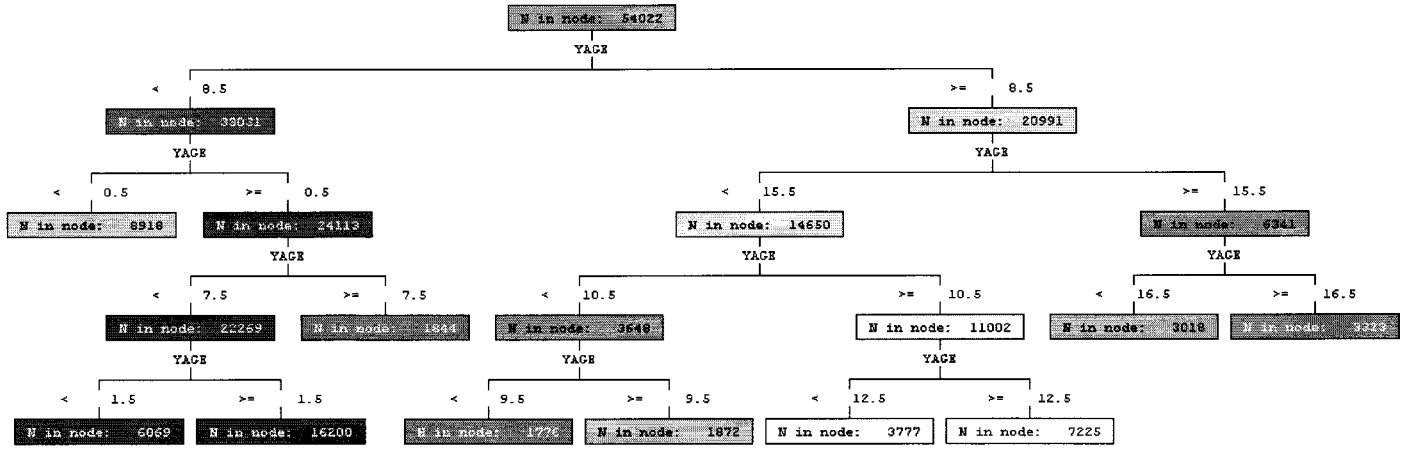
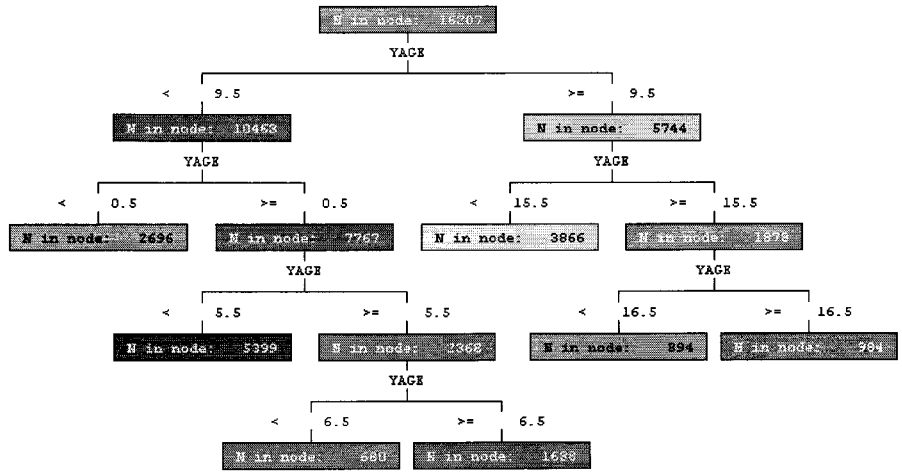


Figure 7: Tree Diagram for Global Age Splits, 'Typical' Dataset - Partition 40/30/30 (Validate dataset shown)



3.3.2 Model 2: CMG-Based Splits

Model 2 contains 123 leaves (i.e., one for each CMG included in the model), and represents the clinical baseline for which further age splits should be compared against. For Model 2, the overall mean squared error for all records (in the 'Special' and 'Typical' datasets) split by CMG only are shown in Table 8. For both datasets, the MSE for Model 2 is higher than for Model 1.

Table 8: Mean Squared Error (MSE) for CMG-Based Splits, 'Special' and 'Typical' Datasets

Dataset	MSE Overall
Special	34916717.00
Typical	29192672.28

3.3.3 Model 3: CMG-Specific Age Splits

For Model 3, the results for the CMG-specific age splits (in the 'Special' and 'Typical' datasets) are shown in Appendix E. The overall mean squared errors for this model are shown below in Table 9. The MSE for Model 3 for both datasets are lower than for Model 1.

Table 9: Mean Squared Error (MSE) for CMG-Specific Age Splits, 'Special' and 'Typical' Datasets

Dataset	MSE Overall
Special	30071623.76
Typical	23254427.99

Of the 123 CMG included in the 'Special' and 'Typical' Datasets for analysis, 64 (or 52%) and 49 (or 40%) of the CMG, respectively had no further age splits indicating that patients within these CMG were not only clinically similar but homogenous with respect to total cost. For the remaining CMG split by age, the number of leaves ranged from two to as many as nine within each dataset (see Figure 8). The frequencies of the first and second-level CMG-specific age splits are illustrated in Figure 9 and Figure 10 below. Both figures illustrate that for some CMG there are differences in the consumption of hospital resources by age. For example, such differences appear in CMG 485 (Nutrition & Miscellaneous Metabolic Disorders) where CART suggests age splits for those <2.5 years, ≥2.5 and <14.5 years, and ≥ 14.5 years of age. This division indicates there are sub-populations within the 0-17 years CMG age group that should be differentiated on the basis of cost or consumption of hospitals resources.

Figure 8: Distribution of the Number of Leaves, CMG-Specific Age Split Model

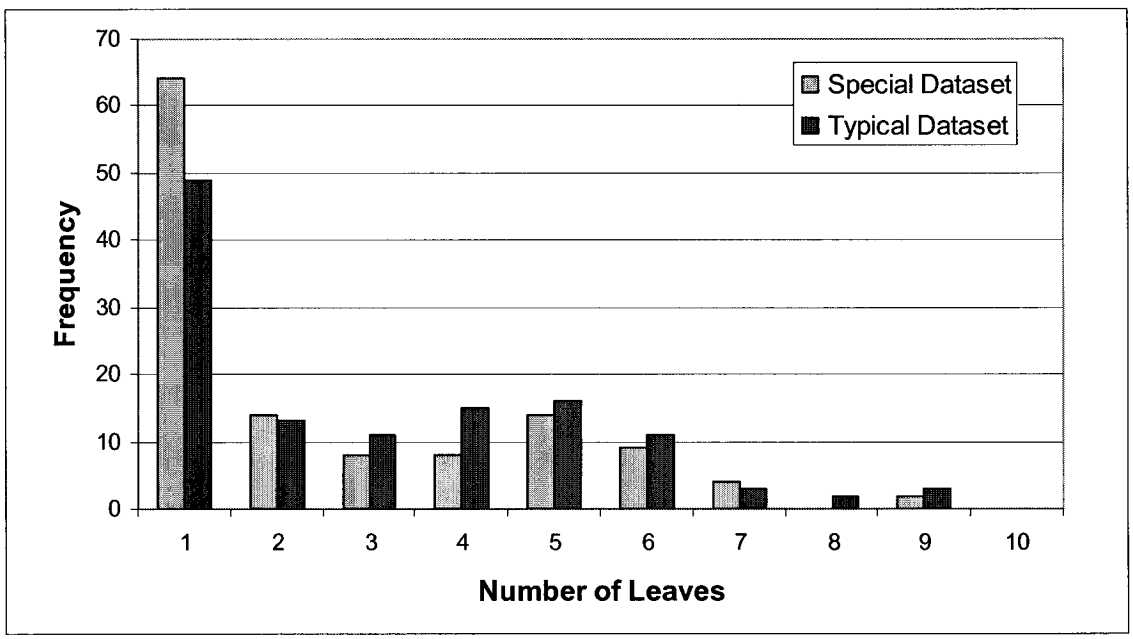


Figure 9: Frequency Distribution of the First and Second Level CMG-Specific Age Splits, 'Special' Dataset

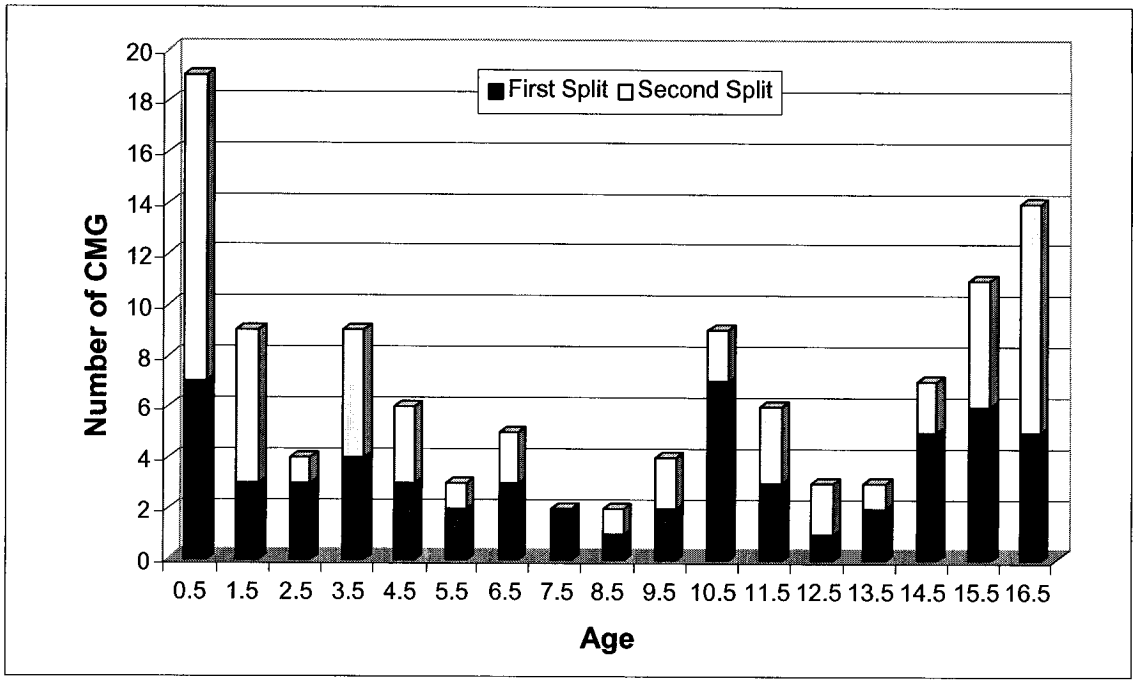
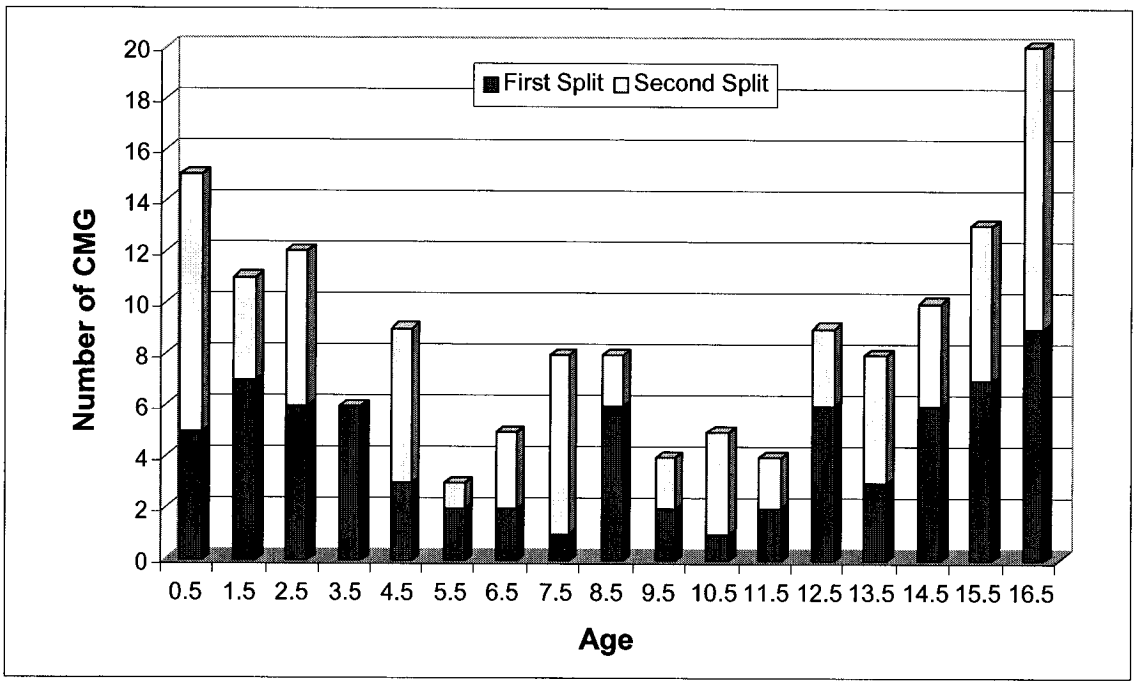


Figure 10: Frequency Distribution of the First and Second Level CMG-Specific Age Splits, 'Typical' Dataset



3.3.4 Model 4: Fixed Age Splits Applied Across CMG

For Model 4, the MSE results for fixed age splits applied across all CMG (for the 'Special' and 'Typical' datasets) are shown in Appendix F. The top ten fixed age splits are shown below in Table 10, and illustrate that age splits should be made among very young children (i.e., those less than 6.5 years of age) to differentiate to degrees of resources consumption. All CMG were split into three age groups, and the smallest MSE resulted when all a CMG were split at 0.5 years and 1.5 years, regardless of which dataset was used. Model 4 leads to a lower MSE than all other models.

Table 10: Top Ten Fixed Age Splits Applied Across CMG, 'Special' and 'Typical' Datasets

Age Split		MSE	
First Split	Second Split	Special Dataset	Typical Dataset
0.5	1.5	27844677.78	21295343.50
0.5	2.5	28498590.28	21862973.76
1.5	2.5	28544848.67	21886330.40
0.5	3.5	28741180.93	22128714.52
1.5	3.5	28797920.37	22163246.55
2.5	3.5	28816047.69	22159152.36
0.5	4.5	29358698.29	22693083.02
0.5	5.5	29368041.59	22696935.25
0.5	6.5	29389539.61	22718198.14
1.5	4.5	29419836.12	22728825.70

3.4 Model Summary

The table below highlights the overall mean squared error calculated for each model using the 'Special' and 'Typical' datasets.

Table 11: Summary of the Mean Squared Error (MSE) by Model

Model		Mean Squared Error	
		Special Dataset	Typical Dataset
1A	Global Age Splits - No Partition	34847570.96	29118396.69
1B	Global Age Splits - Partitioned 40/30/30	34852804.00	29127870.56
2	CMG-Based Splits	34916717.00	29192672.28
3	CMG-Specific Age Splits Using CART - Partitioned 40/30/30	30071623.76	23254427.99
4A	Fixed Age Splits Applied Across All CMG - Smallest MSE Split at 0.5 & 1.5yrs	27844677.78	21295343.50
4B	Validation of 0.5 & 1.5yrs Fixed Age Splits - Partitioned 70/30	27259873.68	20626000.48

The CART algorithm created additional age splits within a CMG only when the total costs associated with the stay in hospital were not homogenous for across all ages with the 0-17 year CMG age group. However, splitting the 0-17 year CMG age group using fixed age splits across all CMG proved to be most practical (i.e., produced the smallest mean squared error) when cases in each CMG were grouped into those less than 0.5 years, greater than 0.5 years and less than 1.5, and greater than 1.5 years. The next chapter will discuss these results in more detail, including the application of the results, limitations of the study, and directions for further research.

Chapter 4: Discussion and Conclusion

4.1 Importance Of The Study

The current ICD-9-based Case Mix Groups or CMG methodology has three broad age groups: 0-17 years, 18-69 years and 70+ years. These age groups were initially created to adjust acute care health resource indicators, such as Expected Length of Stay (ELOS) and Resource Intensity Weights (RIW), by age. However these age groups were not derived from inpatient activity data, but rather were age conventions simply adopted from other international groupers, primarily the Diagnosis Related Groups or DRG system. As a result, an assumption was made about how health care resources were being utilized by patients within a given age group; that is, patients within the same age group utilized similar resources for a given condition regardless of age.

Canadian health care providers who work in the pediatric community in Canada believe that patients in the 0-17 year CMG age group are not homogenous with respect to healthcare management and resource utilization. Hence, the purpose of this thesis was not to calculate or determine the average cost of hospital services for pediatric patients by CMG and age. Rather, this thesis explored whether refining the current 0-17 year CMG age group is practical, and if so to create new age groups that are meaningful to the pediatric community in Canada. More important, since the current CMG methodology is being redeveloped to take advantage of the ICD-10-CA and CCI classifications systems, this thesis will provide useful information on pediatric age splits and how new age splits can be developed.

Since the new age splits were not determined *a priori* and total cost of the case was used to differentiate across age, the classification and regression tree (CART) algorithm was a more straightforward way than other statistical techniques for exploring how the 0-17 year CMG age group could be refined. Below is a discussion of the results from the four models explored:

- Model 1: Global Age Splits;
- Model 2: CMG-Based Splits;
- Model 3: CMG-Specific Age Splits; and
- Model 4: Fixed Age Splits Across All CMG.

4.2 Interpretation of the Model Results

The results of this thesis highlight that the 0-17 year age group in the current CMG methodology is too broad a category and not representative of the resources consumed by pediatric patients in Canada because of heterogeneity.

Of the various models explored, the Global Age Splits model (Model 1) illustrates that, regardless of the diagnosis and treatment a patient may receive, there are indeed differences in cost among patients 0-17 years of age. Model 1 identified several high-level age splits for the various datasets used, including those:

- < 11.5 and ≥ 11.5 years of age (Figure 4: Special Dataset - No Partition);
- < 8.5 and ≥ 8.5 years of age (Figure 6: Typical Dataset - No Partition); and
- < 9.5 and ≥ 9.5 years of age (Figure 7: Typical Dataset - Partition).

Within these age splits, second-level splits often identified were:

- < 0.5 and ≥ 0.5 years of age or < 1.5 and ≥ 1.5 years of age for the lower age split; and

- < 15.5 and ≥ 15.5 years of age for the higher age split.

The results from Model 1 illustrate that resource consumption of hospital and health care services among those 0-17 years of age is similar to a U-shaped distribution. In other words, the patient population at either end of the distribution is unique and different with respect to resource consumption compared to the patient population in the middle of the distribution.

Since the CMG methodology is built on clinically defined groups, splitting age without taking CMG into consideration does not seem practical or feasible from an implementation perspective. For example, Model 1 showed that, in general, those <2 years of age were different in resource consumption than those 2-15 years and those 16-17 years of age. However, this model does not indicate whether or not the distribution was applicable for children diagnosed with asthma, gastrointestinal diseases or common childhood diseases. Global age splits cannot distinguish between age effects and CMG effects. Thus, the 0-17 year CMG age group should be refined within each CMG. The CMG-Based Splits model (Model 2) represents the clinical baseline or starting point from which the 0-17 year CMG age group could be refined. Splitting age within a CMG will only serve to improve the explanatory power on the observed variations in total costs (i.e., reduce the mean squared error or increase the R^2).

The CMG-Specific Age Splits model (Model 3) explored the refinement of the 0-17 year CMG age group within each CMG using CART. Of the 123 CMG included in the study, no further age splits were recommended for 52% of the CMG in the Special dataset and 40% of the CMG in the Typical dataset. In others words, the 0-17 year CMG age group was appropriate for describing the level of resources consumed for all

patients within these CMG. Within these CMG, all cases were relatively clinically and resource homogenous.

For the remaining CMG, CART illustrated that pediatric patients within these CMG were not entirely resource homogenous. The 0-17 year CMG age group could be split into several age groups that represent a more accurate picture of how resources are being consumed. For example, CART suggested that cases in CMG 145 (Tracheobronchitis) could be split into those <2.5 and ≥2.5 years of age; cases in CMG 146 (Asthma) could be split into those <7.5, ≥7.5 and <10.5, ≥10.5 and <15.5, and ≥15.5 years of age; and cases in CMG 294 (Esophagitis, Gastritis and Miscellaneous Digestive Disorders) could be split into those <0.5, ≥0.5 and <1.5, ≥1.5 and <13.5, and ≥13.5 years of age. For certain CMG, CART was able to identify an interaction between CMG and the age of the patient that better differentiates the cost of care and/or treatment.

After generating CMG-Specific Age Splits, it was recognized that implementing the results from Model 3 could be challenging and likely not feasible. Some CMG only warranted 2 splits on age and others had as many as 5 splits on age. The latter number of age splits divides the pediatric population too finely. When calculating health resource indicators, like ELOS and RIW, having more than 2 age splits within a CMG may create a low volume issue for some CMG and an even greater issue for other CMG already experiencing low volume with the current 0-17 year age group.

Although more splits on age may further explain the variation in total costs for some CMG, there was no consistency among the number of age groups or in the age range for the resulting age groups. From the examples above, it is evident that the new age groups could be either small and represent as little as a one-year cohort of pediatric

patients, or large and represent a ten-year cohort of pediatric patients. Given the various numbers of age groups and diverse age ranges, the age splits were not similar and did not follow any type of pattern across CMG. The CMG-Specific Age Splits were perhaps resource homogenous and even clinically meaningful, but were not standardized across CMG.

The Fixed Age Splits Across All CMG model (Model 4) attempted to explore whether or not two standard or fixed age splits could be applied across all CMG. Upon examining all 136 combinations and permutations of two fixed age splits between 0 and 17 years of age, it was apparent that the 0-17 year CMG age group could be split into three age groups across all CMG. The fixed age splits that produced the smallest mean squared error (MSE) were 0.5 years and 1.5 years. As a result, these age splits produced three new age groups: <1 year, ≥1 and <2 years, and ≥2 and ≤17 years. The next four best fixed age splits (listed in ascending order of the MSE) were: i) 0.5 and 2.5 years; ii) 1.5 and 2.5 years; iii) 0.5 and 3.5 years; and iv) 2.5 and 3.5 years. These splits suggest that the very young are different with respect to the consumption of hospital and health care services than older children. Thus, there may be some benefit to placing these patients into their own age group for the purpose of describing resources consumption.

4.3 Which Model Is Better For CMG Age Refinement?

After exploring the models described above, choosing the best model to refine the 0-17 year CMG age group depends on the model selection criteria being applied. A model could be chosen based on clinical relevance, heterogeneity and feasibility. If a model were chosen based on clinical relevance, then Model 3 (CMG-Specific Age Splits)

would be chosen to refine the 0-17 year CMG age group. In Model 3, the resulting age groups were created in context with the most responsible diagnosis (MRDx) and/or most resource intensive procedure coded on the patient's medical chart. The resulting age groups are only applicable to the CMG they were created for, as they describe the resource consumption patterns among patients with a specific disease and/or who receive similar treatment.

Using the other two criteria, the minimum MSE and the ability to create age groups that can be applied to all CMG, would result in Model 4 (Fixed Age Splits Across CMG) being chosen to refine the 0-17 year CMG age group. As shown in Table 11, Model 4 produced the smallest MSE using the same age splits applied across all CMG. The model with a small MSE relative to other models indicates that the model is best able to minimize between group variations, and thus produce relatively homogenous groups. As such, Model 4 was the best model to refine the 0-17 year CMG age group. Model 4 performed the best with respect to minimizing variations in the total cost for pediatric inpatients when splits at 0.5 years and 1.5 years of age. Hence, these age groups provide some insight into the resource consumption patterns among pediatric patients. Specifically, patients <1 year of age have different resource requirements than those ≥1 and <2 years of age. Likewise, both of these age groups were also different with respect to how resources are consumed compared to those ≥2 and ≤17 years of age.

4.4 Results Compared To Age Groups Found In The Literature

The results from this thesis are not only important to the development and evolution of inpatient grouping methodologies, but represent unique findings from an

original piece of research. This thesis is the first known Canadian study to examine the refinement of age groups used in inpatient grouping methodologies. No other known inpatient grouper or study has used or considered using the resulting age groups from the different models explored in this thesis. As discussed in Chapter 1, the DRG system was developed in the United States to group patients into clinically meaningful groups that were homogenous with respect to resource consumption. Other countries adopted the DRG system (including age groupings) despite differences in their demographics, data collection, information standards, and financing methods, compared to the U.S. All versions of the DRG system in use today use the 0-17 year age group to represent the pediatric population other than neonates (those less than 29 days old at admission). As a result, any differences in how resources are consumed among those 0-17 years of age are not being differentiated or reflected in various health resource indicators.

However, the results from this thesis share some similarity with recommendations provided to Canadian Institute for Health Information (CIHI) by the Canadian Pediatric Decision Support Network (CPDSN). The CPDSN suggested that the 0-17 year CMG age group should be sub-divided into seven age groups; those 0-28 days (for neonates – already provided for in the CMG methodology), 29 days-11 months, 12 months-2 years, 3-4 years, 5-9 years, 10-13 years and 14+ years.[5] The results from Model 4 suggest age splits at 0.5 years and 1.5 years, which create age groups similar to CPDSN's age groups of 29 days-11 months and 12 months-2 years. Although CPDSN's age groups were created based on expert committee input, two of these age groups are now supported by data-driven results produced by Model 4. These two age groups were not only recommended and supported by experts in the field, but were uncovered using CART. When applied in combination with the 2-17 year age group

across all CMG, these age groups produced the smallest MSE among all models explored in this thesis and differentiated costs well among pediatric patients.

4.5 Limitations Of The Study

This thesis is a first in Canada as no other study has explored the refinement of age categories, in particular pediatric age, used in the CMG methodology. As stated throughout the thesis, the purpose of the study was to explore whether refining the 0-17 year CMG age group was possible using a recursive partitioning technique such as CART. Although the analyses show that refining the 0-17 year age group is indeed possible and age could be split either at the CMG-level (i.e., age splits specific to the disease and/or treatment – Model 3) or universally across all CMG (i.e., fixed aged splits – Model 4), the study has several limitations.

First, this is a methodological study, and CMG-level analyses are limited. The study was only interested in splitting the 0-17 year population into distinct age bands (where possible) that would better explain or account for the observed variations in patient costs at the CMG-level. Using CART or applying fixed age splits was only a means of splitting age. Any resulting age groups were not assessed for clinical relevance, as this was not within the scope of the thesis. Hence, CMG-specific age groups and age groups resulting from the fixed age splits should be validated with respect to the role of age within a given disease and/or treatment. The clinical relevance of the resulting age groups (from the CMG-Specific Age Splits model and the Fixed Age Splits model) for the 123 CMG included in the analyses could be assessed through literature reviews and through seeking expert opinion from specialists in each area of medicine and/or established physician panels.

Secondly, no confidence intervals were calculated around the MSE. As described in Chapter 2, the 'Special' and 'Typical' datasets had 59,580 and 54,022 records, respectively. As a result, the study had access to cost data for the inpatient pediatric population in Canada over four fiscal years (1997/98 to 2000/01). Given the size of these datasets, these model results are most likely generalizable to the 0-17 year pediatric population in Canada.

Confidence intervals around the MSE could be generated using bootstrapping techniques.[55,56] However, bootstrapping using datasets with approximately 50,000 records, and resampling a minimum of 25 to 200 times would be very time consuming and computationally intensive given the marginal benefit that may be gained. In addition, the size of the dataset would likely produce confidence intervals that were very narrow.[55,56]

Although CART is non-parametric, not probabilistic in nature and did not produce any confidence intervals, the model developed using the 'learn' and 'validate' datasets were assessed or cross-validated on how well the model performed on the 'test' dataset. By doing this, a more realistic estimate of the prediction error could be obtained.[40,41,57]

Thirdly, the age distribution may have skewed the results. As highlighted in Chapter 3, Figure 2, those 0 and 1 year of age account for over 25% of the data. Any splits involving those 0 and 1 year of age could be significant simply due to the size of the datasets and the high volume of cases for these ages. Although these datasets were very large and sampling was not necessary, a balanced distribution across age

could have been achieved by sampling. Further clinical validation of these age splits may be required before they can be put into practice.

Finally, data quality issues around the activity and cost data may affect the results of the thesis. CIHI has conducted several national data quality studies over the last three years. For fiscal year 1999/2000, there was a 13.4% coding discrepancy associated with the MRDx and a 10% coding discrepancy with the principal procedure.[58-61] As a result, the incorrect coding of the MRDx and principal procedure may impact CMG assignment, and thus affect where the costs associated with a specific disease and/or treatment are assigned. In turn, this may create CMG-specific age splits that may not otherwise be significant and vice versa. Also, the costs of hospital services may not be captured or allocated consistently across case costing facilities and/or jurisdictions in Canada. For a given disease and/or treatment, some hospitals may have higher or lower costs than others, depending on the size of the hospital (i.e., teaching hospital versus a small community hospital) and geographical location (i.e., urban versus rural). These facility effects may create CMG-specific age splits that may not otherwise be significant and vice versa. CIHI is currently investigating this issue by exploring ways of adjusting costs relative to the hospital's case mix and creating inclusion/exclusion rules for these data.

4.6 How This Research Is Useful

Despite these limitations, the results of this thesis illustrate that using the CART algorithm can serve to better differentiate costs among acute care patients of different ages within the 0-17 year CMG age category. As such, the CART algorithm may be

useful for refining the 18-69 and 70+ year CMG age groups also being used in the current CMG methodology.

In addition, the CART algorithm and study design of this thesis will be useful in identifying age splits for the acute care grouper currently being redeveloped. With the recent launch of the CMG Grouper Redevelopment Project in Fall 2003, CIHI aims to redevelop the current CMG grouper to reflect current acute care practice patterns in Canada and take advantage of the newly implemented ICD-10-CA and CCI classifications systems. The results of the thesis can also be incorporated into the redeveloped grouper and validated by the various physician panels setup by CIHI. Furthermore, health resource indicators, such as ELOS and RIW, could be recalculated to better reflect the resource utilization characteristics of distinct age groups within the 0-17 year CMG age category. Calculation of these indicators would serve to further validate whether the CMG-specific or fixed age splits better represent the resource consumption patterns of the 0-17 year CMG age group.

The thesis research also has important practical applications. Refining the 0-17 year CMG age category will provide decision-makers and stakeholders with information that better represents how the pediatric population in Canada is utilizing hospital services in the acute care setting. In the province of Ontario, the Ministry of Health and Long Term Care (MoHLTC) would be interested in the refined 0-17 year CMG age group because it would affect the province's hospital funding formula, and thus the allocation of hospital resources for the pediatric population.[31,62] For example, one portion of a hospital's budget in Ontario is determined through recommendations by the Joint Policy and Planning Committee (JPPC) in Ontario.

The JPPC makes its recommendations using the funding formula below.[62]

Total Cost = Rate * Volume

where:

Total Cost = expected cost of services for a given hospital;
Rate = an estimate of the hospital's cost for treating each case;
Volume = an estimate of the annual number of cases a hospital should be treating.

The Volume portion of the formula relies upon hospital activity and is measured in weighted cases (the number of cases in a given CMG, Plx and age group * the RIW value).[62] In the current methodology, the RIW value for a given CMG and Plx level is the same for all ages, making the weighted cases calculation straightforward. However, if the results of Model 4 were implemented, the weighted cases calculation would be affected, as there would be a different RIW value for a given CMG and Plx level depending on the age group the patient falls into. The impact on weighted cases is illustrated in Table 12 below.

Table 12: Comparison of Weighted Cases in Hospital X for CMG 146 (Asthma), Plx 1 Using the 0-17 year Age Group and the Model 4 Age Groups

Age	Number of Cases in Hospital X	RIW Value		Weighted Cases	
		0-17 years†	Model 4‡	0-17 years	Model 4
< 1 Year	140	0.4342	0.6134	61	86
1 Year	60	0.4342	0.4662	26	28
≥ 2 Years	200	0.4342	0.4077	87	82
			Total Weighted Cases	174	195

† Represents the 2002 RIW value for CMG 146, Plx, Age 0-17 years.[2]
‡ The RIW values for Model 4 are hypothetical values used for illustrative purposes. Actual RIW values for Model 4 would have to be calculated by the Case Mix Department at CIHI, and these values may vary.

Although the difference in weighted cases in Table 12 may appear to be quite small, this difference is very large and significant when weighted cases are incorporated into the JPPC funding formula to distribute millions of dollars to Ontario hospitals. In the

example above, the difference in weighted cases would result in increased funding (i.e., 21 more weighted cases) for Hospital X.

In addition to hospital funding, more age groups within the 0-17 year CMG age category would impact program planning activities faced by hospital administrators, those who work in health services, decision support units, resource utilization management or hospital finance, and health professionals who work with the pediatric population.[2,31,62,63] For example, more age groups within the 0-17 year CMG age group would not only affect the number of weighted cases in a hospital, but may also impact long stay outlier trim points, ELOS and RIW. This in turn may affect the allocation of hospital resources and determine whether or not a hospital needs to open more beds or provide more specialized programs and services.

Most important, applying CMG-specific or fixed age splits across all CMG can better represent the resource consumption patterns of the pediatric population in acute care settings across Canada. Acute care facilities in Canada would be interested in the results of this thesis, as they are generalizable to the Canadian pediatric population. Although the thesis used linked cost data from Ontario, Alberta and British Columbia only, these data represent a broad spectrum of diseases/conditions, treatments and patient costs from small community hospitals to large teaching facilities. These data also represent all available Canadian acute care cost data for the 0-17 year population between fiscal year 1997/98 and 2000/01. Therefore, the age distribution for pediatric patients being seen at other Canadian acute care facilities not represented in the analysis would most likely be the same or similar as the age distribution illustrated in Figure 3. Acute care facilities across Canada could easily apply the results of this thesis

to their respective pediatric population. Although these age splits require clinical validation, they also create opportunities for future case mix research.

4.7 Future Research Considerations

Future research in this area should be explored, especially with respect to the following recommendations. Firstly, the CMG-specific and fixed age splits should be clinically validated by physician panels, such as those setup by CIHI for the CMG Grouper Redevelopment Project. Secondly, a similar study should be conducted using ICD-10-CA/CCI-based activity and cost data, as the new classification systems will impact how diseases and/or treatment are grouped, and thus split based on age. Finally, cost for all ages should be used to refine the current CMG age categories. Using data that are representative of all ages may identify new age groups for those in their mid- to late-teens.

4.8 Conclusion

Based on the results of this study and subject to clinical validation, the 0-17 year CMG age group in the current ICD-9-based CMG methodology should be divided into those:

- < 1 year of age;
- ≥ 1 year of age and < 2 years of age; and
- ≥ 2 years of age and ≤ 17 years of age.

The results from this thesis are not only important to the development and evolution of inpatient grouping methodologies, but represent unique findings from an original piece of research. This thesis is first known Canadian study to examine the refinement of age groups used in inpatient grouping methodologies. No other known inpatient grouper or study has used or considered using the resulting age groups from the statistical refinement of the different models explored in this thesis. In comparison to the CMG-Based Splits, the CMG-Specific Age Splits and the Fixed Age Splits produced a lower MSE. Using the 'Typical' dataset, the CMG-Specific Age Splits resulted in a 20% reduction in the MSE, whereas the Fixed Age Splits at 0.5 years and 1.5 years resulted in a 29% reduction. These results emphasize that cases currently in the 0-17 year CMG age group are not entirely homogenous with respect to consumption of hospital resources. Further refinement to this age group is required to better differentiate clinical and age effects related to treatment and care. More important, refinement to the current 0-17 year CMG age group, such as the recommendations above, would be most beneficial to stakeholders and decision-makers who play a central role in hospital funding and program planning.

References

1. Canadian Institute for Health Information. (12-03-2003). Case Mix Groups with Complexity Overlay and Age Adjustment. http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=casemix_cmgplx_e.

2. Canadian Institute for Health Information. (2002). *DAD Resource Intensity Weights and Expected Length of Stay 2002.* Ottawa: Canadian Institute for Health Information.

3. Canadian Institute for Health Information. (2002). *CMG™/PIx™ Directory 2002 (ICD-10-CA/CCI Version).* Ottawa: Canadian Institute for Health Information.

4. Canadian Institute for Health Information. (2003). *CMG™/PIx™ Directory 2003 (ICD-10-CA/CCI Version).* Ottawa: Canadian Institute for Health Information.

5. Canadian Pediatric Decision Support Network. (20-11-2002). Case Mix Group and Complexity Issues in Paediatrics (Unpublished).

6. Personal communication with Dr. Robin Walker, Professor, Department of Paediatrics, University of Ottawa, and Paediatrician, Division of Neonatology, Children's Hospital of Eastern Ontario. (18-03-2003).

7. Averill R. (1991). Development. In Fetter R, Brand D, Gamache D, eds. *DRGs: Their Design and Development.* Ann Arbor: Health Administration Press.

8. Averill, RF, Muldoon, JH, Vertrees, JC, Goldfield, NI, Mullin, RL, Fineran, EC, Zhang, M, Steinbeck, B, and Grant, T. (1998). The Evolution of Casemix Measurement Using Diagnosis Related Groups (DRGs).

9. Fetter R. (1991). Background. In Fetter R, Brand D, Gamache D, eds. *DRGs: Their Design and Development.* Ann Arbor: Health Administration Press.

10. Georgoulakis, J. (10-7-2003) Purpose/Goals of Groupers (Presentation).

11. Kulinskaya, E. (11-10-2003) International Casemix Research: Why and How. Washington, DC. Proceedings of the 19th International Case Mix Conference, 11-10- 2003.

12. Peacock Plomann M. (1982). Diagnosis Related Groups (DRGs). In Peacock Plomann M, ed. *Case Mix Classification Systems: Development, Description and Testing.* Chicago: The Hospital Research and Educational Trust.

13. Averill, RF, Goldfield, NI, Steinbeck, B, Grant, T, Muldoon, JH, Brough AJ, and Gay, J. (1997). Development of the All Patient Refined DRGs (APR-DRGs). Salt Lake City: 3M HIS.

14. Coffey R, Louis D. (2001). Case Mix in the USA: Fifteen Years of DRG-Based Hospital Financing in the United States. In Roger France F, Mertens I, Closon M,

Hofdijk J, eds. *Case Mix: Global Views, Local Actions*. Washington, DC: IOS Press.

15. Campbell L. (1998). CIHI Case Mix Tools. In Johnson L, Richards J, Pink G, Campbell L, eds. *Case Mix Tools for Decision Making in Health Care*. Ottawa: Canadian Institute for Health Information.

16. Freeman J, Fetter R, Park H, Schneider K, Lichtenstein J, Hughes J *et al.* (1995). Diagnosis-Related Group Refinement with Diagnoses- and Procedure Specific Comorbidities. *Med Care*, 338, 806-27.

17. Lichtig L, Knauf R, Parrott R, Muldoon J. (1989). Refining DRGs: The Example of Children's Diagnosis-Related Groups. *Medical Care*, 275, 491-506.

18. Mullin, RL, Vertrees, JC, Freeman, R, Castioni, R, and Tinker, A. (2002). Case-Mix Analysis Across Patient Populations and Boundaries: A Refined Classification System Designed Specifically for International Use. Salt Lake City: 3M HIS.

19. Palmer G. (2001). Case Mix in Australia: AR-DRG - A Smooth Process Involving States and a Commonwealth Government. In Roger France F, Mertens I, Closon M, Hofdijk J, eds. *Case Mix: Global Views, Local Actions*. Washington, DC: IOS Press.

20. Commonwealth Department of Health and Ageing. (2002). Australian Refined Diagnosis Related Groups, Version 5.0, Volume 1. Canberra: Commonwealth Department of Health and Ageing.

21. Duckett S. (2000). The Development of Australian Refined Diagnosis Related Groups: The Australian Inpatient Casemix Classification. *CASEMIX Quarterly*, 24, 115-20.

22. Mountney L. (1998). The Increasing Role of Casemix in Understanding the NHS in England. *CASEMIX Quarterly*, 01, 1-3.

23. NHS Information Authority. (2003) Version 3.5 HRG: Health Resource Groups Documentation Set. Winchester: NHS Information Authority.

24. Rodrigues J, Trombert-Paviot B. (2001). Case Mix in France: Is there a French Way from Problem to Paradigm in Healthcare? The DRG/PMSI SAGA. In Roger France F, Mertens I, Closon M, Hofdijk J, eds. *Case Mix: Global Views, Local Actions*. Washington, DC: IOS Press.

25. Patris A, Blum D, Girardier M. (2001). EfP: A Change in the French Patient Classification System. *CASEMIX Quarterly*, 34, 129-38.

26. Z / I / M Zentrum für Informatik und wirtschaftliche Medizin. (16-8-1999). Use of PCSs for Acute Inpatient Treatment in Europe and Some Other Countries. http://www.fischer-zim.ch/notes-en/PCS-in-Europe-9810.htm.

27. Roger France F, Mertens I, Closon M, Hofdijk J. (2001). *Case Mix: Global Views, Local Actions*. Washington, DC: IOS Press.

28. Paediatric Nursing Study Consortium. (1996). National Paediatric Nursing Study Phase 2. Adelaide: Paediatric Nursing Study Consortium.

29. Hanson R, Phythian M, Jarvis J, Stewart C. (1998). The True Cost of Treating Children. *MJA*, 169, S39-S41.

30. Canadian Institute for Health Information. (15-11-2002). The Graduate Student Data Access Program (GSDAP). http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=reqdata_gsdap_e.

31. McKillop, I, Pink, GH, and Johnson, LM. (2001). The Financial Management of Acute Care in Canada: A Review of Funding, Performance Monitoring and Reporting Practices. Ottawa: Canadian Institute for Health Information.

32. Canadian Institute for Health Information. (2003). *Canadian MIS Database: Moving Torward the Reporting of Hospital Financial Performance Indicators, 1999-2000 and 2000-2001.* Ottawa: Canadian Institute for Health Information.

33. Canadian Institute for Health Information. (2004). The MIS Guidelines and Service Recipient Costing (Unpublished).

34. Canadian Institute for Health Information. (9-4-2003). Discharge Abstract Database. http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=services_dad_e.

35. Canadian Institute for Health Information. (2002). *Privacy and Confidentiality of Health Information at CIHI: Principles and Policies for the Protection of Personal Health Information and Policies for Institution-Identifiable Information, 3rd edition.* Ottawa, Canadian Institute for Health Information.

36. Germanson T, Lanzino G, Kassell N. (1995). CART for prediction of function after head trauma. *J Neurosurg*, 83, 941.

37. Lemsky C, Smith G, Malec J, Ivnik R. (1996). Identifying risk for functional impairment using cognitive measures: An application of CART modeling. *Neuropsychology*, 10, 368-75.

38. Goldman L, Cook E, Johnson P, Brand D, Rouan G, Lee T. (1996). Prediction of the need for intensive care in patients who come to emergency departments with acute chest pain. *N Engl J Med*, 334, 1498-504.

39. Elsolh A, Stubeusz D, Grant G, Grant B. (1996). Outcome of AIDS patients requiring mechanical ventilation predicted by recursive partitioning. *Chest*, 109, I1584-I1590.

40. Breiman L, Friedman J, Olshen R, Stone C. (1984). *Classification and Regression Trees*. New York: Chapman & Hall.

41. Hastie T, Tibshirani R, Friedman J. (2001). *The Elements of Statistical Learning.* New York: Springer-Verlag.

42. Ripley B. (1996). *Pattern Recognition and Neural Networks.* New York: Cambridge University Press.

43. Salford Systems. (30-7-2003). An Overview of the CART Methodology. http://www.salford-systems.com/whitepaper.html - An.

44. StatSoft Inc. (2004). Classification and Regression Trees (C&RT). http://www.statsoftinc.com/textbook/stcart.html.

45. American Medical Rehabilitation Providers Association. (22-3-2002). Letter to Grace Carter with comments on the Draft RAND Report on the Inpatient Rehabilitation Facilities Prospective Payment System, July 13, 2000. http://www.amrpa.org/grace_carter_letter.htm.

46. StatSoft Inc. (2004). Discriminant Function Analysis. http://www.statsoftinc.com/textbook/stdiscan.html.

47. StatSoft Inc. (2004). Cluster Analysis. http://www.statsoftinc.com/textbook/stcluan.html.

48. Aldenderfer, MS and Blashfield, RK. (1984). Cluster Analysis. (07-044). *Sage University Papers series on Quantitative Applications in the Social Sciences*, ed. Lewis-Beck, MS. London: Sage Publications.

49. Everitt B, Landau S, Leese M. (2001). *Cluster Analysis.* London: Oxford University Press.

50. 2000 Annual Meeting of the Society for Academic Emergency Medicine. (23-5-2000). An Introduction to Classification and Regression Tree (CART) Analysis. http://www.saem.org/download/lewis1.pdf.

51. International Food Policy Research Institute. (15-07-2003) Classification and Regression Trees: An Introduction. (Technical Guide #3). http://www.ifpri.org/themes/mp18/techguid/tg03.pdf.

52. The SAS Institute. (2004). Tree Node. In *SAS System Help, The SAS System, Release 8.02 TS Level 02M0.*

53. Moser C, Kalton G. (1985). *Survey Methods in Social Investigation.* Burlington: Ashgate Publishing Group.

54. The SAS Institute. (2004). Predictive Modelling. In *SAS System Help, The SAS System, Release 8.02 TS Level 02M0.*

55. Efron B, Tibshirani R. (1998). *An Introduction to the Bootstrap.* New York: CRC Press LLC.

56. Davidson A, Hinkley D. (1997). *Bootstrap Methods and their Application*. New York: Cambridge University Press.

57. de Ville.B. (1990). Applying Statistical Knowledge to Database Analysis and Knowledge Base Construction. Washington, D.C. Proceedings of the Sixth IEEE Conference on Artificial Intelligence Applications. IEEE Computer Society. 1990.

58. Richards, J, Brown, A, and Homan, C. (2001) The Data Quality Study of the Canadian Discharge Abstract Database. Ottawa. Proceedings of Statistics Canada Symposium 2001. Statistics Canada. 16-10-2001.

59. Canadian Institute for Health Information. (2002). *Discharge Abstract Database Data Quality Study: Preliminary Year 1 Findings*. Ottawa: Canadian Institute for Health Information.

60. Canadian Institute for Health Information. (2002). *Discharge Abstract Database Data Quality Re-abstraction Study: Combined Findings for Fiscal Year 1999/2000 and 2000/2001*. Ottawa: Canadian Institute for Health Information.

61. Brown, A and Richards, J. (2001) Quality Measurements of the Canadian Discharge Abstract Database. Stockholm. Proceedings of the Quality Conference 2001. Statistics Sweden. 14-5-2001.

62. Ontario Joint Policy and Planning Committee. (12-3-2004). *Hospital Funding Report Using 2002/2003 Data*. Toronto: Ontario Joint Policy and Planning Committee.

63. Canadian Institute for Health Information. (1998). *Case Mix Tools for Decision Making in Health Care*. Ottawa: Canadian Institute for Health Information.

Appendix A: Glossary of Terms

Term	Definition
APIx	The analytical cell representing a given CMG, PIx and age category
AP-DRG	All Patient Diagnosis Related Groups
APR-DRG	All Patient Refined Diagnosis Related Groups
AR-DRG	Australian Refined Diagnosis Related Groups
CART	Classification and Regression Trees
CC	Complications and Comorbidities
CCI	Canadian Classification of Health Interventions
CCP	Canadian Classification of Diagnostic, Therapeutic and Surgical Procedures
CIHI	Canadian Institute for Health Information
CMG	Case Mix Groups – inpatient grouping methodology used in Canada
CPDSN	Canadian Pediatric Decision Support Network
Complexity	How complicated a patient is to treat based on diagnoses other than MRDx for which prolonged LOS and/or costly treatment might be reasonably expected.
CPHA	Commission on Professional and Hospital Activities
DAD	Discharge Abstract Database
DRG	Diagnosis Related Groups – inpatient grouping methodology used in the United States
ELOS	Expected Length of Stay – amount of time a patient can expect to stay in hospital for a given disease, condition or treatment

Term	Definition
Grouper	A system for categorizing patients based on: i. clinical criteria (e.g., diagnoses and/or procedures), ii. demographic information (e.g., age and/or sex), and iii. resource consumption (e.g., costs and/or LOS) Typically tries to relate the type of patients a hospital treats to the costs incurred by that hospital.
HCFA-DRG	Health Care Financing Administration Diagnosis Related Groups
H-ICDA-2	Hospital Adaptation of ICDA, Second Edition
HRG	Health Resource Groups – inpatient grouper used in Great Britain
ICD-10-CA	International Statistical Classification of Diseases and Related Health Problems, Tenth Revision – The Canadian Enhancement
ICD-9	International Statistical Classification of Diseases, Injuries, and Causes of Death, Ninth Revision
ICD-9-CM	International Statistical Classification of Diseases, Injuries, and Causes of Death, Ninth Revision - Clinical Modification
ICDA-8	International Classification of Diseases, Eighth Revision – Adapted
IR-DRG	International Refined Diagnosis Related Groups
LOS	Length of Stay – the amount of time a patient spends in hospital from the date of admission to the date of discharge
MCC	Major Clinical Category – used with the CMG methodology
MDC	Major Diagnostic Category – used with the DRG methodology
MRDx	Most Responsible Diagnosis – the diagnosis, which accounts for the greatest proportion of LOS; used with the CMG methodology.
MSE	Mean Squared Error
NACHRI	National Association of Children's Hospitals and Related Institutions
PCS	Patient Classification System (see Grouper)
PDx	Principal Diagnosis – the diagnosis for which the patient was admitted to the hospital; used with the DRG methodology.
Plx	Complexity Overlay (see Complexity)

Term	Definition
PPS	Prospective Payment System – pays a fixed amount for each patient treated within a particular health care setting
R-DRG	Refined Diagnosis Related Groups
Recursive Partitioning	An exploratory technique for uncovering structure in data by successively splitting a dataset into increasingly homogeneous subsets until it is no longer feasible to continue, based on a set of "stopping rules"
RIW	Resource Intensity Weights – relative values that describes the expected resource consumption of the "average" patient within a given CMG, Plx and Age Group
ROM	Risk of Mortality – likelihood of dying
Severity	Qualifies how seriously ill a patient is with respect to loss of function within a given illness/condition (i.e., mild, moderate or severe).
SOI	Severity of Illness – the extent of organ system loss of function or physiologic decompensation

Appendix B: Major Clinical Categories (MCC) versus Major Diagnostic Categories (MDC)

CMG Methodology	DRG System
MCC1 Nervous System	MDC1 Nervous System
MCC2 Eye	MDC2 Eye
MCC3 Ear,Nose,Mouth & Throat	MDC3 Ear, Nose, Mouth and Throat
MCC4 Respiratory System	MDC4 Respiratory System
MCC5A Cardiac -Circulatory System MCC5B Vascular -Circulatory System	MDC5 Circulatory System
MCC6 Digestive System	MDC6 Digestive System
MCC7 Hepatobiliary System & Pancreas	MDC7 Hepatobiliary System & Pancreas
MCC8 Musculoskeletal System & Connective Tissue	MDC8 Musculoskeletal System & Connective Tissue
MCC9 Skin, Subcutaneous Tissue & Breast	MDC9 Skin, Subcutaneous Tissue & Breast
MCC10 Endocrine & Metabolic	MDC10 Endocrine & Metabolic
MCC11 Kidney & Urinary Tract	MDC11 Kidney and Urinary Tract
MCC12 Male Reproductive System	MDC12 Male Reproductive System
MCC13 Female Reproductive System	MDC13 Female Reproductive System
MCC14 Pregnancy & Childbirth	MDC14 Pregnancy & Childbirth
MCC15 Newborns & Neonates	MDC15 Newborns & Other Neonates
MCC16 Blood, Blood-forming Organs & Immunological Disorders	MDC16 Blood & Blood Forming Organs & Immunological Disorders
MCC17A Lymphoma or Leukemia MCC17B Neoplasm –Unspecified	MDC17 Neoplastic Disorders (Haematological & Solid Neoplasms)
MCC18 Multisystemic or Unspecified Site Infections	MDC18 Infectious and Parasitic Diseases
MCC19 Mental Diseases & Disorders	MDC19 Mental Diseases and Disorders
MCC20 Does not exist; cases now included in MCC19	MDC20 Alcohol & Drug Use
MCC21 Injury, Poisoning & Toxic Effects of Drugs	MDC21 Injuries, Poisoning and Toxic Effects of Drugs
MCC22 Burns	MDC22 Burns
MCC23 Other Reasons for Hospitalization	MDC23 Factors Influencing Health Status & Other Contacts w Health Service

CMG Methodology	DRG System
MCC24 HIV Infections (AIDS)	MDC24 HIV Infections
MCC25 Significant Trauma	MDC25 Multiple Significant Trauma
MCC99 Ungroupable Data	MDC 99 Does not exist;

Appendix C: Number of Groups and Subgroups for Selected Patient Grouping Methodologies

System	Version	Country	Base Groups	Subgroups	Total No. Groups
CMG	2002 v3.0	Canada	478	370 w/ 4 Plx Levels 108 w/ No Plx Levels Plus subdivisions by 3 age groups	4,752
HCFA-DRG	1995 v12.0	USA	338	Some groups split by age & CC	492
R-DRG	1995 v10.0	USA	367	Some groups split by age & CC	1,170
AP-DRG	1995 v12.0	USA	N/A	Some groups split by age & CC	641
APR-DRG	2003 v20.0	USA	316	Most groups have 4 severity levels	1,258
IR-DRG	2001 v1.2	International (3M)	321	306 w/3 severity levels 10 w/ no severity levels 3 non-related O.R. w/ 3 severity levels 2 error	939
AR-DRG	2002 v5.0	Australia	665	Some groups split by age & CC	N/A
HRG	2003 v3.5	Great Britain	610	130 groups split by age or CC	N/A

Appendix D: Number of Cost Records for each CMG included in the Special and Typical Datasets

CMG	CMG Description	Special Dataset	Typical Dataset
1	Craniotomy Procedures	466	394
3	Spinal Procedures	174	146
5	Ventricular Shunt Revision	321	279
10	Neoplasm of Nervous System	292	239
18	Viral Meningitis	259	246
19	Infection except Viral Meningitis	148	118
22	Seizure and Headache	2,137	1,958
28	Other Nervous System Diagnoses	250	218
60	Major Eye Infections	127	114
63	Other Ophthalmic Diagnoses (MNRH)	112	103
77	Less Extensive Head and Neck Procedures	133	123
78	Cleft Lip and Palate Repair	552	491
83	Reconstructive ENT Procedures	353	327
85	Mastoid Procedures	194	178
89	Dental Extraction or Restoration (MNRH)	119	111
93	Tonsillectomy and Adenoidectomy Procedures (MNRH)	1,629	1,588
104	Influenza	564	502
109	Other ENT Infections	285	262
114	Sore Throat (MNRH)	431	397
115	Miscellaneous ENT Diagnoses (MNRH)	265	232
116	Croup (MNRH)	597	543
127	Major Respiratory Procedures	134	110
137	Respiratory Infections and Inflammations	189	156
143	Simple Pneumonia and Pleurisy	2,665	2,438
145	Tracheobronchitis	3,362	3,148
146	Asthma	4,014	3,710
147	Other Respiratory Diagnoses	680	606
182	Other Cardio-thoracic Procedures with Heart Pump without Cardiac Cath	522	442
194	Minor Cardio-thoracic Procedures without Heart Pump	220	201
218	Cardiac Cath without Specified Cardiac Conditions	197	187
237	Arrhythmia	124	112
251	Gastrostomy and Colostomy Procedures	389	322
253	Major Intestinal and Rectal Procedures	237	209
255	Less Extensive Esophageal, Stomach and Duodenum Procedures	195	160
258	Laparotomy	243	211
261	Complicated Appendectomy	609	544
262	Simple Appendectomy	1,778	1,626
264	Minor Gastrointestinal Procedures	150	130
269	Bilateral Inguinal/Femoral and Other Hernia Procedures	191	181
271	Unilateral Hernia Procedures (MNRH)	221	211
281	G.I. Hemorrhage	112	98
289	Inflammatory Bowel Disease	234	209

CMG	CMG Description	Special Dataset	Typical Dataset
290	G.I. Obstruction	178	163
294	Esophagitis, Gastroenteritis and Miscellaneous Digestive Disease	4,794	4,391
297	Other G.I. Diagnoses	535	489
317	Laparoscopic Cholecystectomy	140	125
356	Repair Hip and Femur Procedures	218	193
363	Back and Neck Procedures with Fusion	285	239
369	Major Lower Extremity Procedures	395	368
372	Major Upper Extremity Procedures	192	178
374	Minor Lower Extremity Procedures	732	678
375	Minor Upper Extremity Procedures	104	97
376	Miscellaneous Musculoskeletal Procedures	171	153
377	Wound Debridement and Skin Graft for Musculoskeletal Disorders	208	187
378	Soft Tissue Procedures (MNRH)	183	166
379	Other Musculoskeletal Procedures (MNRH)	339	311
380	Other Lower Extremity Procedures (MNRH)	340	313
392	Osteomyelitis	105	90
398	Other Inflammatory Arthritis	264	232
401	Other Musculoskeletal Malignancy	247	198
447	Cellulitis	556	509
454	Minor Skin Disorders	159	139
483	Diabetes	814	728
485	Nutritional and Miscellaneous Metabolic Disorders	1,314	1,229
487	Cystic Fibrosis	191	158
488	Inborn Errors of Metabolism	116	98
489	Endocrine Disorders	114	95
504	Major Urinary Tract Procedures	575	508
508	Minor Upper Urinary Tract Procedures	159	131
524	Nephrotic Syndrome	126	110
525	Nephropathy without Nephrotic Syndrome	109	99
527	Upper Urinary Tract Infection	481	423
529	Lower Urinary Tract Infection	605	555
551	Penis Procedures	100	89
552	Testes Procedures	120	112
554	Miscellaneous Male Reproductive System Procedures (MNRH)	230	216
579	Major Uterine and Adnexal Procedures without Malignancy	144	133
596	Miscellaneous Gynecological Diagnoses (MNRH)	106	101
609	Vaginal Delivery with Complicating Diagnosis	442	410
611	Vaginal Delivery	845	797
619	False Labour LOS < 3 Days (MNRH)	150	146
624	Antepartum Diagnosis	188	176
662	Femur or Pelvic Procedures for Trauma	531	460
664	Wound Debridement and Skin Graft for Trauma	387	348
666	Major Lower and Upper Extremity Procedures for Trauma	661	618
668	Miscellaneous Musculoskeletal Procedures for Trauma	108	97
670	Upper Extremity Procedures for Trauma	1,693	1,564
680	Femur or Pelvic Fractures and Dislocations	152	132

CMG	CMG Description	Special Dataset	Typical Dataset
683	Intracranial Injuries	221	181
687	Thoraco-abdominal Injuries	287	250
688	Weight Bearing Injuries	320	297
692	Wounds	266	246
694	Facial Injuries	130	112
695	Other Cranial Injuries	690	622
696	Upper Extremity Fractures	926	886
703	Other O.R. Procedures of Blood and Blood-forming Organs	130	110
704	Red Blood Cell Disorders	312	281
709	Coagulation Disorders	421	383
710	Reticuloendothelial and Immunity Disorders	670	630
726	Acute Leukemia without Major Procedures	660	539
730	Lymphoma and Chronic Leukemia	102	92
736	Chemotherapy	1,341	1,209
751	Septicemia	358	302
756	Post-operative and Post-traumatic Infections	110	103
757	Viral Illness	894	821
761	Fever of Unknown Origin	415	377
763	Other Infectious Diagnoses	102	97
766	Depressive Mood Disorders without ECT without Axis III Diagnosis	311	268
767	Depressive Mood Disorders LOS < 6 Days	165	147
777	Schizophrenia and Other Psychotic Disorders w/o ECT or Axis III Diagnosis	107	84
786	Disruptive Behaviour Disorders	509	427
787	Eating Disorders	181	150
792	Adjustment Disorders (MNRH)	366	323
804	Non-extensive Procedures for Injury or Complication of Treatment	256	234
813	Drug Reactions	652	585
818	Complications of Treatment	457	419
823	Minor Injuries and Trauma Diagnosis	214	173
832	Non-extensive Burns with Skin Graft	113	95
834	Non-extensive Burns without Burn Procedures	127	116
846	Aftercare following Surgery or treatment	658	603
851	Other Factors Causing Hospitalization	188	150
852	Procedures Cancelled (MNRH)	181	153
901	Non-extensive Unrelated O.R. Procedures	260	225
	Total	**59,580**	**54,022**

Appendix E: MSE for CMG-Specific Age Splits (Model 3), 'Special' and 'Typical' Datasets

CMG	Number of Leaves		MSE	
	Special Dataset	Typical Dataset	Special Dataset	Typical Dataset
001	1	4	63596864.00	146330000.00
003	2	7	60949099.00	43865043.00
005	1	1	4679496.00	4018794.00
010	5	1	17013939.00	22371585.00
018	1	4	1082064.00	2230422.00
019	1	1	28701971.00	16696443.00
022	6	9	4187128.00	2660962.00
028	1	1	49402892.00	21236543.00
060	1	1	59647905.00	66413410.00
063	1	1	1764633.00	2376508.00
077	1	3	10078652.00	8404272.00
078	1	1	3704518.00	7508595.00
083	2	1	67320009.00	63580985.00
085	5	3	49697863.00	55506932.00
089	1	8	192670000.00	3927959.00
093	1	5	487127.90	285984.50
104	1	1	17103452.00	3037301.00
109	1	4	1235441.00	3246809.00
114	2	3	1296767.00	3452912.00
115	5	3	4090818.00	4935547.00
116	1	1	13011762.00	14181969.00
127	1	3	552740000.00	152490000.00
137	3	1	110570000.00	110830000.00
143	1	5	14504072.00	9533704.00
145	2	2	17269848.00	12947626.00
146	5	4	13622236.00	14728194.00
147	1	5	28051587.00	8149319.00
182	5	5	93823485.00	237560000.00
194	3	3	13407062.00	8951754.00
218	3	4	7651109.00	6354062.00
237	1	4	16492428.00	5255648.00
251	1	2	1289900000.00	64852983.00
253	7	5	75529559.00	98646916.00
255	1	6	55140994.00	98194490.00
258	1	1	37765500.00	39775715.00
261	4	5	8129014.00	5645939.00
262	1	9	2240652.00	9322517.00
264	1	1	11139895.00	1421065.00
269	6	2	3956426.00	2915695.00
271	5	1	1381280.00	3953428.00

CMG	Number of Leaves		MSE	
	Special Dataset	Typical Dataset	Special Dataset	Typical Dataset
281	2	2	36398270.00	10842121.00
289	1	5	6199956.00	8756243.00
290	1	5	6842473.00	17309931.00
294	4	4	14236977.00	17032391.00
297	1	1	16850708.00	6810322.00
317	1	2	5996545.00	12278091.00
356	7	1	17045900.00	19173749.00
363	6	3	55771522.00	224860000.00
369	2	2	4570668.00	17717042.00
372	1	6	3938269.00	41347477.00
374	7	6	3951942.00	1573624.00
375	2	2	748693.90	1071509.00
376	1	5	30099462.00	30108408.00
377	1	4	9202968.00	21014436.00
378	1	8	2128325.00	2060637.00
379	1	1	2940344.00	1748272.00
380	1	4	2830793.00	32068535.00
392	1	3	22710755.00	28306872.00
398	1	3	20044230.00	21779412.00
401	6	1	21269615.00	20434237.00
447	5	7	3337588.00	2086545.00
454	5	1	5063101.00	7458039.00
483	1	5	3219719.00	4180460.00
485	3	6	7588704.00	5690236.00
487	5	9	44923581.00	30695419.00
488	5	1	28693551.00	53539351.00
489	2	1	11774041.00	77994636.00
504	5	1	20860635.00	12519987.00
508	6	1	11140757.00	5958980.00
524	1	7	10658042.00	8377833.00
525	1	5	1765779.00	2365007.00
527	5	4	4450420.00	4617042.00
529	5	5	5113169.00	3312367.00
551	1	1	5003841.00	2634176.00
552	9	1	826751.40	743860.20
554	2	1	1438957.00	648374.40
579	1	2	1584191.00	2377118.00
596	1	1	114710000.00	116270000.00
609	2	4	1494324.00	802216.20
611	1	4	7221661.00	7155201.00
619	1	1	818941.60	795940.00
624	3	1	311462.50	343367.80
662	1	4	24188587.00	26015403.00

CMG	Number of Leaves		MSE	
	Special Dataset	Typical Dataset	Special Dataset	Typical Dataset
664	4	1	58919035.00	42932323.00
666	1	5	36155549.00	49078624.00
668	1	2	62303641.00	65164554.00
670	1	6	1668992.00	1438676.00
680	1	1	20792427.00	14594006.00
683	6	1	39714465.00	51888269.00
687	6	1	168500000.00	23490623.00
688	1	1	2435224.00	3299535.00
692	1	5	1725113.00	1321180.00
694	5	1	1844473.00	2144389.00
695	1	6	9391485.00	3831660.00
696	1	1	10490743.00	47340232.00
703	1	1	6806540.00	45477024.00
704	1	1	10010613.00	8870479.00
709	1	1	10346903.00	12155516.00
710	9	6	8196047.00	9731856.00
726	4	4	97983897.00	393510000.00
730	1	2	22963210.00	11881788.00
736	6	6	6187095.00	10003622.00
751	1	3	17630015.00	172420000.00
756	4	1	17997470.00	14112232.00
757	3	4	1566149.00	8859674.00
761	7	6	1019063.00	1617485.00
763	4	1	2174054.00	2481393.00
766	6	5	44904774.00	66841594.00
767	1	2	11572995.00	11977840.00
777	2	1	39591460.00	43369395.00
786	4	6	181890000.00	182430000.00
787	2	1	315970000.00	217420000.00
792	1	5	9972964.00	13941879.00
804	2	1	25457153.00	17956431.00
813	3	1	2241953.00	3289675.00
818	4	2	46266585.00	45896726.00
823	1	2	66310402.00	99134353.00
832	1	1	383520000.00	864520000.00
834	1	1	3909916.00	5910834.00
846	1	6	1289801.00	1962026.00
851	3	3	4384004.00	2099419.00
852	1	1	293181.90	25889212.00
901	2	1	835680000.00	818160000.00

Appendix F: MSE for Fixed Age Splits (Model 4), 'Special' and 'Typical' Datasets

Fixed Age Split Model Number	Age Split		MSE	
	First Split	Second Split	Special Dataset	Typical Dataset
1	0.5	1.5	27844677.78	21295343.50
2	0.5	2.5	28498590.28	21862973.76
3	0.5	3.5	28741180.93	22128714.52
4	0.5	4.5	29358698.29	22693083.02
5	0.5	5.5	29368041.59	22696935.25
6	0.5	6.5	29389539.61	22718198.14
7	0.5	7.5	29420531.40	22755251.49
8	0.5	8.5	29422612.62	22747820.13
9	0.5	9.5	29960458.25	23186724.28
10	0.5	10.5	29942824.21	23174805.49
11	0.5	11.5	29933559.94	23217867.56
12	0.5	12.5	29957115.10	23241409.93
13	0.5	13.5	29938651.05	23214803.46
14	0.5	14.5	30024584.48	23320277.82
15	0.5	15.5	30030809.59	23322397.03
16	0.5	16.5	30049855.55	23363346.26
17	1.5	2.5	28544848.67	21886330.40
18	1.5	3.5	28797920.37	22163246.55
19	1.5	4.5	29419836.12	22728825.70
20	1.5	5.5	29425299.08	22728961.96
21	1.5	6.5	29439645.41	22744202.77
22	1.5	7.5	29461905.38	22770875.63
23	1.5	8.5	29460491.95	22762030.14
24	1.5	9.5	29997475.50	23201757.35
25	1.5	10.5	29979816.64	23185768.79
26	1.5	11.5	29964141.44	23224063.11
27	1.5	12.5	29987929.24	23249715.67
28	1.5	13.5	29969116.42	23223526.97
29	1.5	14.5	30050480.10	23325624.58
30	1.5	15.5	30058298.13	23329899.65
31	1.5	16.5	30085341.10	23372024.44
32	2.5	3.5	28816047.69	22159152.36
33	2.5	4.5	29427365.26	22721925.88
34	2.5	5.5	29434616.45	22722308.58
35	2.5	6.5	29444201.04	22729006.99
36	2.5	7.5	29465234.63	22754772.23
37	2.5	8.5	29462691.83	22745741.17
38	2.5	9.5	30000413.53	23188402.45
39	2.5	10.5	29981898.11	23170856.80
40	2.5	11.5	29968310.43	23210025.06
41	2.5	12.5	29995813.94	23238285.07
42	2.5	13.5	29978519.11	23210268.49
43	2.5	14.5	30060657.19	23310073.87
44	2.5	15.5	30062747.34	23304254.50

Fixed Age Split Model Number	Age Split		MSE	
	First Split	Second Split	Special Dataset	Typical Dataset
45	2.5	16.5	30086487.61	23345054.49
46	3.5	4.5	29433707.58	22734611.43
47	3.5	5.5	29440139.27	22733845.54
48	3.5	6.5	29447315.00	22740078.47
49	3.5	7.5	29470657.23	22769056.43
50	3.5	8.5	29468716.90	22761453.40
51	3.5	9.5	30005353.65	23203021.66
52	3.5	10.5	29989278.04	23186555.66
53	3.5	11.5	29976714.49	23227207.20
54	3.5	12.5	30000346.46	23253958.93
55	3.5	13.5	29980818.12	23226816.09
56	3.5	14.5	30065042.14	23331553.27
57	3.5	15.5	30068571.01	23328883.13
58	3.5	16.5	30096903.99	23369989.21
59	4.5	5.5	29439112.99	22737879.77
60	4.5	6.5	29443794.51	22743757.51
61	4.5	7.5	29465937.74	22771201.84
62	4.5	8.5	29463442.37	22762321.66
63	4.5	9.5	30000928.18	23205361.79
64	4.5	10.5	29982253.76	23189006.21
65	4.5	11.5	29971343.39	23231033.54
66	4.5	12.5	29994501.74	23255675.65
67	4.5	13.5	29975616.73	23226816.08
68	4.5	14.5	30059205.47	23328887.38
69	4.5	15.5	30058626.99	23327966.37
70	4.5	16.5	30078440.41	23364357.28
71	5.5	6.5	29463357.39	22758969.58
72	5.5	7.5	29484483.03	22784714.72
73	5.5	8.5	29481115.31	22775347.56
74	5.5	9.5	30019241.61	23219871.60
75	5.5	10.5	30003405.74	23206015.52
76	5.5	11.5	29993322.79	23247842.19
77	5.5	12.5	30014651.16	23270636.76
78	5.5	13.5	29997290.26	23243545.13
79	5.5	14.5	30077951.32	23344100.52
80	5.5	15.5	30080251.53	23345538.85
81	5.5	16.5	30102228.01	23380180.94
82	6.5	7.5	29501060.73	22798662.71
83	6.5	8.5	29497953.15	22789828.70
84	6.5	9.5	30036053.45	23236198.52
85	6.5	10.5	30022128.23	23222529.08
86	6.5	11.5	30013689.57	23264956.38
87	6.5	12.5	30032971.90	23285486.07
88	6.5	13.5	30018042.35	23261749.91
89	6.5	14.5	30095088.07	23358036.47
90	6.5	15.5	30095529.62	23357114.50

Fixed Age Split Model Number	Age Split		MSE	
	First Split	Second Split	Special Dataset	Typical Dataset
91	6.5	16.5	30115332.06	23392072.59
92	7.5	8.5	29511630.99	22808075.23
93	7.5	9.5	30049671.74	23255292.12
94	7.5	10.5	30036579.05	23241866.55
95	7.5	11.5	30028035.08	23283785.02
96	7.5	12.5	30046804.16	23301982.19
97	7.5	13.5	30030426.19	23279126.12
98	7.5	14.5	30103614.15	23369813.92
99	7.5	15.5	30103861.30	23369549.79
100	7.5	16.5	30126619.82	23407336.81
101	8.5	9.5	30057503.34	23258352.67
102	8.5	10.5	30043410.65	23243622.73
103	8.5	11.5	30034914.42	23285224.54
104	8.5	12.5	30051888.03	23301805.95
105	8.5	13.5	30036278.51	23278735.60
106	8.5	14.5	30106369.20	23366613.05
107	8.5	15.5	30107355.75	23366152.10
108	8.5	16.5	30130722.44	23402409.73
109	9.5	10.5	30046147.37	23237621.88
110	9.5	11.5	30036220.67	23278181.69
111	9.5	12.5	30051556.10	23292006.41
112	9.5	13.5	30034965.29	23265311.64
113	9.5	14.5	30108253.91	23356560.82
114	9.5	15.5	30104301.15	23351295.57
115	9.5	16.5	30128022.09	23386263.38
116	10.5	11.5	30035418.51	23284191.38
117	10.5	12.5	30050010.16	23295650.19
118	10.5	13.5	30039720.23	23275003.04
119	10.5	14.5	30114773.96	23363328.85
120	10.5	15.5	30110251.53	23355721.50
121	10.5	16.5	30125172.61	23386870.70
122	11.5	12.5	30054065.29	23301339.33
123	11.5	13.5	30041895.08	23280840.06
124	11.5	14.5	30118448.80	23369515.05
125	11.5	15.5	30113910.74	23361149.57
126	11.5	16.5	30125147.69	23386925.25
127	12.5	13.5	29998253.89	23227320.29
128	12.5	14.5	30072529.74	23309916.75
129	12.5	15.5	30071076.04	23304040.71
130	12.5	16.5	30083806.27	23329602.62
131	13.5	14.5	30061649.42	23300535.62
132	13.5	15.5	30052519.72	23284512.66
133	13.5	16.5	30066992.50	23306677.93
134	14.5	15.5	30077304.18	23324505.57
135	14.5	16.5	30095320.07	23344121.06
136	15.5	16.5	30102279.69	23354439.99