2018-10-23

The Economics of Big Data in Healthcare: A Case Study of the ICD-9 to ICD-10 Transition

Khair, Shahreen

UNIVERSITY OF CALGARY

The Economics of Big Data in Healthcare: A Case Study of the ICD-9 to ICD-10 Transition

by

Shahreen Khair

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF ARTS

GRADUATE PROGRAM IN ECONOMICS

CALGARY, ALBERTA

OCTOBER, 2018

Abstract

This thesis attempts to provide a methodological and qualitative analysis of the trend in data usage in the healthcare industry and highlight the importance of economic evaluation in big data for healthcare. The paper explores the need for economic evaluation of the current data usage scenario in Canada via a case study of the ICD-9 to ICD-10 transition and illustrates the gaps in data availability. Detailed formulae for calculating the costs and benefits of the transition is provided, while a comprehensive guideline illustrates how the cost calculations would differ between the different provinces of Alberta, Manitoba, Saskatchewan and British Columbia. The paper also presents an overview of the upcoming ICD-11 transition and provides a framework for conducting an economic evaluation on the ICD-10 to ICD-11 transition. In conclusion, the thesis finds that although considerable progress has been made, significant improvements in data utilization is required to reap the full benefits of health data available and enable efficient economic evaluation.

Keywords: Big Data, Health, Economic Evaluation, ICD-9, ICD-10, Digitization, Costs, Benefits, Research, Data Quality, Cost-Benefit Analysis.

Acknowledgements

I am grateful to my professor, Dr Mingshan Lu, for her constant guidance and for helping me formulate my thoughts and ideas into words on this paper. She has helped me utilize my knowledge and shown me the way to combine healthcare and economics in a comprehensive manner. I have learnt a lot from her and hope to continue learning.

I am also very grateful to my committee members Dr Hude Quan and Dr Lasheng Yuan for their insightful feedback and guidance, without which this paper would not have become what it is today.

I would like to thank Dr Catherine Eastwood and Chelsea Doktorchik, whose feedback has enriched the content of this paper.

I am grateful to all my professors at the University of Calgary and BRAC University, who have taught me through my graduate and undergraduate studies.

My parents and sister have been an amazing support system for me throughout all my academic endeavours and my gratitude towards them is endless. I am very grateful to my husband for his encouragement, help and for being an unyielding pillar of strength for me. I would like to thank my parents-in-law for their constant encouragement and prayers.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Abbreviations

ADT: Admissions, Discharges and Transfers
AHS: Alberta Health Services
ANP: Alberta Netcare Portal
BD2K: Big Data to Knowledge Initiative
CAC: Computer Assisted
CBA: Cost-Benefit Analysis
CDVC: Care Delivery Value Chain
CEA: Cost Effectiveness Analysis
CHR: Calgary Health Region
CIHI: Canadian Institute for Health Information
CUA: Cost Utility Analysis
DI: Diagnostic Imaging
DIMR: Analytics, Data Integration, Measurement and Reporting
EHR: Electronic Health Record
EHRDSC: EHR Data Stewardship Committee
EMR: Electronic Medical Records
HEOR: Health Economics and Outcomes Research
HIA: Health Information Act
HIEC: Health Information Executive Committee
HIM: Health Information Managers
HIS: Health Information System
HISCA: Health Information Standards Committee for Alberta
IAM: Identity and Access Management System
IBM: International Business Machines Corporation
ICD: International Classification of Diseases and Related Health Problems.
ICS: International research Institute for Climate and Society
ICTC: Information and Communication Technology Council
OIPC: Office of the Information and Privacy Commissioner of Alberta
PACS: Picture Archiving and Communication System
PCORnet: Patient Centered Research Network
PHANSC: Provincial Health Analytics Network Sponsor Committee
pHIE: Provincial Health Information Exchange
PHP: Personal Health Portal
PHPSC: Personal Health Portal Steering Committee
PIN: Pharmaceutical Information Network
PPMS: Pharmacy Practice Management Systems
PTSD: Post-traumatic Stress Disorder
QALY: Quality-Adjusted Life-Year
S2S: System to System

SDP: Seniors Drug Program
SNOMED CT: Systemized Nomenclature of Medicine- Clinical terms
WHO: World Health Institution
XDS: Cross Enterprise Document Sharing

1 Introduction

The Canadian healthcare system has recently seen a spurt of growth in digitization. Over the last decade, the advent of data usage in various sectors including healthcare, has been increasing at a fast pace. This has largely been caused due to a large amount of data being recorded by pharmaceutical companies and hospitals. With digitization of medical records, the data repositories are becoming more and more valuable for researchers, medical practitioners as well as the government. Digital records of healthcare aid in better disease management, payment and funding of healthcare facilities, allocation of resources and in research and innovation. The synthesizing and documentation of data on a large scale is being referred to as big data. Big data has been an emerging concept in a broad spectrum of activities and industries worldwide.

Canadian healthcare spending amounts to 11.5% of overall GDP. In 2018, the healthcare expenditure budget is expected to amount to $242 billion or $6604 per person. (Canadian Institute for Health Information, 2018). As economic activities become vaster and faster paced, it is important that budgetary decisions are taken and implemented accurately. Data driven health records are the best way to ensure the effective and economic utilization of limited resources.

This thesis provides an extensive review of how big data is changing the way research and clinical practices are being conducted, as well as highlighting the need for economic evaluation in big data for the healthcare industry. It will use the International Classification of Diseases and Health Problems (ICD)-9 to ICD-10 transition, as a case study to illustrate the direction of current data usage and the way forward, including addressing the gaps in data availability and providing recommendations. ICD is a system of codes for health data and related diseases and serves a myriad of professionals including healthcare practitioners, researchers, hospital personnel and the government. Coded data helps in planning budgets, allocating resources efficiently, formulating policies, in research work and as an indication and prevention of upcoming diseases. The paper provides detailed formulae for calculating the costs and benefits of the ICD-9 to ICD-10 transition. The formulae would incorporate coded data from mortality and morbidity statistics from hospitals. A comprehensive guideline showing how the cost calculations would differ among different provinces for a Canada wide cost-benefit analysis is provided, which has been compiled based on focus group interviews with the coding managers in the provinces of Alberta, Manitoba, Saskatchewan and British Columbia. The case study focuses on the challenges faced in conducting an economic evaluation on big data, specifically due to the lack of data availability for the specific timeline of the ICD transition, as well as due to the difficulty in retrieving the required data related to cost and benefit calculations.

As Canada moves towards implementing the ICD-11 coding practice this year, the ICD-9 to ICD-10 transition case study is the best example to illustrate the big data utilization state in Canada, as it provides a foundation for economic evaluation of the ICD-10 to ICD-11 transition. An evaluation of the economic impacts of the ICD-9 to ICD-10 transition hence provides an in-depth analysis for understanding whether the benefits of implementing the new coding system would exceed the costs or not.

The cost estimation could not be completed as data for the years 2000 to 2003 is not available for certain variables in the formulae. Hence, for the costs, the estimations by Calgary Health Region (CHR) Business Case (2001) is provided in Appendix D. An extensive benefit estimation methodology (see Appendix E) has been created using benefit scenarios, based on the RAND report by Libicki & Brahmakulam (2004). These scenarios depict the potential benefit of using the ICD-10 coding system for healthcare and hospital services, instead of ICD-9 coding system. However, after numerous interviews with coding specialists, researchers, physicians and clinicians, it was found that information and data sources for the required variables in the benefit estimation formulae is not available. Moreover, it is difficult to apply the benefit scenarios in the Canadian context. This is because, ICD coding is not used by physicians to deliver healthcare or for hospital management. For funding and billing purposes, Canada still uses ICD-9 codes. Hence, it is difficult to construct outcome measures without the required structured data. This shows the lack of usable data availability in healthcare research. In addition, the report by Calgary Health Region also did not outline measures to calculate the benefits and described the benefits from implementing ICD-10. Hence, in order to calculate the benefits, a crude estimation for a data quality and quantity measure, which depicts that a higher number of published papers used ICD-10 codes for research, compared to ICD-9 codes, has been constructed. Thus, the data quality and quantity measure made the argument that ICD-10 has better data for research use due to the higher number of codes, specificity, ease of usage and understanding of the ICD-10 coding system, compared to the ICD-9 coding system. However, this simple method of enumerating data quality has various pitfalls and cannot be taken as a standalone depiction for data quality. This is because, the increase in the number of publications could have been due to an increase in research funding and/or a rising interest in the economic evaluation of the healthcare sector in recent years.

The results of the case study show that data is currently only utilized for a limited amount of research by academicians and is not applied by clinical practitioners in the provision of healthcare. With advancements in machine learning and the huge availability of health data, it is vital that the healthcare industry exploits the advantages big data brings to the table. Moreover, the case study also provided insights on how the cost and benefit calculations would change in the coming years due to payment reforms and changes to funding allocation in hospitals. Recommendations were made on how the actors in the healthcare industry, manufacturers, pharmaceutical companies, physicians and other end users of healthcare could keep up the momentum of data analytics in healthcare delivery.

The paper has been divided into six chapters. Chapter 1 provides an overview of the thesis and introduces the arguments made in the paper. Chapter 2 talks about the movement and value of big data in healthcare. Chapter 3 explains the importance for economic evaluation in big data in healthcare. Chapter 4 introduces the case study on ICD-9 to ICD-10 transition and provides the cost and benefit calculation formulae. This chapter validates the RAND and CHR cost and benefit framework and highlights the challenges in employing the cost and benefit estimations in the Canadian context. The chapter analyses the case study on ICD-9 to ICD-10 transition to illustrate the current gaps in data usage and economic evaluation and explains the policy implications of big data in healthcare services. Chapter 5 provides recommendations the health sector can pursue in the future to ensure the momentum of big data in health and Chapter 6 concludes the paper.

2 The Movement of Big Data in Healthcare

As medical records, pharmaceutical research and healthcare analytics are being aggregated into electronic platforms, Canada is experiencing a paradigm shift in digitization in the healthcare industry. In the last decade, data collection and synthesis of information has increased dramatically. Consequently, our use of technology has risen as well. The amalgamation of these two increases has led to the formation of big data and has resulted in industries and businesses gaining allocative and resource efficiency.

Big data is now being utilized to predict upcoming epidemics, infections and diseases and as a signal to act before diseases strike by taking preventative actions. Dutton (2014) reports that after analysing wind and dust conditions between the years 1986 and 2006 by the International research Institute for Climate and Society (ICS), it was found that the severity of meningitis outbreaks in Niger, a country in Western Africa, was linked to stronger winds and heavy dust. This allowed the health authorities to administer the vaccines at the correct time in order to ensure that the medicine was most effective to counter the meningitis outbreak which was expected to occur one to two months later. Climate change data analysis helped in predicting outbreaks of leishmaniasis, a parasite killer which affected 12 million people in 90 countries by causing skin ulcers and attacking the spleen and liver. ICS also predicted malaria outbreaks by monitoring rainfall and irrigation, which are favourable conditions for mosquitos to breed. If the weather is between 19C to 32C, malaria develops inside mosquitos; if the weather is cooler, malaria takes longer to develop. Hence, health authorities are able to prepare for the outbreak up to two months before with accurate rainfall forecast. (Dutton, 2014).

Moreover, the treatments are changing rapidly due to improvements in technology, which in turn is being driven by advancements in data. The increase in the usage of devices such as Fitbit, Samsung Gear Fit, Jawbone, etc. have resulted in patients to be able to detect diseases, plan their diet, monitor critical health indicators and thus live a healthier lifestyle. Apple and IBM have come together to establish a health data platform which will enable Apple and iPhone users to share data to Watson Health, a cloud healthcare service by IBM.

Adams (2018) also reports the use of data and wearable technology in combating cancer by storing data responses on the patient's quality of life, sleep patterns and general mobility, to educate the physicians and other health professionals on how best to provide therapy for a certain cancer patient.

2.1 Big Data in Health

Healthcare analysis using big data is expected to increase in the next few years as advancements are made in the usage of data to power research, enhance patient outcomes, influence the growth in predictive analysis and enable patient and hospital payments (including performance-based pay). These are few of the recurring themes in the discussion for improving the health sector. Usage of big data in the health industry would enable enhanced patient empowerment via availability of information, efficient utilization of resource and funding allocation to hospitals by identifying at-risk patients and population health hazards ahead of time. Taking preventative steps and adopting holistic measures to follow up with patients will result in better patient outcomes in terms of healthcare. (Weir, 2017).

The most profitable retailer in Japan, Seven-Eleven Japan, changed the course of its business decision-making and marketing by turning to data-driven information from its 200,000 part-time sales clerks in the 1970s. This was done by sending information to the sales clerks regarding which items were being sold in every store, which enabled the clerks to order inventory based on available aggregate information. (Ross, Beath & Quaadgas, 2013).

Tibben and Wamba (2018) explore the potential of big data in healthcare by using the care delivery value chain (CDVC) framework. This framework re-organizes healthcare delivery to increase and improve the patient outcomes and reduce costs. They find that there is a need for management and policy development in big data in health in order to increase patient healthcare outcomes as well as to achieve efficiency in health services delivery. (Tibben & Wamba, 2018).

The National Institutes of Health in the United States has launched Big Data to Knowledge Initiative (BD2K) to increase access to data for the biomedical research community. Academic institutions, healthcare systems, life sciences companies and advocacy organizations have also been brought together by collaborations between Optum Labs and National Patient-Centered Research Network (PCORnet). (Shah & Pathak, 2014).

In 2011, around $2 million USD was provided to hospitals and doctors to establish digital records and EMRs. Philips, with the help of its HealthSuite Digital Platform, which collects patient health and lifestyle information and allows healthcare providers to analyse data, has generated over one-third of its sales from healthcare. Insurance companies such as Humana and UnitedHealth also use data analytics on claims data to target patients with high risk of health problems and provide them with information about nutrition, exercise, etc. (Gerdeman, 2015).

In Canada, Alberta has been at the forefront of establishing a data movement in its province's hospitals and clinics. The Alberta Electronic Health Record (Alberta EHR), a provincial program established in 1997, is a comprehensive and integrated health information bank which allows Health information access to authorized personnel. All of the data sharing is done in a secured environment. Alberta Health and Alberta Health Services (AHS) are responsible for the e-health systems. It is different from the usual electronic health record information system which only provides access to a limited number of healthcare clinics and hospitals. The Alberta Netcare Portal (ANP) and in some instances the System to System (S2S) grants access to the Alberta EHR data. (Robillard & Tolfree, 2015).

In spring 1999, Alberta Wellnet, the predecessor of Alberta Netcare, employed the Seniors Drug Program (SDP) in two hospitals of Edmonton. The SDP system provided a way for viewing medical information of seniors by pharmacists, hospital staff and physicians. The medical staff could thus have access to accurate information about the patients in their care. (Robillard & Tolfree, 2015).

In 2004, the Capital Health netCARE platform was formed which evolved to the Alberta Netcare Portal 2006 (ANP 2006). In January 2008, Capital Health netCARE and ANP 2006 joined to form ANP. The Alberta EHR connects more than fifty e-health systems. It is comprised of the following four components, which support the network in various ways:

1. Access Tools: these do not have any data by themselves but allow users to view health data from repositories and registries in the Alberta EHR.
2. Registries: store and provide data regarding places, events, identities and tools for identification purposes.
3. Repositories: store and maintain data on diagnostics, treatment and care information
4. Infrastructure: enables repositories, registries and access tools to transmit health data to users.

Plans to provide patients with drug information and allow the patients to enter personal health information (such as data from blood glucose monitors, blood pressure and track theraputic activities such as cardio exercises) are in the pipeline as Alberta Health incorporates interactive features in the Personal Health Portal (PHP). PHP is a web-based site which will allow Albertans to access their own personal health information in Alberta EHR as well as record their health information.

In addition, the report by Robillard and Tolfree (2015) outlines the following plans to integrate the health system with data in the future:

I. ePrescribing will replace the faxed and paper prescriptions in the Pharmaceutical Information Network (PIN). Plans to integrate laboratory information by standardizing the terminology used for lab test panels, including results and orders, will increase the accuracy and ease of usage of the laboratory information in Alberta EHR.

II. Cross Enterprise Document Sharing (XDS) is a digital file sharing network which would provide an interoperable and vendor-neutral platform for sharing information across provinces.

III. The eReferral Project will enable patient referral and transition across different Alberta health systems. This would result in an efficient method of reffering critically ill patients to designated and required care and physicians, reducing the wait times.

IV. The pHIE (Provincial Health Information Exchange) is a routing system used internally by AHS and externally to physician office systems, to exchange health information. This system would help to deliver messages accurately between different systems. Other networked health information systems include Identity and Access Management System (IAM), Diagnostic Imaging (DI)/Provincial Picture Archiving and Communication System (Provincial PACS) and Electronic Referral Systems. (Robillard & Tolfree, 2015).

The Alberta Peer-to-Peer Electronic Medical Records (EMR) Network Program was launched in October 2015 by Canada Health Infoway in partnership with Alberta Medical Association. This program supported workshops and development and growth of Peer Leaders in order to optimize the usability of EMR in Alberta. The program enabled sharing of EMR process information among peer leaders (physicians and team members). (Alberta Peer-to-Peer EMR Network Program, 2016).

However, use of data and technology has its disadvantages. The collection of information contributes to concerns regarding privacy of patient information as well as the risk of cyber-attacks since the newer technology uses the int ernet for most clinical data and for mobile wearable devices. Diabetic patients have been warned by the Johnson and Johnson Company about possible changes in insulin doses due to cyber-attacks by hackers. Moreover, confidentiality of patient information and data piracy are critical concerns. (Adams, 2018).

Some pharmaceutical companies may try to misuse this information by targeting customers to increase sales. The Office of the Information and Privacy Commissioner of Alberta (OIPC) has determined a method to ensure privacy and protection of health information collected by the Pharmacy Practice Management Systems (PPMS). This includes collaboration between the Pharmacies, the PPMS vendor and Alberta Health. The result of the collaboration is the "PIA Template for Real Time Integration of PPMS with the PIN" document, which will ensure real time integration of PPMS and privacy, overseen by OIPC. The Health Information Act (HIA) governs the structure and development of Alberta EHR and protects the collection and disclosure of health information from healthcare providers. (Robillard & Tolfree, 2015).

The Ministry of Health along with EHR Data Stewardship Committee (EHRDSC), Health Information Executive Committee (HIEC), EHR Sponsors Committee, Personal Health Portal Steering Committee (PHPSC), Provincial Health Analytics Network Sponsor Committee (PHANSC), IM/IT Strategy Committee and Health Information Standards Committee for Alberta (HISCA) oversee the privacy and security of Alberta EHR. De-identifying patient information and authentication of accessing sensitive data provide enhanced privacy and security of data in Alberta EHR. Masked information only reveals the first and last name, date of birth, gender and personal health number in the patient's file. (Robillard & Tolfree, 2015).

2.2 Value of Big Data

As defined by Armah (2013), big data has four main attributes. These are volume, variety, velocity and value. Big data encompasses *volumes* of data which are usually too large for typical databases to capture, process and analyse. Analysing and storage of such an enormous amount of data requires specialized and planned software. In addition, the designing and planning of such specialized software requires time and effort and is also very expensive most times. The data sources are *varied* and only about 10% of the data is in structured format which can easily be analysed. Structured data is data which is collected by companies, time series data collected by statistical agencies for the specific purposes of statistical and economic analysis. Unstructured data includes emails, road traffic information and audiovisual data. These unstructured data cannot be processed as is and requires to be brought into a structure.

Data gathered from social media, internet and networked transactions are continuously being updated. This results in a high *velocity* and volume of data which can only be handled by specialized software. The fourth attribute of big data is the way in which insightful information can be gleaned from the data, which will be *valuable* to corporations and for research. (Armah, 2013).

Groves, Kayyali, Knott & Kuiken (2013) note that the demand for big data has been escalating in the recent years. Due to rising costs for healthcare and shifts in reimbursement trends, a new wave of change in the clinical health provision is underway. In the US, the cost of healthcare represented 17.6 percent of GDP, which was about $600 billion more than the benchmark.

It is a usual practice for physicians who are rewarded under a fee-for-service system to be concentrated on the volume of treatment provided, rather than on the quality and healthcare outcomes. As a result, physicians do not review the results of the outcomes which would enable them to understand if patients are satisfied or not. As the healthcare scenario changes with risk-sharing models entering the payment structures, where payors provide reimbursement for drugs which produce measured results in patient care improvement, there is now an incentive for health providers to store and disseminate health data more eagerly. (Groves, Kayyali, Knott & Kuiken, 2013).

Moreover, with rising demand for information and data, supply of health information data has increased. This data is collected from claims and cost reimbursement information, pharmaceutical research and development information on drugs, side effects of drugs and the effect on the body, patient behaviour data on preferences, patient's finances, etc. In the US, different large firms have joined to collaborate on utilizing big data. These include Premier, OptumInsight for United Health, ActiveHealth for Atena, and HealthCore for WellPoint. (Groves, Kayyali, Knott & Kuiken, 2013).

The applications of big data in the real world is extensive. It is possible to gain useful insights into household behaviour, future expectations of customers, direction of research and product demand, financial stability and other economic indicators of a country by successfully combining big data and analytics. Walmart and Target collect data on consumer spending, inventories and wholesale prices, while banks collect data on household and businesses. The collection of these data fuel analysis about consumer preferences and result in targeted advertisements to consumers. (Einav and Levin, 2014). As we move into an era of digitization, for example into using more cashless transactional methods, the data generated from electronic payments could be used to analyse the trends in people's purchasing patterns. This could give indications of how the labour and housing markets will behave in the future, prompting policy makers to predict future economic growth and formulate policies accordingly. Mclaren and Shanbhogue (2011) have found from their research that including the results from internet search for labour market growth analysis improved the forecast performance. European countries use the data from point of scale scanners to compile the consumer price indexes. Statistics Netherlands incorporated super market scanner data in its CPI in June 2002, while Swiss Federal Statistical Office utilized the same method in calculating its price index. (Armah, 2013).

Technological advancements make it easier to share information. The advancements have made the processes of collecting, storing and analysing data cheaper and faster. This has resulted in more affordable EMR systems and has allowed the hospitals and pharmaceutical companies to ensure privacy. The specialized computer systems allow for more advanced data analytics by combining data across different data pools. Advanced analytics could for example show that which of the three treatments a patient has been given, resulted in the highest response rate in reducing a cancer tumor. (Groves, Kayyali , Knott & Kuiken, 2013).

For healthcare, the applications of big data have translated to healthcare analytics encompassing patient information, funding options and evaluation, and evidence-based research. Companies and start-ups are using the electronic data driven platform and machine learning technology to build user friendly applications which enable the patients' easier access to professional physician help. Patients can hope to convey their concerns to physicians even without the need to be physically present at clinics. Telemedicine, a recently built technology, is a way for providing care to patients using telecommunication technology. It would improve healthcare access for patients in rural areas since it allows transmission of medical as well as health informatics data and imaging between different locations. It would be possible to virtually receive feedback on patient's symptoms and be diagnosed. This would enable faster care and reduce the long wait times currently plaguing the Canadian healthcare system.

The value of healthcare system will come from adjusting and balancing the costs of healthcare with patient outcomes, in other words when the costs and benefits are balanced. This value will come about through the collaboration of patients, care providers, quality healthcare provision and innovation through new therapies and approaches to deliver care. It is estimated that collaborations would lead to early successes in clinical trials, which would drive down costs from $450 billion USD to $300 billion USD in healthcare spending in the US. Groves et al. (2013) have found in their study that preventive actions such as early cholesterol screening and hypertension screening for patients with family histories, aspirin use for patients with risk of coronary heart disease, reduce total cost of care by over $38 billion USD. (Groves, Kayyali, Knott & Kuiken, 2013).

Canada is positioned to dominate in the data analytics landscape in the areas of healthcare and mining. The analytics market in Canada is estimated to be $1.1 billion per year. (Bourgi, Castillo & Freund, 2015). Moreover, Bourgi et. al (2015) also say that Ontario contributes to one-third of the ICT sector's gross domestic product. Furthermore, government initiatives in taking advantage of the data initiatives have been increasing, with steps taken by the current government in launching national strategies for digitizing Canada. Along with the growth in data analytics, employment opportunities have also increased in the sector. Information and Communication Technology Council (ICTC) estimates that about 33,600 data analysts were employed in Canada in 2016 who contributed about $5 billion to the Canadian GDP. (Bourgi, Castillo & Freund, 2015).

However, the healthcare sector has lagged behind in using data and data analytics to help make medical decisions. This is partly due to the traditional way of thought of medicine practitioners. The physicians would rather base their decisions on their own knowledge and judgement compared to data driven solutions. A second reason for slow health data analytics is the lack of availability of data. It is difficult to share data among organizations and hospitals due to barriers relating to privacy concerns. Lengthy and time-consuming documents have to be drawn up before any sharing of information can take place. These documents also require to be approved, which further constrain time and thus hamper efficiency for hospitals.

In order to sustain the momentum of the current big data revolution and to reap the benefits of data analytics, stakeholders such as the government, pharmaceutical companies and tech innovation companies have to remain committed. Data governance should be established to ensure that privacy of patients is retained, and information is available to be shared securely. The definitions of healthcare data demographics must be standardized so that interpretation is uniform across data pools and new research. Lastly, it has to be ensured that the load of data management and analysis is shared by all participating entities to capture the full value of data analytics.

3 The Importance of Economic Evaluation in Big Data in Healthcare

Big data provides a pathway for real time evaluation and analysis of data. This requires the companies to manage the data it collects by integration and coordination. Cloud sharing, B2B data exchange enable data sharing with vendors, outsourcing, etc. Ultra-messaging and data masking enable distribution, management and security of sensitive data. (Informatica, 2012).

In order to glean the value of data and put it to good use, it is important that the information from stored data pools is linked to provide solutions and create new knowledge. An economic evaluation of big data will help understand the value in terms of dollars, which allows a cost-benefit comparison.

Moreover, to enumerate the value of big data, it is important to evaluate the usefulness of big data. A thorough economic evaluation of the state of big data in health care would enable us to understand the value of data analytics in the healthcare system, as well as provide some perspective into weather the costs of implementing and undertaking the numerous digitization strategies in healthcare systems as identified in section 2.1, would be worthwhile. It is vital to know whether standardizing and structuring the huge volume of unstructured data is valuable to the healthcare system or not.

3.1 What is an Economic Evaluation?

Economic evaluation refers to the systematic comparison between the input (costs) and the outputs (benefits, outcomes, results) for a project. Analysing the cost and benefits would give some insight into whether a project would be feasible to pursue. A systematic evaluation specifies the details about the courses of action, the outcomes and the exact perspective for the results while a comparative evaluation examines alternatives to identify relative efficiency of actions. A complete economic evaluation examines all the relevant costs and results of actions so that trade-offs can be examined.

An economic evaluation is a necessary step in understanding where big data, in its current form and shape, stands in the healthcare industry. As mentioned previously, healthcare in Canada, and especially in Alberta, has made important strides in the inclusion of digital recording and electronic patient records. All of the advancements made in digitizing the healthcare system has resulted in a huge volume of data and data repositories. While data is available, much of it is not in a usable form. Unstructured data constitutes a large part of the collection of patient records, transactions and billing of hospital facilities and services. In order to determine the usability of the data available and whether the benefits of implementing data analytics in healthcare is higher than the costs, it is important to conduct a systematic evaluation of big data in healthcare.

The tools used in economic evaluation vary based on context and detail. There are three main forms of economic evaluation, these are as follows:

- Cost-Benefit Analysis (CBA)
- Cost Effectiveness Analysis (CEA)

- Cost Utility Analysis (CUA)

3.1.1 Cost-Benefit Analysis (CBA)

This is the most common form of analysis. It is usually done in monetary terms. According to welfare economics, a net excess of monetary benefits compared to costs represents a net welfare gain to society. A cost benefit analysis determines whether the intervention or project provides a net welfare gain to society and how the net gain is different and better from an alternative intervention or project. Hence, CBA analyzes both allocative and productive efficiency. (Palmer, S., Byford, S. & Raftery, J. 1999).

Cost Benefit Ratio is calculated as follows:

$$BCR = \frac{Benefits_{Netdicounted}}{Costs_{Netdicounted}}$$

Source: Leung (2016); Health Economic Evaluation: A Primer for Healthcare Professionals

3.1.2 Cost Effectiveness Analysis (CEA)

This method compares the costs of achieving a particular nonmonetary objective, for example measuring outcomes in natural cases like lives saved, heart attacks prevented, natural and caesarean births, etc. Outputs are measured in the same unit across alternatives and results are expressed as a ratio of the cost to outcome. As costs and benefits are measured in different units, a CEA ratio provides a comparable measure to assess the effectiveness of the project. However, this ratio does not provide a measure of relative efficiencies between two intervention or project. (Palmer, S., Byford, S. & Raftery, J. 1999).

3.1.2.1 *Cost Effectiveness Ratio (CER)*

Cost effectiveness ratio is calculated using the following formula:

$$CER = \frac{CostAlt\_1 - CostAlt\_2}{EffectAlt\_1 - EffectAlt\_2}$$

Source: Leung (2016); Health Economic Evaluation: A Primer for Healthcare Professionals

In the above formula, CostAlt_1 and CostAlt_2 are net discounted costs of two alternate projects 1 and 2 respectively while EffectAlt_1 and EffectAlt_2 are net discounted effects of project 1 and project 2 respectively. Leung (2016) reports that using CBA to compare outcomes in terms of dollars will enable the president to decide and allocate budget between healthcare, welfare, defense and infrastructure. At the provincial levels, CUA can guide Public Health Director to refer resources into different disease preventing initiatives using population health outcomes while CEA measures justify the choice of the best glaucoma screening program.

3.1.3 Cost Utility Analysis (CUA)

This form of evaluation takes preferences and time factor into account. For example, impacts of different health programs can be compared with a 'quality of life' measure also known as "Quality-Adjusted Life-Year" or QALY. It is measured as follows: Quality-Adjusted Life-Year (QALY): Weight of a year of life which is of low quality minus weight of a year of life of higher quality.

By analyzing the patterns and trends in healthcare, big data can potentially improve the opportunities in healthcare services, while reducing costs and making healthcare provision more efficient and accessible. Raghupathi and Raghupathi (2014) report that analyzing big data can lead to higher quality of care at lower costs. Raghupathi et al. (2014) also say that by analysing patient characteristics and profiling, health provider behavior can be influenced. Patients receive better healthcare via preventive care, lifestyle changes and disease profiling.

Cost minimization can be achieved by publication of medical records and options, so as to enable patients to identify and choose the procedures and doctors; thereby reducing the cost of having patients shift from one doctor to another as well as saving time. Currently, the waiting time for patients to receive a call after being referred to a doctor is very long, with emergency and surgery wait times even longer. A study on Canada's healthcare performance by Bacchus, Hasan and Timmermans (2017) found that patients had to wait an average of four months or more for an appointment with a specialist, and four months for an elective surgery. (Bacchus, Hasan and Timmermans, 2017). This waiting period causes patients to undergo unnecessary and prolonged suffering. The 19th edition of the wait times study by Esmail (2009) reports 49.2 to 70.8 percent of around 2.8 million people in waiting for a certain procedure or meeting a physician were affected by anxiety, worry and stress. (Esmail, 2009).

3.2 The Scarcity of Economic Evaluation in Big Data

A scoping study conducted by Losev (2017) across MedLine, Scopus, EconLit, and PubMed databases, using keywords such as 'economics,' 'economic evaluation,' or, 'cost-benefit estimation,' or, 'analysis,' or, 'evaluation,' or, 'projection.' Combining the prior terms with 'big data,' and 'health care,' or, 'medicine,' or, 'health services,' or, 'hospital,' resulted in only seven peer-reviewed journals and 2 grey literature. This shows the lack of research that has been conducted in evaluating big data in healthcare. Einav and Levin (2014) discuss the use of data in economic empirical research and point out how the availability or unavailability of data may change how an economic study is done. Chen, Guzauskas, Gu, Wang, Furnback, Xie, Dong and Garrison (2016) discuss the role of big data in health economics and outcomes research studies (HEOR). Chen et al. (2016) say that big data analytics can make the health economic studies more specific by employing "patient similarity analysis." Precision medicine will also gain momentum with big data analysis and economic evaluation. Precision HEOR would bring the benefit of better assessment of economic value of treatment. Phillips, Douglas, Trosman & Marshall (2017) conducted research on economic evaluation of digital biomarker technologies and found that only a few digital medicine technologies have been evaluated regarding costs or outcomes. They point out that there is a lack of effective evidence in digital medicine, which makes it harder to conduct economic evaluations. Phillips et al. (2017) also suggest the need for standardized subheadings in PubMed, which would allow for easier and efficient searches for relevant research. Given the lack of literature and evidence-based studies in economic evaluation in big data in healthcare, this paper adds to the literature and provides a case study to illustrate the lack of structured data availability for research.

3.3 The Challenges of Conducting Economic Evaluation using Big Data

Gartner, an IT research firm, estimated that total software and IT services spending on big data analytics was above $28 billion USD worldwide in 2012. Business performance improves significantly when data analytics is used. Even then, many companies have not yet adopted this evidence-based decision-making practice. The reasons, as cited by Ross, Beath and Quaadgas (2013) are the need for redefined work processes, re-establishing business rules and the cleaning of data; this is a big 'cultural shift' for most companies. Another reason why companies cannot reap the benefits from big data is the different sources of data available, which result in different interpretations. It is important that metrics come from a single source and is cleaned and structured so as the data analytics paint a clear picture of the company's performance. (Ross, Beath & Quaadgas, 2013).

Hanna (2018) reports of a partnership between Kraft Precision Medicine Accelerator, the Harvard Business School and MIT. This was founded with a $20 million USD gift, with the intention of eliminating bottlenecks and inefficiencies in precision medicine. The accelerator co-chair claims business analysis approaches will help speed up the processes and reduce inefficiencies in the development of precision medicine. (Hanna, 2018).

Pharmaceutical companies may identify patients for clinical trials if given access to big data. Pharmaceutical firms have been gathering vast amounts of data regarding patient information in order to perform research activities, as well as supply chain monitoring. Informatica (2012) in its report claims that roughly 70 percent of a data project involves the integration, cleaning and transformation of data, before analyzing it.

However, analysing and utilizing data is difficult for pharmaceutical companies since they have to balance the costs of extracting data as well as comply with ethical concerns. Collecting enough patients for clinical trials and assembling the ethics report requires time and money. Moreover, monitoring the supply chain and research and development in order to keep the products safe from theft and counterfeit require extra time, effort and expense. On the other hand, as the world evolves, and the spread of infectious diseases increase, pharmaceutical companies must be at par in order to optimize the supply and demand of medicines. (Informatica, 2012).

Patients are becoming more aware of available treatments and spend a substantial amount of time researching available options. When it comes to participating in clinical trials, the patients are also evaluating the value of their time and money and opportunity costs of participating in clinical trials. For pharmaceutical industries to succeed in advancing research and development, real-time industry information needs to be combined with historic data for greater understanding of the demands and needs for enhanced patient outcomes. Analysing real time data provides insights into treatment pattern, clinical outcomes, etc. in a real-world environment.

Copping, R. and Li, M. (2016) observe a different aspect regarding the barrier to utilizing big data in pharmaceuticals: the need for upgrading skill sets. They say that the current challenge is the inability for pharmaceuticals to handle big data due to a lack of skills for the clinicians and coders in extracting clinical data from charts, physician's notes, scans, images and pathology reports. The ICD-11 coding system will come into play in this regard as it includes a vast amount of codes for new diseases, disorders and a new and easy structure for coders to navigate through the system. Another challenge in conducting economic evaluation using big data is finding a way to utilize the vast amount of unstructured, missing information in an effective manner. Hence, better and analytical methodologies and tools are required. (Li & Copping, 2016).

A structured economic evaluation of the data utilization in healthcare would pave the way for improved efficiency, both in terms of the service delivery, reduction in wait times, better communication and allocation of resources and an enhanced range of choices available to patients in the future. This paper uses the ICD-9 to ICD-10 transition study to illustrate the extent to which data is being utilized to predict and provide health services. The case study highlights the gaps in big data in terms of structure and availability of data for research due to privacy concerns, and the challenges faced in trying to conduct the economic evaluation on the ICD-9 to ICD-10 transition. Literature review on the cost benefit analysis of the ICD-9 to ICD-10 transition resulted in only two reports from the years 2001 and 2004 which provided guidelines for cost-benefit estimation of transitioning to ICD-10. These are the RAND report, *The Costs and Benefits of Moving to the ICD-10 Code Sets (2004),* from the United States and a business case from the Calgary Health Region (CHR), *ICD-10-CA/CCI Business Case (2001)*. The only cost projection study for ICD-9 to ICD-10 transition, in the Canadian context, is the business case which was completed sixteen years ago. This shows the lack of literature and data needed for cost projections. Currently, there are no national cost projections for the newest ICD-11. It is imperative to know the cost implications of ICD-11 as it is presently being reviewed by the World Health Organization for an expected implementation date in 2023. In order to understand the economic impacts of the ICD-10 to ICD-11 transition, the case study on the ICD-9 to ICD-10 transition is very timely and would provide the required cost and benefit estimation frameworks. The results of the case study show that data is currently only utilized for a limited amount of research by academicians and is not applied by clinical practitioners in the provision of healthcare. With advancements in machine learning and the huge availability of health data, it is vital that the healthcare industry exploits the advantages big data can provide.

4 The Economic Evaluation of ICD-9 to ICD-10 transition in Calgary: A Case Study

4.1 What is ICD and ICD-9 to ICD-10 coding system transition?

ICD assigns codes to medical diagnosis that are part of an internationally recognized registry of diagnostic codes set out by the World Health Organization (WHO). The current ICD-10 is a health data management computer system that includes hundreds of diagnostic codes. These codes are used for research and analytics by the health industry, as well as to pay healthcare providers. The Deputy Minister of Health for provincial, federal and territorial governments decided to adopt the International Statistical Classification of Disease and related Health Problems (ICD)-10 in 1995. The date of implementation for Alberta was set to be on April 1, 2002. ICD-9, which was published by the World Health Organization in 1977, was deemed inadequate in reflecting the newest advances in medical care, such as genetic research, international data sharing and biomedical informatics while ICD-10 provided greater detail for coding inpatient hospital procedures and diagnoses. ICD-10 increased the original volume of approximately 18,000 codes in ICD-9 coding system to over 140,000 codes; for the United States, the transition meant a five-fold increase to 68000 codes. (Deloitte, 2010; Health Affairs, 2014). By 1998, approximately 38 countries had adopted the ICD-10 coding standards.

ICD-10 included an enlarged coding frame, hierarchic and logical presentation of codes using alphanumeric characters and the use of combination codes, which changed and improved the structure of ICD-10 compared to ICD-9.

Due to the collection of ambulatory care data, the scope for implementing ICD-10 in Alberta was larger than other provinces. The Calgary Health Region, in its business case study, estimated the total cost of implementing ICD-10 to be approximately $4,551,376 CAD. (Calgary Health Region, 2001). It is estimated that the cost for implementing ICD-10 in the United States for a small practice was expected to be $226,000 USD and above $8 million USD for a large practice. In addition, U.S. had to implement ICD-10 for the health insurance companies as well, which was around $2-$3 billion USD for total system-wide cost. (Health Affairs, 2014).

ICD codes are used by a large facility of people ranging from clinical practitioners, physicians, nurses, researchers and for billing systems. A substantial amount of information from these coded data is used by Canadian Institutes of Health Information and the World Health Organization to organize and allocate budget and resources to hospitals. As Alberta moves towards activity-based payments, the use of ICD coding information is expected to increase. Moreover, a number of research publications uses ICD codes. As explained later, the interest of more numerical and empirical research has given rise to a surge of demand for ICD coded information, which can be used by specialized statistical packages to perform statistical analysis.

4.2 Why do we need to estimate costs and benefits of ICD transition?

The ICD-10 is a health data management computer system that includes hundreds of diagnostic codes used for research and analytics by the health industry, as well as for funding allocation in the healthcare sector. The only cost projection study which was available on transitioning from ICD-9 to ICD-10, is a business case which was completed sixteen years ago. This shows the lack of literature and data needed for cost projections for ICD-11. As mentioned before, at the moment there are no national cost projections for the newest ICD-11 coding system which is imperative since the countries implementing ICD-11 want to know understand the feasibility of ICD-11 implementation.

The paper uses the ICD-9 to ICD-10 health indicator coding system transition as a case study to illustrate the lack of data availability in the healthcare sector as well as the challenges faced in conducting an economic evaluation. At the onset, an investigative study is conducted on the economic evaluation of the transition from ICD-9 to ICD-10, using a cost-benefit analysis.

A cost-benefit analysis is chosen so that the costs and benefits of the transition may be calculated in dollar terms. Analysing the economic impacts of the transition in monetary terms provides the best basis for comparison for both within the project as well as with other projects which will be implemented in the future, such as the ICD-11 coding system.

The challenges in conducting the economic evaluation include a lack of data availability for the specific timeline of the ICD transition, as well as the difficulty in constructing benefit scenarios in the Canadian context. As Canada moves towards implementing the ICD-11 coding practice, the ICD-9 to ICD-10 transition case study is the best example to illustrate the big data utilization state as it provides a foundation for economic evaluation of the ICD-10 to ICD-11 transition. Hospitals implementing the ICD-11 system are interested in understanding whether or not the benefits of implementing the system would exceed the costs.

4.3 Literature Review for Cost and Benefit Estimation Methodology

Two reports were reviewed for cost projections and methods of estimating the cost and benefits of implementing ICD-10 from the years 2001-2003. The RAND report, *The Costs and Benefits of Moving to the ICD-10 Code Sets (2004),* from the United States was compared with a business case from the Calgary Health Region (CHR), *ICD-10-CA/CCI Business Case (2001)*. Both reports project costs and benefits for implementing ICD-10 and show the main benefits of newer ICD code sets. In the RAND report, the benefit from ICD-10 coding system were improved productivity and improved processes which led to more comprehensive information with less billing errors. The RAND report also calculated benefits accumulating from a better medical data management system under ICD-10. Additionally, the report noted that more refined codes of ICD-10 enabled more detailed data entry, expanding research capabilities for the health care industry.

Among the studies identified, the RAND technical report (Libicki & Brahmakulam, 2004), "The Costs and Benefits of Moving to the ICD-10 Code Sets" is the only study that provided a comprehensive methodology for cost-benefit estimation in dollar terms. A cost-benefit estimation (see table 2 and table 3, Appendix C) produces dollar estimations that can help decision-makers in assessing the value of moving from ICD-9 to ICD-10 in the U.S. in contrast to its cost. The RAND report calculated that the dollar value of the benefits of moving to new code sets is greater than the costs of the transition. This report presented a theoretical framework that has been revised and adapted to form the costs-benefits estimation methodology for ICD-9 to ICD-10 transition in Canada in this thesis.

The benefits outlined in the CHR business case report were not calculated and it was expected that there would be a general benefit from increased scope of codes, improved specificity, international compatibility, effective structure and presentation of the ICD-10 coding system.

The two reports reviewed outline the necessary information to begin a cost projection for ICD-11. The cost estimation formulae are based primarily on the robust RAND report, as the CHR business case lacks details on the methodology used to estimate project costs. The data for cost estimation is received from Health Information Managers and Directors at Alberta Health and the report on ICD-9 to ICD-10 transition by Calgary Health Region. In person interviews with coding managers from Health Information Management (HIM) and Canadian Health Information Management Association (CHIMA) were conducted in order to gain more insight into the process of ICD-10 implementation in Calgary Health Region and to inform on the processes involved in the economic evaluation framework of the ICD-9 to ICD-10 transition. Moreover, skype interviews were conducted with Saskatchewan, Manitoba and British Columbia for another project estimating costs of ICD-10 to ICD-11, to understand the differences in estimation methodology which need to be implemented (see table 1). Although many different coding managers were contacted for data on enumerating the benefits, it has been very difficult to avail such information. The interviews and focus group discussions form the basis of the study design and formulae to calculate the costs and benefits of the ICD-9 to ICD-10 transition.

4.4 Methodology of Economic Evaluation of ICD transition

The economic evaluation case study in this paper borrows on the methods developed by the RAND study (which was conducted in the U.S.), on ICD-9 to ICD-10 transition. The methods for benefit and cost estimations described in this paper have been in part based on the formulae developed by RAND and changed according to the Canadian universal healthcare system.

The benefit calculations are quite different due to the fact that RAND uses billing systems and insurance, while Canada does not implement an insurance system as the basic healthcare services are provided by the government free of cost to the citizens of Canada. The cost calculations are similar but have been differentiated in accordance to a standardized requirement of different provinces. Detailed descriptions of differences are provided in table 1.

4.4.1 Cost Estimation Methodology

The cost estimation methodology developed in this paper is an improvement of the cost calculation conducted in the RAND report and the CHR business case. The calculation includes various new aspects and variables which make it more pertinent and recent.

4.4.1.1 *Cost Estimation Framework by Calgary Health Region*

Prior to the implementation of ICD-10 in Canada in 2002, the Calgary Health Region conducted a cost estimation analysis for evaluating the implementation of ICD-10. Figure 2 illustrates the framework by CHR. While the benefits were not enumerated, the team made a rigorous estimation of the costs that was expected to be incurred.

Alberta Health and Wellness used three survey questionnaires to collect the base data. The questionnaires addressed impact to health authorities' custom computer applications, impact to health authorities when implementing new versions of packaged software that incorporates ICD-10-CA/CCI and information on software vendors with regards to change in license and annual maintenance fees for software for Inpatient, Public Health, Continuing Care and Ambulatory Care packages. Project management and administration costs were calculated at a rate of 15%. AH & W assumed that there would be no changes in morbidity data flow for simplicity. (Calgary Health Region, 2001). Detailed explanations of the cost estimation, including the dollar amounts for each category, is presented in tables 5, 6 & 7 in Appendix D.

4.4.1.2 *Cost Estimation Framework by RAND*

The variables included in the cost estimation in the RAND report were costs of training, productivity losses and system change costs. Table 2 in Appendix C gives detailed variables with corresponding dollar amounts. Costs were estimated by identifying the relevant coders, physicians and code users, along with the time required to train them. Productivity loss estimates included initial and long-term productivity losses associated with ICD-9 to ICD-10 transition. Costs for reprogramming the systems and software updates were calculated for a sample and extrapolated for the population. It was estimated that the one-time cost of conversion would be between $425 million USD to $1150 million USD, while costs in lost productivity would accumulate to between $5 million USD to $40 million USD per year. (Libicki & Brahmakulam, 2004).

4.4.1.3 *Validation of the CHR and RAND Frameworks for Cost Estimation*

In order to validate the economic evaluation framework used, a comparison of the cost estimation methodology used in the RAND report and the one used in CHR ICD-10-CA/CCI Business Case, was conducted. Detailed side-by-side comparisons were done to check the methods of cost estimation in both reports, as well as variables and data used to conduct the projections (see Table 4, Appendix C). The comparison showed that the framework and methodology for economic evaluation of ICD transition adopted in the RAND report could be modified and adopted to be used in a Canadian context. (Losev, 2017).

The last ICD-9 to ICD-10 transition study was done in 2001, prior to the implementation of ICD-10. The cost calculations in the CHR business case is provided in Appendix D. The cost estimations were done based on the data available at the time. The calculations were based on answers to questionnaires addressing the expected training and project management costs, productivity losses and software application costs. Detailed information of the exact method of calculations are not provided in the tables or the report by CHR. The methods and data used for the cost estimation, such as data on short-term and long-term productivity impacts, the bids by software vendors for incorporating system changes, were not available. Appendix A- the detailed project plan was not provided, and the reference list of the staff interviewed, stakeholders, and other information sources were not produced to us. The comparative tables show that CHR did not have as much data for cost estimations though it presented some cost amounts in the business case, without any background information of the data collection and cost estimates. Although the CHR report stated that more information was available upon request, it has been difficult to collect this information due to a lack of proper storage, filing as well as privacy concerns regarding sensitive data.

This made it very difficult to try and conduct an analysis which could evaluate the actual costs that occurred during the implementation phase. It has also been impossible to collect information on the productivity losses and training hours required after implementation. Interviews with project and coding managers conducted by us in the past few months point to sufficient amount of time lost in productivity. This was due to the time lost in coding during which coders had to struggle to retrieve accurate codes and fix the broken code pathways which were associated with the then ICD-10 coding system. Performance appraisal reports were retrieved from the years 2016 and 2017, but these are far too recent to accurately predict the productivity loss which occurred during and after ICD-10 implementation in 2002 and 2003.

4.4.1.4 *Cost Estimation Method for ICD Transition in Canada*

Based on the CHR report and the RAND report, this paper has improved upon the cost estimation methodology, by adding newer variables which are used in the current coding system. Suggestions from the focus group interviews have also been incorporated into the formulae. It must be noted that the cost framework contains both the costs during implementation, as well as indirect costs. Indirect costs refer to the costs all the other users of the data will incur when ICD-10 was implemented, the productivity losses which could extend well beyond the implementation phase as well as the opportunity costs of implementing and allocating healthcare budget. The opportunity costs, however, can only be calculated accurately after the costing and budgeting of ICD-10 implementation has been completed. Hence, this framework does not explicitly calculate the opportunity cost in terms of healthcare budgets, of moving to ICD-10. Costs of ICD transition had been classified into three broad categories: costs of re-training, productivity losses and cost of system changes (see Figure 1). Using data collected from the interviews and focus group meetings, total cost of moving from ICD-9 to ICD-10 coding system has been formulated. The main components of the cost projection framework are further broken down and calculated as follows:

Costs of re-training has been divided into equivalent full-time coders, coding support unit, physicians/clinicians and end users of the coded data. Other users of the data include patient Quality and Safety, Infection Control, Business and Data Analysts and Project Managers. Cost of re-training is calculated based on number of people who need to be re-trained, time it requires to re-train them, and hourly wage in each category. Here, the people retrained are those who record hospital data using ICD-10 codes. The framework does not include billing or physician claims coding data personnel.

Formula 1: Simple formula for cost estimation calculation for cost of re-training

Total Cost (CAD $)	Estimation Method
$RT_{\text{equivalent full-time coders}}$	Wages per hour for equivalent full-time coders X number of equivalent full-time coders
$RT_{\text{physicians/clinicians}}$	Wages per hour for physicians/clinicians X number of physicians/clinicians
$RT_{\text{other users (coding support unit, Patient Quality and Safety, Infection Control, Business and Data Analysts and Project Managers)}}$	Wages per hour for other users X number of other users
Total Cost of re-training	$RT_{\text{equivalent full-time coders}} + RT_{\text{physicians/clinicians}} + RT_{\text{other users}}$

X denotes the multiplication sign.

Here, RT denotes the cost of re-training coders. This calculation can be done for the time of implementation of ICD-10 and ICD-11, to depict true costs, adjusted with inflation. It is difficult to collect data for the ICD-9 implementation timeline at this point in time since ICD-9 was implemented a long time back. The total cost of re-training is calculated by adding all the components outlined in the above formula.

Productivity losses: The productivity of coders, code users and technology support units are affected during the transition phase. Short-term and long-term productivity losses is estimated, considering break-in period of three to six months. (Break-in period refers to the transition period after ICD-10 is implemented, including intensive training and learning period for all users). Data on the number of hospital discharges per month and extra coder hours required during the break-in period, are included in the estimation by Calgary Health Region of total cost due to productivity losses. Implementation of ICD-10 may coincide with new fiscal year, which would result in the coders to work additional time to take care of backlog of data. The productivity loss is estimated using the calculation depicted in the following grid:

Formula 2: Simple formula for cost estimation calculation of productivity loss during system transition

Total Cost (CAD $)	Estimation Method
$M_{\text{coding ICD9 per coder}}$	Minutes required to code a chart using ICD 9 codes
$M_{\text{coding ICD10 per coder}}$	Minutes required to code a chart using ICD 10 codes
M_{increase}	$M_{\text{coding ICD10 per coder}} - M_{\text{coding ICD9 per coder}}$
Total amount increases in remuneration	($M_{\text{coding ICD10 per coder}}$ X wages/ hour) - ($M_{\text{coding ICD9 per coder}}$ X wages/hour)

X denotes the multiplication sign.

Here, M denotes the minutes required to code a chart by each coder.

This calculation can be done using one-year coding data collected for one coder under each coding system. Performance appraisal reports are a good source for this data. The M_{increase}

measure, if positive, will indicate the higher inefficiency of coding using ICD-10 compared to ICD-9. In addition, we can find the dollar amount related to productivity loss by multiplying the wages of coders. If the total amount increase in remuneration is positive, we can conclude that the ICD-10 coding system is less efficient and has resulted in a cost for the hospital management by increasing the amount of money paid in terms of total wages given to coders.

For example, it is expected that the coding using ICD-10 would be slow in the beginning allowing for an adjustment to the newer code definitions and classifications. However, gradually, this time difference will be closed as coders become more efficient and familiar with the ICD-10 system. The short term (3 months) and long-term (6 months) productivity break-in periods will allow for accurate calculation of productivity losses.

On the other hand, this productivity loss may be offset in some amount by the benefits which may be accrued from higher productivity of workers (Formula 2 in benefit estimation methodology) due to the better structure and enhanced software available with the ICD-10 coding system.

Other code users and technology support units will require extra time to perform their daily tasks during the transition phase, gradually becoming used to the new system. Orientation and training of new hires would also cost the province time. Moreover, time is required to convert data which takes away time from other projects. Reinforcement training after implementation of the ICD-10 system is an added time required which the coders have had to invest, which in turn resulted in lost work time and thus productivity loss. This extra lost time is added to the per coder time estimate using appropriate adjustments.

Costs of System Changes is calculated using bids by software vendors. In order to accumulate the cost of software application changes for the ICD business case report by Calgary Health Region, questionnaires were given to vendors to bid for contracts. For this study, system change costs are expected to include estimations of time to work, total cost of designing and adapting/updating software, data conversion cost, data ware house, data dictionary and decision support systems, licensing fees and cost of adapting the new codes and reprogramming logic.

Formula 3: Simple formula for cost estimation calculation for cost of system changes

Total Cost (CAD $)	**Estimation Method**
$W_{\text{data conversion}}$	Cost per unit of data conversion
$W_{\text{hours of work by vendor}}$	Number of hours required to complete per unit of data conversion
$W_{\text{wages per hour}}$	Wages per hour for software vendor
Total cost of data conversion	$W_{\text{data conversion}} \times W_{\text{hours of work by vendor}} \times W_{\text{wages per hour}}$
$W_{\text{other costs}}$	cost of designing and adapting/updating software, data warehouse, data dictionary and decision support systems, licensing fees and cost of adapting the new codes and reprogramming logic.
Total Cost of system change	Total cost of data conversion + $W_{\text{other costs}}$

X denotes the multiplication sign.

Here, W denotes the costs of variables related to system changes.

In addition, the cost estimation involves accommodation expenses, travel & facilities expenses, data quality checks, hours of technical support needed to fix broken pathway (validation and correction), quality assurance reports, clinical quality of documentation, coding quality support teams, orientation of new hires. Other secondary users such as Manitoba Center for Health Policy, Surgical Information Management System, HIS, ADT, Grouper Software users and institutions offering online courses will have been affected by the system change and will have incurred costs in terms of both software changes and retraining of workers. For ICD-11 transition cost estimation, IMO and coding automation would be added variables. Moreover, the cost estimations will vary in different provinces due to the prevalence of different needs and regulations

From discussions and interviews with Health Information Managers and other data coding directors and managers in Alberta, Manitoba, Saskatchewan and British Columbia, the differences in cost estimations for different provinces is detailed in table 1.

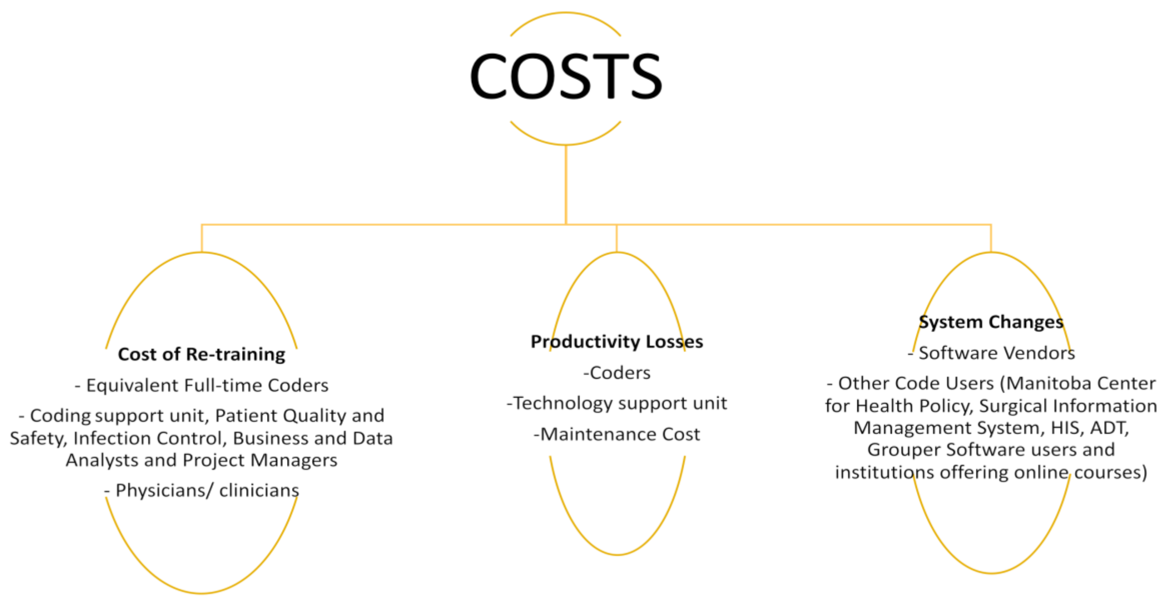

Figure 1: Cost Estimation Framework

4.4.2 Benefit Estimation Methodology

4.4.2.1 *Benefit Estimation Framework by RAND*

In the RAND report by Libicki & Brahmakulam (2004) the benefits of moving to ICD-10 are the following:

a. More-accurate payments for new procedures
b. Fewer miscoded, rejected, and improper reimbursement claims
c. Better understanding of the value of new procedures
d. Improved disease management
e. Better understanding of health care outcomes (considered, but not estimated)

In RAND report, benefits were estimated using benefit scenarios. A detailed summary of the benefit estimation calculation as estimated in the RAND report is provided in Appendix C.

4.4.2.2 *Benefit Estimation Framework by Calgary Health Region*

The CHR business case highlighted that benefits from the ICD-10 implementation is expected in terms of a more comprehensive coding system which includes many new procedure codes, terminology and clinical concepts; increase in the specificity of codes which would result in better quality of data for epidemiological research as well as higher sensitivity of classification; a Canadian version of ICD-10 coding system which ensures that the coding is relevant to the Canadian healthcare system; and international compatibility of codes which will result in international comparability of data. However, CHR did not enumerate the benefits into dollar terms in the report. (Calgary Health Region, 2001).

4.4.2.3 *Validation of the CHR and RAND Framework for Benefit Estimation*

The RAND report estimated benefits in terms of benefit scenarios. While the RAND report provided a comprehensive background for the methodology for estimating the benefits, it is important to note that the operational systems of the billing process in Calgary is very different compared to the U.S. These benefit scenarios which employed the billing and payer methods, such as those resulting from insurance payments, value of new procedures and improved disease management, are difficult to construct in the context of the Canadian healthcare system. This is because, the Canadian healthcare system is a single payer system, which does not use ICD-10 codes for billing purposes. It still uses ICD-9 codes for some of its hospital bills and payments.

Moreover, CHR, in its ICD-9 to ICD-10 transition report (2001), did not mention less fraud as a benefit of newer codes. RAND mentioned less fraudulent claims as a benefit. It is quite plausible that the issues of miscoded and reimbursement of claims does apply even in Canada, however, since ICD-10 has not been implemented and insurance or third payer parties is not widely used in Canada, this benefit cannot be calculated. Therefore, it would be valuable to find some information about the rate that this may occur in Alberta or Canada to calculate if new codes could reduce over-billing by improper coding. Unfortunately, due to a lack of information, economic evaluations or investigations of any sort, it is not possible to track mis-coding, improper billing or fraudulent claims.

Moreover, physicians and hospital administration does not use ICD codes for diagnoses or for hospital management, thus the benefit scenarios arising from better disease management cannot be constructed. This thesis has however constructed a few of the benefit scenarios and has given detailed formulae for benefit calculation in the Canadian context (see Appendix E). Appendix E also provides a discussion of the challenges due to which the benefit scenarios cannot be applied in practice.

Furthermore, the Canadian healthcare management varies by each province and territory in its processes and healthcare funding allocation. Since the U.S has a large variance across their vast multi-payer health system, the RAND report offered wide confidence ranges to accommodate for those uncertainties and acknowledged further research opportunities, while also accounting for their small sample sizes. The CHR business case, however, did not account for these variances across provinces.

4.4.2.4 *Benefit Estimation Method for ICD Transition in Canada*

Although the benefit estimation scenarios do not apply in the case of Canada, one benefit calculation formula, which applies to the Canadian context is illustrated below:

ICD-10 has an enlarged coding frame and uses the grouping software which makes coding more efficient and faster, resulting in a reduction in time required to code. This would result in higher productivity of workers over time. The calculation for this benefit is done using the following formula:

Formula 1: Simple formula for benefit estimation calculation for reduced time required to code after ICD-10 implementation

Total Benefit (CAD $)	**Estimation Method**
$M_{\text{coding ICD9 per coder}}$	Minutes required to code a chart using ICD 9 codes
$M_{\text{coding ICD10 per coder}}$	Minutes required to code a chart using ICD 10 codes
Minutes saved	$M_{\text{coding ICD9 per coder}} - M_{\text{coding ICD10 per coder}}$
Total amount saved in remuneration	$(M_{\text{coding ICD9 per coder}}$ X wages/ hour) - $(M_{\text{coding ICD10 per coder}}$ X wages/hour)

X denotes the multiplication sign.

Here, M denotes the minutes required to code a chart by each coder.

This calculation can be done using one-year coding data collected for one coder under each coding system. Performance appraisal reports are a good source for this data. If the total amount saved in remuneration is positive, we can conclude that the ICD-10 coding system is more efficient and has resulted in benefitting the hospital management by saving money in terms of total wages given to coders. The minutes saved measure, if positive, will also indicate the higher efficiency of coding using ICD-10 compared to ICD-9.

However, the estimation of productivity gains from ICD-10 coding, a stark contrast to the productivity loss estimation, may be a scale which could easily tip in favour of productivity loss instead. This productivity gain may be realised only when a considerable number of years has passed after ICD-10 implementation.

Other benefits arising from ICD-10 implementation include higher number of classifications and specificity and the inclusion of newest procedures. This results in increased coding efficiency and reduced time required to code, which can be measured using formula 1 above.

4.5 The Data Quantity and Quality Approach for Benefit Estimation

The other benefit from ICD-10 implementation is the enhanced data quality and increase in quantity of research papers, arising due to the use of combination codes, better system structure and addition of newer procedures and codes. Since the data for the variables for benefit scenarios in Appendix E is unavailable for the years from 2000 to 2003, which is when the ICD-9 to ICD-10 transition took place, as well as due to the difficulty in constructing the benefit scenarios in the Canadian context, the benefit from increased data quantity and quality is measured. This has been done by observing the number of peer reviewed research published using ICD-10 code data compared to ICD-9 between the years January 2000 to May 2018. I have gone through articles found on PubMed and used search terms related to ICD-9 to ICD-10 transition, which included the terms 'coding system,' 'ICD-9,' 'ICD-10,' 'data quality,' 'data quantity,' and 'ICD transition,'. Searching with all the search terms mentioned above, we see that a higher number of journal articles have used ICD-10 codes compared to ICD-9 codes for research purpose.

However, this is a very small scale and simplistic survey. There could be various pitfalls to assuming that the rise in research report publications is due to a better structure of the ICD-10 codes. It should be noted that many of these papers sought to understand whether the transition to ICD-10 coding system brought about higher specificity and sensitivity in understanding the coding of diseases, compared to ICD-9.

This method of estimation is vague since there is no threshold or limit to the total number of publications that can be accounted for. The list is virtually exhaustive, making it very difficult to enumerate all publications that have utilized ICD-9 and ICD-10 codes. Other factors may affect the increase in economic evaluation and resulting increase in publication of research overall. These are explained as follows:

4.5.1 Increase in research funding:

There may have been an increase in research funding in the years after 2000, which has led to a rise in publications in the field of healthcare. Tamblyn, McMahon, Girard, Drake, Nadigel and Gaudreau (2016) report that in the years from 2001 to 2011, 80163 applications for grant were submitted to CIHR, which marks a threefold increase; from 327 to 1137 applications per year for health services and policy research. Annual funding for accepted research increased from \$12.6 to \$48.0 million between 2001 to 2011.

4.5.2 A heightened interest in the economic evaluation of health sector and healthcare budget:

I. More seminars and conferences may have been arranged as a response to the demand for higher dissemination of information. Seminars usually constitute of presentation by researchers, policy makers and academic personnel, which results from published reports and synthesized statistics.
II. Voters are more aware of the effect public policy changes have on their daily lives. This creates an interest in learning more about the healthcare policies and the ways in which healthcare is funded. There has been a rise in the debates for privatizing healthcare to make it more efficient. As discussed previously in this paper, Canada has high wait times for physician appointment, surgery, etc. and studies are being conducted to determine exactly how long these wait times are. Activity-based funding, cost-sharing and user fees

are upcoming arguments which the people of the country are interested to know more about.

III. Academia has become more competitive than ever before. Universities give higher value to professors who publish in high ranking journals, while the community gives more recognition to researchers publishing a higher number of reports and commenting in tabloids regarding policy options. This has resulted in the academic community to engage in publishing more research reports or op-eds which in turn may have contributed to an increase in the number of publications.

IV. The recent trend in accepted journal publication is a demand for intensive and more convincing empirical evidence to support claims and hypothesis by the authors. This trend in trying to obtain and incorporate more data has resulted in researchers turning to data-driven research, which means a heightened use of ICD coded data. This may not have been the case even in the early 90s, and hence acts as a contributing external factor in increased journal publication using ICD-10 codes.

The above factors contribute to an increase in the number of publications using ICD-10 codes compared to ICD-9. Thus, although we may be inclined to conclude that the detail and structure of codes in ICD-10 has backed a rise in better quality research judging from the higher number of published papers, it could quite well be a result of the above factors.

Other benefits which are being considered, albeit just in theory for this paper are international compatibility in codes, effective structure due to enlarged coding frame and higher number of codes resulting in better and easier coding practices with fewer errors, better understanding of health conditions and healthcare outcomes and rapid detection of emerging diseases.

4.6 Lessons learned from the ICD-9 to ICD-10 Transition Case Study

As the WHO rolls out the ICD-11 coding system, it is important to understand the milieu of the economic impacts such a system change will bring about. An economic evaluation which weighs the costs of a system change such as the ICD-9 to ICD-10 coding system transition against the benefits, will provide a platform to prepare and understand the processes the ICD-10 to ICD-11 transition will go through.

The case study illustrates the methods and formulae required to calculate the benefits and costs of the transition in monetary terms. In doing so, the study highlights the challenges and drawbacks of completing a cost benefit analysis. These impediments mainly hinge around the lack of availability of structured data, both due to the timeline of ICD-9 to ICD-10 transition, as well as due to issues surrounding privacy of data, collection and storage.

Although the costs estimated by Calgary Health Region in September 2001 had been measured using interview questionnaires and secondary data, most of it was an extrapolation of estimates. It is challenging to calculate the exact cost measures since the data for the periods from 2000 to 2003 is unavailable. Moreover, from interviews for the ICD-10 to ICD-11 transition study (which is being conducted by The Methods Hub at The University of Calgary and tries to formulate a cost benefit analysis of the ICD-11 implementation in Canada) it was found that most of the data is too sensitive to be collected. A cost benefit evaluation thus seems to be a very difficult task to accomplish. We have been asked by DIMR to provide them with information on what data it needs to collect, in order to facilitate a study like the ICD transition evaluation.

On the other hand, the HIM did not have access to the methodology and figures used by CHR, so the methodologies of the cost estimations in the CHR business case could not be checked. In the absence of tests for miscoding, over-coding, or under-coding, for example, a physician may opt for the most general ICD code when billing despite having the choice of more detailed codes to choose from. This may impede the potential benefit of ICD-10 providing a more robust information pool for medical research.

Despite the formulation of an extensive estimation technique to evaluate the transition, it has been a challenging task to employ the cost-benefit analysis. This has been due to a lack of availability of data and considerable gaps in usage of codes in practice. Interviews and discussion with experts have shown that the ICD health codes are not used for disease management or referred to by physicians to identify possible solutions for the patients. However, the coding managers have strongly recommended that the codes could be used to inform on the progress of patients and may even facilitate early discharge depending on an analysis of the charts by the physicians.

Due to the significantly different health system in Canada, which is a single-payer system compared the multi-payer system in the U.S., it has been difficult to construct the benefit scenarios outlined in the RAND report. Hence, the cost-benefit analysis applied in the RAND report is not feasible to be applied for Canada. The benefit measured in this thesis in terms of data quality, is not in dollar terms and cannot be taken as a sole measure to compare against cost of ICD-10 implementation. Hence, another lesson learned from this study is the direction of economic evaluation for the next ICD-10 to ICD-11 study, which would lean more towards a cost-effectiveness analysis, rather than a cost-benefit analysis. Given the context of the health system in Canada, a cost-effectiveness analysis is more feasible. This next transition study would estimate the costs using a refined cost estimation framework and estimate the benefits in terms of data quality, by calculating the sensitivity, specificity, positive predictive value and negative predictive value for each of the Charlson (Charlson, Pompei, Ales & MacKenzie, 1987) and the Elixhauser (Elixhauser, Steiner, Harris & Coffey, 1998) conditions.

4.7 The Way Forward

The utilization of data in healthcare has vast implications for the future. This includes the possible advancements in activity-based funding for hospitals, information management of patients, policy implications regarding privatization of healthcare, payment reform, risk adjustment, budgeting for healthcare, and using ICD codes for disease management and practice.

4.7.1 Activity based funding

With activity-based funding making strides in Alberta in the next year as reported by the Fraser Institute, the province and the country has to ask the very pertinent questions of whether the province of Alberta is prepared to adopt such methods.

Activity based funding refers to allocation of money to hospitals based on the number of procedures and treatments provided from the hospital. It is deemed to be a positive stride in ensuring higher efficiency and accountability for hospitals. Canada currently has the highest rate of waiting times among countries providing universal healthcare. The wait time for seeing a specialist is almost four months (Bacchus, Hasan and Timmermans, 2017). It is expected that a change in methodology would help the hospitals become more efficient due to enhanced monitoring and activity-based funding.

The current system of payment of funds to hospitals is via a global budget. This means that hospitals are paid a lumpsum amount, regardless of the number of patients treated. This creates a disincentive for hospitals to efficiently treat patients and mobilize resources for faster care. It also results in hospitals trying to outsource patients who need extensive or long-term care in order to reduce expenses and remain within budget.

According to a report by CIHI, activity-based care involves health system funders to gather information from clinical and financial data. This data is then categorized into clinically similar and cost homogeneous groups to map activity-based payment price. Hospitals and healthcare providers are reimbursed for services provided based on this data. (Canadian Institute for Health Information, 2013).

The proponents for activity-based funding cite increased efficiency and higher quality of healthcare provision under this system. However, the disadvantages include lower services provided by some regions if providers are unable to attain required efficiency levels for certain services and redirect resources to other more profitable health services. Moreover, this method of funding would require increased monitoring and information requirement to adequately assess funding for hospitals and services. Policies need to be enacted to ensure that information is readily accessible, and codes are recorded without glitches for smooth operation.

4.7.2 Payment Reform and Auto Coding

With the advent of data analytics in health care, important policy strides may be made towards using coded information for payment reform. Recent debates regarding cost-sharing and implementing user-fees may become more accessible and closer to reality given readily accessible and detailed data.

Data from ICD coding may also be used to facilitate budgetary concerns in greater detail. ICD-11 codes have recently been launched live and is expected to be used in hospitals by 2023. The aim of WHO in launching ICD-11 is a set of codes which will be compatible with the existing software system. The new set of codes incorporate PTSD, mental and reproductive issues at a greater extent than previous codes. This would enable government and health care practitioners to understand the extent of mental health disorders and the amount of funds which need to be allocated to this area. ICD-11 also differs in structure from its precursors; it incorporates clustering and will address concerns of mapping between different coding systems in use, quite possibly SNOWMED. SNOWMED is a system which gives multilingual and complete terminology for clinical procedures. Southern et. al (2016) report that clustering will 'enable the capture of the complete clinical picture and benefit the quality improvement efforts and patient care.' Data from the different codes of PTSD, mental health etc. could also allow potential re-allocation of government funding based on evidence.

Auto coding could be used in the coming years. Auto coding would involve usage of ICD codes and data analytics to inform on disease management. This could potentially alleviate wait times through better tertiary management of patients. There would be implications for the cost framework and estimations outlined above if auto coding is adopted, such as a reduced need for coders and thus re-training costs would be lower. However, this may result in coders being laid off, although the nature of work of coders may change from just coding to analysing the coded charts. Gibson (2012) reports that EHR and auto coding result in higher efficiency for medical billing processes. Although, he notes that auto coding does not eliminate the need for thorough review by medical coders. Computer Assisted Coding (CAC), is a software which generates auto codes using SNOMED CT (Systemized Nomenclature of Medicine- Clinical terms). Since coding software are prone to creating codes which are considered medically unlikely for third-party payers, the codes generated require to be reviewed by professional and experienced medical coder. (Gibson, 2012).

On the other hand, in the U. S., as reported by Holland (2018) auto coding and EMR can also cost the medical system a huge amount of money instead of cutting medical costs. This is because of the need for doctors to have to enter more codes in the automated system to correct mistakes. This takes time away from treating patients and may cost a doctor more than $200 USD to process one bill payment. (Holland, 2018).

If the data is used by physicians as an extra help in addition to their diagnosis; for example, in providing information regarding a particular set of drugs and its uses, side effects and percentage of people cured using a certain drug, the chances of a patient being prescribed the correct drug may increase. On one hand this would enable faster diagnosis, and on the other, patients can be made more aware of which drugs are being prescribed to them and why.

4.7.3 The move towards implementation of ICD-11

A parallel study by The Methods Hub at University of Calgary, is working on estimating the cost and benefit frameworks for ICD-11. The cost framework is given in table 1, along with the differences in estimation variables between Alberta, Manitoba, British Columbia and Saskatchewan.

It is expected that the cost estimation framework developed in section 4.4.1.4 will help policy makers to address some key questions surrounding the adoption as well as implementation of the ICD-11 classification system. These are:

I. Who will be affected by the ICD-11 transition?

II. What would be the economic impacts of the ICD-11 transition and what are the resources and support needed?
III. How much will this transition cost?
IV. What are the benefits of this transition?

Codes in ICD-11 are very different and easier to incorporate compared to ICD-10, due to the more effective structure of ICD-11 system. This is expected to bring in an easier transition for the coders. CIHI also expects ICD-11 to be more widely used due to the introduction of many new codes such as mental health, gaming disorders, etc. in ICD-11. Due to clustering in ICD-11, which makes it easier to read and connect with other systems, it may be a possibility that ICD-11 codes would bring about more machine learning and automated systems. It may also be possible to incorporate physician claims and billing for certain regions to use ICD-11 codes.

The RAND report on ICD-9 to ICD-10 transition used billing information and insurance information to calculate benefits. Although the RAND report was used as a guide, it was not possible to adapt the disease management and payment benefits for estimation of ICD-9 to ICD-10 transition in the Canadian context, since Canada has a universal healthcare system where the government pays for healthcare except for optometric and dental care. As mentioned earlier, clinicians or physicians do not use ICD codes for disease management either.

It is expected that Canada will adopt ICD-11 by 2023, while other countries are expected to follow. One of the many reasons for estimating the cost and benefits for this transition is an understanding of the economic evaluation methodology to present to the countries who are contemplating the adoption of ICD-11. It is important to understand the pros and cons of a system overhaul before implementation. An economic evaluation also provides a perspective for the government and other respective institutions involved in the process to gauge the potential benefits, possible resource allocation and plan a budget. And this is the importance and uniqueness of this thesis. This thesis contains an extended and elaborate methodological framework for such an evaluation.

The previous evaluation by Calgary Region for the estimation of costs was very simplistic. More over the business case had not tried to enumerate the benefits. This thesis provides an extensive method for the cost calculation, including the data sources which can be used.

This thesis has used the original variables of productivity and cost of training, and included more variables such as Coding Support Units, Data Sharing (such as with Employees of government organizations such as Analytics at Alberta Health & Wellness, CIHI, Federal Government Vital Statistics Registry, finance department, researchers, etc.), Educational Institutions, Patient Quality and Safety, Infection Control, Business and Data Analysts, Project Managers, Accommodation and Facilities expenses and Data Quality Check Unit. The framework also includes Maintenance Cost, Quality Assurance Reports, Clinical Quality of Documentation, Coding Quality Support Teams, Orientation of New Hires and Reinforcement training after implementation under productivity losses in addition to the already accounted for coding productivity losses. System Conversion costs include Software Vendors, Data Conversion Cost, Data Ware house, Data Dictionary and Decision Support Systems, Licensing fees, Health Payers, Coding Automation, Privacy and Security Impact, Intelligent Medical Objects and other user costs (such as Analytics, CIHI, etc.), as well as secondary users:

- Manitoba Center for Health Policy
- Surgical Information Management System
- HIS (Health Information System)
- ADT (Admissions, Discharges and Transfers)
- Grouper Software
- Online courses

Table 1 provides a detailed picture of what each category of cost entail and the possible differences in cost calculation for British Columbia, Alberta, Manitoba and Saskatchewan.

Table 1: ICD Transition Cost Estimation Items and Variables: Framework for ICD-10 to ICD-11 transition

COST OF RE-TRAINING	**Alberta**	**Manitoba**	**British Columbia**	**Saskatchewan**
Equivalent Full-time Coders	Number of full time coders in hospitals Number of retraining hours Hourly wage		Took 1 year to implement ICD 10; 100 to 200 coders Coding system in place for 35 hospitals	One regional Health Authority Cost of retraining: 119 full time coders, 31 part time coders, 17 analysts, 11 in e-health, 7 in cancer, 16 in health quality council 38 in Saskatchewan, 35 in Regina, 25 in other regions
Coding Support Unit			Three health authorities: Fraser Health, Vancouver Coastal Health, PSA	3 trainers, travel expense 15 sessions of 2-day workshops: 13 participants at each workshop
End Users: Data Sharing (such as with Employees of government organizations such as Analytics at Alberta Health & Wellness, CIHI, Federal Government Vital	Number of users Number of retraining hours Hourly wage Lost work-time		Nurse practitioners	

COST OF RE-TRAINING	Alberta	Manitoba	British Columbia	Saskatchewan
Statistics Registry, finance department, researchers, etc.)				
Educational Institutions			Canadian Healthcare Institute	Saskatoon Polytechnic: Trains 3 to 4 instructors and 8 to 14 graduates every second year
Physicians/ Clinicians	Number of physicians/ clinicians Number of retraining hours Hourly wage Lost work-time			Uses ICD 9 for physician claims; will consider using ICD 11
Patient Quality and Safety				
Infection Control				
Business and Data Analysts	Number of analysts Number of retraining hours Hourly wage			Analysts currently use ICD-10
Project Managers	Number of managers Number of retraining hours Hourly wage		4 portfolios: one manager and coding director for each region	
Accommodation, Travel & Facilities expenses				
Data Quality Checks	Number of people Number of retraining hours Hourly wage			E-Health

PRODUCTIVITY LOSSES	**Alberta**	**Manitoba**	**British Columbia**	**Saskatchewan**
	Considering break-in period of 3-6 months	No matter how well trained the coders are, there is still a backlog of work. This might affect productivity even after a few years into ICD 11 implementation and hence productivity loss may not be accurately estimated.	Prenatal and Neonatal Database, National Rehab Data, Trauma, Cardiac, Home based Care. Data available from last 16 years. The province is creating a provincial productivity standard Should consider both short-term and long-term productivity losses	The province did not go into adopting ICD 10 as a province; each region was separate. By 2004 every region had adopted ICD 10. Regina was the last to adopt and caused 2 extra years of lost productivity in Saskatchewan due to coding time lags in Regina. Lagged coding work was completed by contract coders during implementation. Ministry managed overtime costs
Coders	Time difference to code with ICD-11 from ICD-10 Number of hospital discharge patients/month Hourly wage Extra coder hours required for ICD-11	ICD 10 implementation took 2 years. It took the province 9 months for coders to get back to previous productivity levels.		
Physicians	% of patient records not routine and need clarification of physician Physician time needed per record Number of discharge per month Combined cost physicians and coders Additional physician hours			

PRODUCTIVITY LOSSES	**Alberta**	**Manitoba**	**British Columbia**	**Saskatchewan**
Reinforcement training after implementation				
Maintenance Cost	Hours of technical support needed to fix broken pathway (Validation and correction) Hourly wage Lost work time	Folio for ICD 10 is not working properly	Implementation of ICD 11 may coincide with new fiscal year, which would result in the coders to work additional time to take care of backlog of data. Orientation and training of new hires would also cost the province time	More time is required to convert data which takes away time from other projects The province hopes the algorithm/ conversion tables will work better this time
Quality Assurance Report				
Clinical Quality of Documentation				
Coding Quality Support Teams				
Orientation of New Hires				
Other Users				Readers, scanners, windows systems may not support new programs. Technical educational programs may be necessary to teach professionals on how to use electronics such as computers

COST OF SYSTEM CHANGES	**Alberta**	**Manitoba**	**British Columbia**	**Saskatchewan**
		The province has a centralized vendor but different regions	Software updates are up to the individual organization	Updates for Med2020, Folio, Encoder, 3M, SCM

COST OF SYSTEM CHANGES	Alberta	Manitoba	British Columbia	Saskatchewan
		have different contracts, which means cost estimations will be different for each health region		for clinical documentation Physician Billing System Hospital Reciprocal Billing System. Current software capable to support ICD 11and updates were done during ICD 9 implementation
Software Vendors	Estimated time of work Estimated total cost of designing and adapting/updating software		Should be divided into Coding and Abstracting Clinical information codes Intervention codes	There are 3 software vendors for Med2020 and e-health vendors for others
Data Conversion Cost	Possible to get an estimated amount from previous records			
Data Ware house, Data Dictionary and Decision Support Systems		Software is updated every 3 to 5. The interviewees emphasized the need to update software and hardware to handle ICD 11 data; For example, the Folio system software and Abstracting software		
Licensing fees				
Health Payers	Cost of adapting the new codes and reprogramming logic			

COST OF SYSTEM CHANGES	**Alberta**	**Manitoba**	**British Columbia**	**Saskatchewan**
	Impacted IT application			
Other Users (such as Analytics, CIHI, etc.) Secondary users: Manitoba Center for Health Policy Surgical Information Management System HIS ADT Grouper Software Online courses	Impacted IT application cost			
Privacy and Security Impact				
Intelligent Medical Objects				
Coding Automation				

The table above shows implementation techniques, resources and methods will need to be different for the different provinces. This would likely entail very different costs of implementation of ICD-11. Some provinces may therefore have higher costs. For example, Manitoba mentioned requiring massive software updates compared to British Columbia, which is already very advanced in the usage of technology in healthcare. As the ICD-11 implementation moves forward, this thesis and the subsequent economic evaluations would provide substantial context into the implementation process of ICD-11 and the resource allocation required for Canada and other countries which will implement ICD-11 in the near future.

5 Recommendations for Sustaining and Enhancing the Big Data Revolution

With data analytics entering the healthcare sector, it seems that the big data revolution is well underway. Bourgi et. al (2015) claim that the potential impact of big data can be measured. In the 21st century, the increased amount of data amassed can be analysed with greater efficiency and accuracy due to the higher performing computing power.

Furthermore, despite the vast amount of data being collected and coded by health coders, there has been little impact of the data in research. Most of the data is sent to Canadian Institutes of Health Research to allow for decisions regarding funding allocation. However, there is great potential for this data to be used in active research involving trends of diseases, disease management, analysis and possible breakthroughs in technological advancements in healthcare.

It is vital for the data to be organized in a manner which facilitates data analysts to glean information for analysis and evaluation. Enhanced monitoring of hospital activity, which may be necessary with newer forms of hospital funding such as activity-based funding, would require more information and an even better surveillance of data input from hospital operations which requires a strong data system. Given the results from the ICD case study, we find that the healthcare industry still has a lot to grasp on to with regards to efficient conveyance of data for research and activity-based funding.

The usage of data analytics in healthcare has been slow, especially in the case of physicians using the data to inform them on their treatment decisions and usage of data analytics in hospital management and funding decisions. Moreover, many companies have not yet grasped onto the potential of existing data that they have collected. It is important that the momentum in the usage of data in healthcare is sustained in order to facilitate effective economic evaluation and research. Groves et al. (2013) encourages a collaborative environment between patients, pharmaceutical companies, researchers, health payers and drug manufacturers would be the most efficient way of ensuring this momentum. Below are a few ways this could be done:

I. Healthcare providers such as physicians, clinicians and hospitals should strive to assess the trends in treatments which are best for patients, use the EMR data to reinforce their medical care for patients. Adopt care practices which capture data from embedded medical devices, mobile devices, etc. for added information. Participate and share information easily and securely for enhanced patient health outcomes.

II. Information should be available in a centralized location, with standardized practices for sharing and using sensitive data securely. The Office of the Information and Privacy Commissioner of Alberta has planned actions to ensure the privacy of patient information and such approaches must be adopted by all the provinces in Canada.

III. There should be a focus on balancing the cost and patient health outcomes, so that the value of big data is fully realized.

IV. Manufacturers and pharmaceutical companies should try to keep abreast of patient demand and treatment data available to try to understand what works best. They should align the research and development approaches for product development to include data driven solutions.

V. If and when payment reforms occur, the payers should try to negotiate rates by leveraging data driven information from both health providers as well as from patients claims data. This would encourage health providers to use the data available to inform on patient demand, as well as analyse which treatments work best for certain kinds of diseases affecting patients. This would also incentivise better and clearer analytical methods for understanding treatment and drug performances. Payers should also try to partner with manufacturers and healthcare providers to understand whether costs for certain health conditions were lower than usual and assessing the trends which drive the costs.

VI. Companies with data-based initiatives must strive to make concentrated efforts in pushing their products forward. This would entail hiring the best data scientists and business individuals who aim to ensure a data driven healthcare industry. Companies must also develop different big data capabilities which will ensure that businesses have the highest effect to take big data analytics to the next edge. (Groves, Kayyali, Knott & Kuiken, 2013).

The above recommendations are just some simple guidelines which may help take the Canadian healthcare industry in becoming more integrated in using data analytics to drive patient's healthcare outcomes, which would enable better and efficient economic research to make a positive impact on patients' lives.

6 Conclusion

Economic Evaluation is an important way of knowing whether a project is feasible to be implemented. The method identifies costs and benefits and outlines the various ways a project can be conducted to bring about efficiency. The evaluation report thus enables the businesses to plan a budget and operate in a timely manner.

As data becomes more readily available, it would be easier to employ advanced statistical analysis for economic evaluation. The healthcare industry currently has a collection of raw data due to extensive chart recording practices throughout Canada. However, it still lacks the infrastructure and guidance for the proper utilization of the vast information. This thesis elaborates on the importance of efficient utilization of data to reap the benefits. The paper has used a case study on ICD-9 to ICD-10 transition to illustrate the extent of data utilization that is currently under way in Canada, what kind of data analytics the country lacks, the ways forward and the importance of economic evaluation in big data.

The thesis provides a comprehensive guideline of the benefit and cost estimation methodology, including potential sources of data from DIMR data dictionary, in Appendix A. The methodology could not be completed as data for the years 2000 to 2003 is not available and construction of benefit scenarios have not been possible in the Canadian context. For the benefit calculation, a crude measure of estimation has been conducted using a data quality measure which depicts a higher number of published papers used ICD-10 codes compared to ICD-9 codes and thus making an argument that ICD-10 has better quality of data in terms of numbers of codes, specificity and ease of usage and understanding. However, this simple method of enumerating data quality has various pitfalls and cannot be taken as a standalone depiction for data quality.

The paper also outlined the gradual movement towards the adoption of ICD-11, which is more advanced, contains more codes with higher specificity and is compatible to connect with previous systems of coding. A table outlines the variables ICD-11 transition study will look at and also includes the differences in cost estimation which have to be considered for provinces such as Manitoba, Alberta, Saskatchewan and British Columbia.

Both payment reform and auto coding will change the cost estimations. As we move towards an era of information technology, it is important that data is available and accurately organized to facilitate evaluation. As experienced during the cost and benefit estimations case study, there is a considerable lack of data available for calculations. If data is not available due to ethical concerns or inadequate data structure, research into healthcare would be slow to materialize. Furthermore, government expenditure requires empirical evidence to be aligned with activity-based funding. It is evident that only accurate data analytics can ensure efficient allocation of resources for a better managed universal healthcare system for Canada.

Given the advancements in technology and the recent strides being made towards a more data-oriented health system, it is of utmost importance that available and synthesized data is integrated into the healthcare sector, while enhancing the current momentum of data analytics.

7 References

1. Adams, A. (2018, April 9). 20. The Role of Big Data and Mobile Apps in Healthcare. Dataconomy. Retrieved July 18, 2018 from http://dataconomy.com/2018/04/the-role-of-big-data-and-mobile-apps-in-healthcare/
2. Alberta Peer-to-Peer EMR Network Program. (2016). *Alberta Medical Association*. Retrieved August 15, 2018, from https://www.albertadoctors.org/leaders-partners/emrs/emr_resources/Peer-to-Peer-EMR
3. Armah, N. A. (2013). Big data Analysis: The next frontier. *Canadian Economic Analysis*. Retrieved August 8, 2018.
4. Bacchus, B. & Esmail, N. (2013). Waiting your turn: Wait times for healthcare in Canada, 2013. Fraser Institute.
5. Bacchus, B., Hasan, S. & Timmermans, I. (2017). Comparing performance of universal healthcare countries, 2017. Fraser Institute.
6. Bourgi, S., Castillo, D. D. & Freund, G. (2015). Big Data and the Intelligence Economy. Information and Communications Technology Council. Retrieved September 20, 2018 from https://www.ictc-ctic.ca/wp-content/uploads/2015/12/BIG-DATA-2015.pdf
7. Canadian Institute for Health Information (2013). Evaluating the Impact of Activity-Based Funding on Health System Performance.
8. Cattel, J., Chilukuri, S., & Levy, M. (2013). How big data can revolutionize pharmaceutical R&D. Mckinsey&Company. Retrieved July 17, 2018 from https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d
9. Charlson M. E., Pompei P., Ales K. L., MacKenzie C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 40:373- 383.
10. Chen, Y., Guzauskas, G. F., Gu, C., Wang, B. C. M., Furnback, W. E., Xie, G., … Garrison, L. P. (2016). Precision Health Economics and Outcomes Research to Support Precision Medicine: Big Data Meets Patient Heterogeneity on the Road to Value. *Journal of Personalized Medicine*, *6*(4), 20. http://doi.org/10.3390/jpm6040020
11. Copping, R., & LI, M. (2016). 18. The Promise and Challenge of Big Data for Pharma. Informatica. Retrieved July 17, 2018, from https://www.informatica.com/content/dam/informatica.com/global/amer/us/collateral/executive-brief/big-data-pharmaceutical-industry_ebook_2341.pdf
12. Deloitte (2010). ICD-10 Implementation for Health Care Providers: The business imperative for compliance.
13. Dutton, G. (2014). Big data and climate clues are helping health officials predict disease in order to prevent it. Forbes. *EMC BRANDVOICE*. Retrieved September 6, 2018, from https://www.forbes.com/sites/emc/2014/06/20/big-data-and-climate-clues-are-helping-health-officials-predict-disease-in-order-to-prevent-it/

14. Einav, L., & Levin, J. (2014). Economics in the age of big data. Science, 346(6210), 1243089–1243089. https://doi.org/10.1126/science.1243089
15. Elixhauser, A., Steiner, C., Harris, D. R., Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Med Care*. 36(1):8-27.
16. Esmail, N. (2009). Activity Based Funding is Good for Alberta. Fraser Institute. Retrieved from https://www.fraserinstitute.org/article/activity-based-funding-good-alberta
17. Esmail, N., Hazel, M. & Walker, M. A. (2013). Waiting your turn: Wait times for healthcare in Canada, 2008. Fraser Institute.
18. Gerdeman, D. (2015). Consumer-centered Health Care Depends on Accessible Medical Records. *Harvard Business School.* Retrieved September 20, 2018 from https://hbswk.hbs.edu/item/consumer-centered-health-care-depends-on-accessible-medical-records
19. Gibson, H. (2012). EHR does not eliminate Medical Coders and Billers. *M-scribe.* Retrieved September 9, 2018 from https://www.m-scribe.com/blog/bid/231536/EHR-Does-Not-Eliminate-Medical-Coders-and-Billers
20. Groves, P., Kayyali, B., Knott, D. & Kuiken, S. V. (2013). The Big Data Revolution in Healthcare. Center for US Health System Reform Business Technology Office.
21. Hanna, J. (2018). A better business model for fighting cancer. *Harvard Business School.* Retrieved September 20, 2018 from https://hbswk.hbs.edu/item/a-better-business-model-for-fighting-cancer
22. Health Affairs (2014). Transitioning to ICD-10. Health Policy Brief. Retrieved September 20, 2018 from https://www.healthaffairs.org/do/10.1377/hpb20140403.830190/full/
23. Health Spending (2018). Canadian Institute for Health Information. Retrieved July 12, 2018 from https://www.cihi.ca/en/health-spending
24. Holland, R. (2018). Electronic Health Records were supposed to cut medical costs. They haven't. *Harvard Business School.* Retrieved September 20, 2018 from https://hbswk.hbs.edu/item/electronic-health-records-were-supposed-to-cut-health-costs-they-haven-t
25. *ICD-10-CA/CCI Business Case* (pp. 1-19, Rep.). (2001). Calgary, AB: Calgary Health Region. Unpublished Manuscript.
26. Leung, L. (2016). Health Economic Evaluation: A Primer for Healthcare Professionals. *Primary Health Care Open Access*, 6(2). doi:10.4172/2167-1079.1000223
27. Libicki, M. & Brahmakulam, I. (2004). 'The Costs and Benefits of Moving to the ICD-10 Code Sets.' RAND Corporation. TR-132-DHHS.
28. Losev, T. (2017). Developing a Cost Projection for ICD-11 Implementation in Canada. Unpublished manuscript.
29. Losev, T. (2017). Estimating the Dollar Value of Big Data Analytics in Healthcare: A scoping Study of Cost-Benefit Estimation of Big Data in Healthcare. Unpublished manuscript.

30. McLaren, N. and R. Shanbhogue. (2011). "Using Internet Search Data as Economic Indicators." Bank of England *Quarterly Bulletin* (Q2): 134–40.
31. Palmer, S., Byford, S., & Raftery, J. (1999). Types of economic evaluation. Bmj, 318(7194). doi:10.1136/bmj.318.7194.1349-a
32. Phillips, K. A., Douglas, M. P., Trosman, J. R., & Marshall, D. A. (2017). "What Goes Around Comes Around": Lessons Learned from Economic Evaluations of Personalized Medicine Applied to Digital Medicine. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *20*(1), 47–53. http://doi.org/10.1016/j.jval.2016.08.736
33. Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems. Retrieved from http://www.hissjournal.com/content/2/1/3
34. Robillard, W. & Tolfree, M. (2015). An Overview of Alberta's Electronic Health Record Information System. Retrieved August 2, 2018 from http://www.albertanetcare.ca/documents/An_Overview_of_Albertas_ERHIS.pdf
35. Ross, J. W., Beath, C. M. & Quaadgras, A. (2013). You may not need big data after all. *Harvard Business Review.* Retrieved September 20, 2018 from https://hbr.org/2013/12/you-may-not-need-big-data-after-all
36. Shah, N. D. & Pathak, J. (2014). Why healthcare may finally be ready for big data. *Harvard Business Review.* Retrieved September 20, 2018 from https://hbr.org/2014/12/why-health-care-may-finally-be-ready-for-big-data
37. Southern, D. A., Hall, M., White, D. E., Romano, P. S., Sundararajan, V., Droesler, S. E., Pincus, H. A., Ghali, W. A. (2016). Opportunities and challenges for quality and safety applications in ICD-11: An international survey of users of coded health data. *International Journal for Quality in Health Care. 28 (1). 129-135.* Retrieved October 23, 2018 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4767047/pdf/mzv096.pdf
38. Tamblyn, R., McMahon, M., Girard, N., Drake, E., Nadigel, J., & Gaudreau, K. (2016). Health services and policy research in the first decade at the Canadian Institutes of Health Research. *CMAJ Open*, *4*(2), E213–E221. http://doi.org/10.9778/cmajo.20150045
39. Tibben, W. & Wamba, S. F. (2018). Exploring the potential of big data on the health care delivery value chain (CDVC): a preliminary literature and research agenda. *Proceedings of the 51st Hawaii International Conference on System Sciences.* Retrieved September 20, 2018 from https://www.researchgate.net/profile/Samuel_Fosso_Wamba/publication/323380677_Exploring_the_potential_of_big_data_on_the_health_care_delivery_value_chain_CDVC_a_preliminary_literature_and_research_agenda/links/5a931e1ba6fdccecff059d1a/Exploring-the-potential-of-big-data-on-the-health-care-delivery-value-chain-CDVC-a-preliminary-literature-and-research-agenda.pdf
40. Universities Canada (2018). Budget 2018 makes historic investment in Canada's research future. Retrieved June 20, 2018 from https://www.univcan.ca/media-room/media-releases/budget-2018-makes-historic-investment-canadas-research-future/

41. Weir, G. (2017). Leveraging big data in the healthcare industry. IT world Canada

Appendix A: Probable codes and data sources for estimating Cost and Benefits Variables for ICD-9 to ICD-10 transition

Available codes from DIMR data dictionary

Variables for Benefit Estimation	**Possible Data Sources**
1. Reduction in resource cost	AHSDATA: a high-level grouping of clinically similar clinical
Hospital funding on different disease types	Resource intensity weight: values provide a measure of a patient's relative resource consumption compared to an average typical ambulatory care case
Prevalence/ Occurrence of diseases: increase or decrease?	
Funding increased/ decreased?	
2. Gains in specificity	AHSDATA
Higher number of classification	Code that identifies additional specificity for diagnosis codes.
Newest procedures	DAD: Code for service most responsible for caring for the patient during his/her hospital stay, determined by the most responsible diagnosis.
	MPSERV: Code for service most responsible for caring for the patient during his/her hospital stay, determined by the most responsible diagnosis.
	MPSUBSERV: Code that further defines types of patients treated within the main patient service
3. Number of Patients sent back home earlier	AHSDATA
Length of stay: has it reduced after ICD 10 coding?	Total length of stay in minutes for the patient's visit
Number of hospital discharges	DXCODE: ICD-10-CA diagnosis code that describes the diagnoses/conditions/problems/circumstances of the patient during the length of stay in the health care facility
	ED Length for Admitted Patients in minutes

	Inpatient: Date and Time the patient physically left the ER and was moved to the inpatient unit or the OR or diagnostic area.
	Total Length of Stay
	Length of stay: Difference in days between the admission date and the discharge date.
4. Ambulatory care: NACRS info on emergency and outpatient visits (demographics, clinical, administrative, service-specific)	AHSDATA; AMBARRIVAL_DTTM; AMBTRANSFR_DTTM; ARRIVE_TIME_SRC
Faster service	Date and time when ambulance pulls into hospital driveway
Time required to reach patient	
5. Diabetes disease management	
Number of Patients identified	ALT_INDIVIDUALIZED_AEDC_1011: Diabetes
Reduction in Hb1Ac levels	AB_POP_COST_UTL_DSP: Kidney Infections
Other diseases due to high blood sugar: kidney problems, cataract?	Kidney Malignancy
	Kidney and Urinary Tract Stones
	Kidney or Ureter Trauma
6. Reduction in time required for coding	*Unable to find relevant data source
Number of hours saved due to disease management	
Remuneration of coders per hour	
7. Effective structure	AHSDATA; DXCLUSTER

Has the use of combination codes increased?	A problem cluster is an alpha character assigned to two or more valid ICD-10-CA diagnosis codes to indicate when more than one ICD-10-CA code is required to describe a circumstance/condition
8. Data Quality	*Estimation conducted by enumerating published reports on PUBMED
Number of published reports using coded data before and after ICD 10	

Variables for Cost Estimation	**Possible Data Sources**
1. Cost of Re-training	AHSDATA
Number of coder staff in Calgary	CODER_NO: Coder number of staff that abstracted the data
Cost of Re-training	*estimation from Calgary Health Region Business Case Report
Coding staff wages per hour	
Coding manager wages per hour	
Hours in a work day for each coder	
Number of retraining hours needed	
Retraining expenses/facility and travel expenses	
Number of facilities affected by ICD implementation	
Overtime charged	
2. Software provider fees	*Unable to find relevant data source

Number of hours required to establish new software: affected productivity?	
3. Productivity estimates	*performance appraisal and productivity reports yet to be received
Extra hours required to complete task/ number of reports completed in a day before and after ICD transition	
4. Total number of weeks in retraining: 6 weeks to 6 months	*estimation from Calgary Health Region Business Case Report
5. Maintenance cost	*performance appraisal and productivity reports yet to be recieved
Hours devoted to fix broken pathways	
Remuneration of coders	
6. Billing and administrative expenses	*Unable to find relevant data source
Time to evolve	
Cost of Integration	

Other Data Sources

Other Data	PROJFLD
	AEDC015: Acute Mental Health Diagnoses
	ALL_AMB_COSTS: Ambulatory Care Costs, which combines all 3 ambulatory physician claims and AMBCOSTS after subtracting ED_UCC_COSTS_1011 from AMBCOSTS. If this subtraction results in a negative number, the result is set to zero. This occurs because the methodology used to create the ED_UCC_COSTS_1011 differs from the one used to create the AMBCOST1011.

	ALT_PREM_ARRANGEMENT_1011: Alberta Health Care Insurance Plan Premium Payment Arrangement
	AMBCOST1011: Each field used to collect codes as defined by CIHI or Province/Territory, or user defined
	Ambulatory Care Costs (include hospital day procedures, ED/UCC, hospital clinics, and some stand-alone clinics costs)
	CIHI_IT_CATEGORY_CODE: Investigative technology includes CCI codes for interventions such as X-Ray, Ultrasound, CT Scan and MRI. These codes are also used to define Direct Diagnostic Imaging CACS cells. Categories of Investigative Technology. Refer to worksheet 9 - Processed Data Codes for description of code values.
	CRG_1011_METHOD_1: CRG Codes Produced according to DIMR Method 1
	DC_AMB_OTHER_ALLIED_1011: Allied Professional Claims for Ambulatory Services
	DC_AMB_OTHER_GP_1011: General Practitioners' (GP) Claims for Ambulatory Services
	DC_AMB_OTHER_SPEC_1011: Specialists' (SP) Claims for Ambulatory Services
	DC_COMMUNITY_ALLIED_1011: Community Allied Professional Claims
	DC_COMMUNITY_GP_1011: Community GP Claims
	DC_COMMUNITY_SPEC_1011: Community SP Claims
	VND_ANES_CATEGORY Vendor Anaesthetic Category: This data element is to be populated by the vendor assigned anaesthetic category value. When reported, the value will be compared against the CIHI assigned anaesthetic category value. If they do not match, a warning message will be generated
	IP_PSY_COST_1011: The Hospital Portion of Inpatient Care Costs Including Stand-alone Psychiatric Hospitals
	LTC1011: LCT Clients Indicator if a patient had LTC costs recorded for 2010/11.
	SCUHOURS: Hours calculated for time patient spent in a special care unit
	CMG: The Code for case mix groups are designed to aggregate acute care inpatients that are similar clinically and in terms of resource use.
	DXCODE: ICD-10-CA diagnosis code that describes the diagnoses/conditions of the patient during the length of stay in the health care facility
	DXTYPE: Code that identifies most responsible diagnosis, any pre-admit or post-admit comorbidity, secondary diagnosis, morphology, etc.

Appendix B: List of published papers for data quantity and quality estimation

Papers which used ICD-9 codes

1. Doshi, R., Shah, J., Jauhar, V., Decter, D., Jauhar, R., & Meraj, P. (2018). Comparison of drug eluting stents (DESs) and bare metal stents (BMSs) with STEMI: Who received BMS in the era of 2nd generation DES? *Heart & Lung,47*(3), 231-236. doi:10.1016/j.hrtlng.2018.02.004
2. Hisle-Gorman, E., Susi, A., Stokes, T., Gorman, G., Erdie-Lalena, C. & Nylund, C. M. (2018). Prenatal, perinatal, and neonatal risk factors of autism spectrum disorder. *Pediatric Research*. doi: 10.1038/pr.2018.23
3. Howard, J. T., Sosnov, J. A., Janak, J. C., Gundlapalli, A. V., Pettey, W. B., Walker, L. E., & Stewart, I. J. (2018). Associations of Initial Injury Severity and Posttraumatic Stress Disorder Diagnoses With Long-Term Hypertension Risk After Combat InjuryNovelty and Significance. *Hypertension,71*(5), 824-832. doi:10.1161/hypertensionaha.117.10496
4. Kenna, M. A., Irace, A. L., Strychowsky, J. E., Kawai, K., Barrett, D., Manganella, J., & Cunningham, M. J. (2018). Otolaryngologic Manifestations of Klippel-Feil Syndrome in Children. JAMA Otolaryngology–Head & Neck Surgery, 144(3), 238. doi:10.1001/jamaoto.2017.2917
5. Quan, H., Parsons, G. A. & Ghali, W. A. (2004). Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administration data. *Medical Care, 42* (8), doi: 10.1097/01.mlr.0000132391.59713.0d
6. Xie, L., Semenciw, R., & Mery, L. (2015). Cancer incidence in Canada: trends and projections (1983–2032). *Health Promotion and Chronic Disease Prevention in Canada : Research, Policy and Practice*, *35*(Suppl 1), 1–187.

Papers which used ICD-10 codes

1. Al-Jaishi, A. A., Moist, L. M., Oliver, M.J., Nash, D.M., Fleet, J.L., Garg, A.X. & Lok, C. E. (2018). Validity of administrative database code algorithms to identify vascular access placement, surgical revisions, and secondary patency. *The Journal of Vascular Access,* doi: 10.1177/1129729818762008
2. Amram, O., Schuurman, N., Pike, I., Friger, M., & Yanchar, N. L. (2016). Assessing access to paediatric trauma centres in Canada, and the impact of the golden hour on length of stay at the hospital: an observational study. *BMJ Open*, *6*(1), e010274. http://doi.org/10.1136/bmjopen-2015-010274
3. Berthelot, S., Lang, E. S., Quan, H. & Stelfox, H. T. (2014). *Identifying Emergency-Sensitive Conditions for the Calculation of an Emergency Care Inhospital Standardized Mortality Ratio.* Annals of Emergency Medicine. 63 (4), 418-424.
4. Chan, V., Thurairajah, P., & Colantonio, A. (2013). Defining traumatic brain injury in children and youth using International Classification of Diseases version 10 codes: a systematic review protocol. *Systematic Reviews*, *2*(102). http://doi.org/10.1186/2046-4053-2-102

5. Chan, V., Thurairajah, P., & Colantonio, A. (2015). Defining pediatric traumatic brain injury using International Classification of Diseases Version 10 Codes: A systematic review. *BMC Neurology*, *15* (7). http://doi.org/10.1186/s12883-015-0259-7
6. Costantino, G., Ruwald, MH., Quinn, J., Camargo, CA. Jr., Dalgaard, F., Gislason, G., Goto, T., Hasegawa, K., Kaul, P., Montano, N., Numé, AK., Russo, A., Sheldon, R., Solbiati, M., Sun, B. & Casazza, G. (2018). *JAMA Internal Medicine, 178 (3), 356-362.* doi: 10.1001/jamainternmed.2017.8175
7. Frolova, N., Bakal, J. A., McAlister, F. A., Rowe, B. H., Quan, H., Kaul, P. & Ezekowitz, J. A. (2015). Assessing the Use of International Classification of Diseases-10th Revision Codes From the Emergency Department for the Identification of Acute Heart Failure. *JACC: heart failure. Elsevier Inc.*, 3 (5), doi: http://dx.doi.org/10.1016/j.jchf.2015.02.004
8. Gentry, C. A., Greenfield, R. A., Slater, L. N., Wack, M. & Huycke, M. M. (2000). Outcomes of an antimicrobial control program in a teaching hospital. *American Journal of Health-System Pharmacy,* 57 (3), 268-274.
9. Hennessy, D. A., Quan, H., Faris, P. D., & Beck, C. A. (2010). Do coder characteristics influence validity of ICD-10 hospital discharge data? *BMC Health Services Research*, *10*, 99. http://doi.org/10.1186/1472-6963-10-99
10. Hohl, C. M., Karpov, A., Reddekopp, L., & Stausberg, J. (2014). ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. *Journal of the American Medical Informatics Association: JAMIA*, *21*(3), 547–557. http://doi.org/10.1136/amiajnl-2013-002116
11. Hopkins, R. B., Burke, N., Von Keyserlingk, C., Leslie, W. D., Morin, S. N., Adachi, J. D., … Tarride, J. (2016). The current economic burden of illness of osteoporosis in Canada. *Osteoporosis International: A Journal Established as Result of Cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, *27*(10), 3023–3032. http://doi.org/10.1007/s00198-016-3631-6
12. Hussain, M., Alali, A., Mamdani, M., Saposnik, G., Salata, K., & Nathens, A. (2018). Risk of Intracranial Hemorrhage After Carotid Artery Stenting Versus Endarterectomy: A Population-Based Study. Journal of Vascular Surgery, 67(5), 1628-1629. doi:10.1016/j.jvs.2018.03.005
13. Januel, J.-M., Luthi, J.-C., Quan, H., Borst, F., Taffé, P., Ghali, W. A., & Burnand, B. (2011). Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. *BMC Health Services Research*, *11*, 194. http://doi.org/10.1186/1472-6963-11-194
14. Jean, S., O'Donnell, S., Lagacé, C., Walsh, P., Bancej, C., Brown, J. P., … Leslie, W. D. (2013). Trends in Hip Fracture Rates in Canada: An Age-Period-Cohort Analysis. *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research*, *28*(6), 1283–1289. http://doi.org/10.1002/jbmr.1863
15. Jolley, R. J., Quan, H., Jetté, N., Sawka, K. J., Diep, L., Goliath, J., … Doig, C. J. (2015). Validation and optimisation of an ICD-10-coded case definition for sepsis using

administrative health data. *BMJ Open*, *5*(12), e009487. http://doi.org/10.1136/bmjopen-2015-009487
16. McChesney-Corbeil, J., Barlow, K., Quan, H., Chen, G., Wiebe, S., & Jette, N. (2017). Validation of a Case Definition for Pediatric Brain Injury Using Administrative Data. Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques,44(2), 161-169. doi:10.1017/cjn.2016.419
17. Pang, J. X., Ross, E., Borman, M. A., Zimmer, S., Kaplan, G. G., Heitman, S. J., … Myers, R. P. (2015). Validation of coding algorithms for the identification of patients hospitalized for alcoholic hepatitis using administrative data. *BMC Gastroenterology*, *15*, 116. http://doi.org/10.1186/s12876-015-0348-5
18. Patel, A. B., Quan, H., Welsh, R. C., Deckert-Sookram, J., Tymchak, W., Sookram, S., … Kaul, P. (2015). Validity and utility of ICD-10 administrative health data for identifying ST- and non-ST-elevation myocardial infarction based on physician chart review. *CMAJ Open*, 3(4), E413–E418. http://doi.org/10.9778/cmajo.20150060
19. Peng, M., Sundararajan, V., Williamson, T., Minty, E. P., Smith, T. C., Doktorchik, C. T. A., & Quan, H. (2018). Data on coding association rules from an inpatient administrative health data coded by International classification of disease - 10th revision (ICD-10) codes. *Data in Brief*, *18*, 710–712. http://doi.org/10.1016/j.dib.2018.02.043
20. Quan, H., Eastwood, C., Cunningham, C. T., Liu, M., Flemons, W., De Coster, C., … for the IMECCHI investigators. (2013). Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open*, *3*(10), e003716. http://doi.org/10.1136/bmjopen-2013-003716
21. Roston, T. M., Tran, D. T., Sanatani, S., Sandhu, R., Sheldon, R., & Kaul, P. (2018). A Population-Based Study of Syncope in the Young. Canadian Journal of Cardiology, 34(2), 195-201. doi:10.1016/j.cjca.2017.12.006
22. Schwarzinger, M., Pollock, B. G., Hasan, O. S. M., Dufouil, C., Rehm, J.(2018). Contribution of alcohol use disorders to the burden of dementia in France 2008-13: a nationwide retrospective cohort study. *Lancet Public Health,* 3(3), 124-132. QalyDays Study Group. doi: 10.1016/S2468-2667(18)30022-7.
23. Smartt, P., & Chalmers, D. (2010). Searching for ski-lift injury: An uphill struggle? Journal of Science and Medicine in Sport, 13(2), 205-209. doi:10.1016/j.jsams.2009.01.004
24. Southern, D. A., Doherty, C., De Souza, M. A., Quan, H., Harrop, A. R., Nickerson, D., & Rabi, D. (2015). Charts versus Discharge ICD-10 Coding for Sternal Wound Infection Following Coronary Artery Bypass Grafting. *Perspectives in Health Information Management*, *12*(Summer), 1e.
25. Tanno, L. K., Bierrenbach, A. L., Simons, F. E. R., Cardona, V., Thong, B. Y.-H., Molinari, N., … on behalf the Joint Allergy Academies. (2018). Critical view of anaphylaxis epidemiology: open questions and new perspectives. *Allergy, Asthma, and Clinical Immunology: Official Journal of the Canadian Society of Allergy and Clinical Immunology*, *14*, 12. http://doi.org/10.1186/s13223-018-0234-0

26. Tarride, J.-E., Hopkins, R. B., Burke, N., Guertin, J. R., O'Reilly, D., Fell, C. D., Dion, G., Kolb, M. (2018). Clinical and economic burden of idiopathic pulmonary fibrosis in Quebec, Canada. *Clinico Economics and Outcomes Research: CEOR*, *10*, 127–137. http://doi.org/10.2147/CEOR.S154323
27. Verma, A., Ritchie, MD. (2017). Current Scope and Challenges in Phenome-Wide Association Studies. *Current Epidemiology Reports,* 4(4), 321-329. doi: 10.1007/s40471-017-0127-7.
28. Youngson, E., Welsh, R. C., Kaul, P., McAlister, F., Quan, H., & Bakal, J. (2016). Defining and validating comorbidities and procedures in ICD-10 health data in ST-elevation myocardial infarction patients. *Medicine*, *95*(32), e4554. http://doi.org/10.1097/MD.0000000000004554
29. Yu, A. Y. X., Quan, H., McRae, A., Wagner, G. O., Hill, M. D., & Coutts, S. B. (2017). Moderate sensitivity and high specificity of emergency department administrative data for transient ischemic attacks. *BMC Health Services Research*, *17*, 666. http://doi.org/10.1186/s12913-017-2612-6

Papers which used ICD-9 & ICD-10 codes

1. Aymé, S., Bellet, B., & Rath, A. (2015). Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. *Orphanet Journal of Rare Diseases*, *10*, 35. http://doi.org/10.1186/s13023-015-0251-8
2. De Coster, C., Quan, H., Finlayson, A., Gao, M., Halfon, P., Humphries, K. H., … Ghali, W. A. (2006). Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. *BMC Health Services Research*, *6*, 77. http://doi.org/10.1186/1472-6963-6-77
3. Feudtner, C., Feinstein, J. A., Zhong, W., Hall, M., & Dai, D. (2014). Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatrics*, *14*, 199. http://doi.org/10.1186/1471-2431-14-199
4. Jolley, R. J., Sawka, K. J., Yergens, D. W., Quan, H., Jetté, N., & Doig, C. J. (2015). Validity of administrative data in recording sepsis: a systematic review. *Critical Care*, *19*(1), 139. http://doi.org/10.1186/s13054-015-0847-3
5. Khokhar, B., Jette, N., Metcalfe, A., Cunningham, C. T., Quan, H., Kaplan, G. G., … Rabi, D. (2016). Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open*, *6*(8), e009952. http://doi.org/10.1136/bmjopen-2015-009952
6. Knight, M., Callaghan, W. M., Berg, C., Alexander, S., Bouvier-Colle, M.-H., Ford, J. B., … Walker, J. (2009). Trends in postpartum hemorrhage in high resource countries: a review and recommendations from the International Postpartum Hemorrhage Collaborative Group. *BMC Pregnancy and Childbirth*, *9*, 55. http://doi.org/10.1186/1471-2393-9-55

7. Leisher, S. H., Teoh, Z., Reinebrant, H., Allanson, E., Blencowe, H., Erwich, J. J., … Flenady, V. (2016). Seeking order amidst chaos: a systematic review of classification systems for causes of stillbirth and neonatal death, 2009–2014. *BMC Pregnancy and Childbirth*, *16*, 295. http://doi.org/10.1186/s12884-016-1071-0
8. Li, B., Evans, D., Faris, P., Dean, S., & Quan, H. (2008). Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. *BMC Health Services Research*, *8*, 12. http://doi.org/10.1186/1472-6963-8-12
9. McCormick, N., Bhole, V., Lacaille, D., & Avina-Zubieta, J. A. (2015). Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review. *PLoS ONE*, *10*(8), e0135834. http://doi.org/10.1371/journal.pone.0135834
10. Metcalfe, A., Neudam, A., Forde, S., Liu, M., Drosler, S., Quan, H., & Jetté, N. (2013). Case Definitions for Acute Myocardial Infarction in Administrative Databases and Their Impact on In-Hospital Mortality Rates. *Health Services Research*, *48*(1), 290–318. http://doi.org/10.1111/j.1475-6773.2012.01440.x
11. Quan, H., Li, B., Duncan Saunders, L., Parsons, G. A., Nilsson, C. I., Alibhai, A., & for the IMECCHI Investigators, W. A. (2008). Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database. *Health Services Research*, *43*(4), 1424–1441. http://doi.org/10.1111/j.1475-6773.2007.00822.x
12. So, L., Evans, D., & Quan, H. (2006). ICD-10 coding algorithms for defining comorbidities of acute myocardial infarction. *BMC Health Services Research*, *6*, 161. http://doi.org/10.1186/1472-6963-6-161
13. Southern, D. A., Norris, C. M., Quan, H., Shrive, F. M., Galbraith, P. D., Humphries, K., … the APPROACH Investigators. (2008). An administrative data merging solution for dealing with missing data in a clinical registry: adaptation from ICD-9 to ICD-10. *BMC Medical Research Methodology*, *8*, 1. http://doi.org/10.1186/1471-2288-8-1
14. Southern, D. A., Pincus, H. A., Romano, P. S., Burnand, B., Harrison, J., Forster, A. J., … Ghali, W. A. (2016). Enhanced capture of healthcare-related harms and injuries in the 11th revision of the International Classification of Diseases (ICD-11). *International Journal for Quality in Health Care*, *28*(1), 136–142. http://doi.org/10.1093/intqhc/mzv099
15. St. Germaine-Smith, C., Metcalfe, A., Pringsheim, T., Roberts, J. I., Beck, C. A., Hemmelgarn, B. R., … Jette, N. (2012). Recommendations for optimal ICD codes to study neurologic conditions: A systematic review. *Neurology*, *79*(10), 1049–1055. http://doi.org/10.1212/WNL.0b013e3182684707
16. Sundararajan, V., Henderson, T., Perry, C., Muggivan, A., Quan, H. & Ghali, w. A. (2004). New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *Journal of Clinical Epidemiology*, 57 (12), 1288-94. doi: https://doi.org/10.1016/j.jclinepi.2004.03.012
17. Tonelli, M., Wiebe, N., Fortin, M., Guthrie, B., Hemmelgarn, B. R., James, M. T., … For the Alberta Kidney Disease Network. (2015). Methods for identifying 30 chronic

conditions: application to administrative data. *BMC Medical Informatics and Decision Making*, *15*, 31. http://doi.org/10.1186/s12911-015-0155-5

*1010 articles in PubMed by search term ‘ICD in Canada’

Appendix C: The RAND report

A summary of the RAND report benefit estimation methodology

The benefits had been calculated using a duration of 10 years. Since the USA has a privatized and insurance-based healthcare system, the value of new procedures and enhanced specificity which the ICD 10 system brought could be calculated in monetary terms such as a higher accuracy in payment procedures and reimbursement claims. Due to specificity and the addition of codes for newer medical procedures in ICD 10, it is possible to pay for different procedures differently instead of having to lump a few procedure payments together. Before ICD 10, certain newer procedures were not performed on patients since they were classified into the wrong reimbursement categories. The opportunity cost of not doing certain procedures for patients as the procedures could not be coded under ICD 9 came to 750 million USD over a 10-year period. (Libicki & Brahmakulam, 2004).

The second benefit was a calculation in terms of fewer miscoded claims. Coding errors are defined as the difference between actual code and the reported code. Due to the more logical sequence and standardized definition in ICD 10 codes, it is expected that the number of errors will be lower as coders became accustomed to the terms. A reduction in errors results in a reduction in erroneous claims sent back, and thus reduce the amount of work for payers. The benefits accrued to 500 million USD over a ten-year period. (Libicki & Brahmakulam, 2004).

The third benefit was in terms of fewer improper claims which stems from the fact that more specific codes would result in fewer opportunity for fraud as there are less ‘gray’ areas. This resulted in a final amount of 250 million dollars. (Libicki & Brahmakulam, 2004).

The fourth benefit is an understanding of the value of procedures. Physicians may use discharge statistics to look at which particular technique was used to treat a patient; these charts could be pulled out faster when specific codes are used. The value of procedures due to greater knowledge and efficiency resulted in benefits accruing to 800 million USD. (Libicki & Brahmakulam, 2004).

Better disease management of diabetic patients is done by identifying a diabetic patient and placing the patient into a disease management program, which uses ICD codes as key variable. It is expected that as a patient goes through a disease management program, his/her quality of life improves. The final figure, after adjustments, amounted to 1.2 billion USD. (Libicki & Brahmakulam, 2004).

Other benefits from ICD 10 implementation which the RAND report discussed are better understanding of health conditions and healthcare outcomes, enhanced ability to evaluate providers and timely intervention for emerging diseases. (Libicki & Brahmakulam, 2004).

Table 2: Cost Estimation in the RAND report; Source: The Costs and Benefits of Moving to the ICD-10 Code Sets (2004)

Summary of Estimated One-Time Costs and Cumulative Annual Costs[a]

	Personnel	Cost Estimate ($ million)	Additional Cost of Sequential Change ($ million)
Training	Full-time coders	100–150	0–20
	Part-time coders	50–150	
	Code users	25–50	0–10
	Physicians	25–100	
Productivity losses	Coders	0–150[a]	
	Physicians	50–250[a]	
System changes	Providers	50–200	5–50
	Software vendors	50–125	5–20
	Payers	100–250	5–50
	CMS	25–125	5–20

[a] Cumulative total of ten years of annual costs (undiscounted).

Table 3: Benefit estimation in the RAND report; Source: The Costs and Benefits of Moving to the ICD-10 Code Sets (2004)

Summary of Estimated Benefits over a Ten-Year Period[a]

Category	Benefit ($ million)	Largely Due to
More-accurate payment for new procedures	100–1,200	ICD-10-PCS
Fewer rejected claims	200–2,500	both
Fewer fraudulent claims	100–1,000	both
Better understanding of new procedures	100–1,500	ICD-10-PCS
Improved disease management	200–1,500	ICD-10-CM

[a] Benefits are not discounted over time.

Table 4: Comparative Cost Estimation for ICD 9 to ICD 10 transition from the RAND report and ICD-10 Business Case

COSTS ESTIMATIONS FOR TRANSITIONING FROM ICD-9 TO ICD-10	
RAND (USD $)	**ALBERTA HEALTH & CALGARY HEALTH REGION (CND $)**
Full-time Coders: Number of coders: 50 000 coders in hospitals. Number of retraining hours: 4-7 (8-hour) days Hourly wage: $50 Retraining Expenses: $500 Lost work-time: 1 work week ($2000) **Total cost of retraining: $2500 per coder**	**Alberta Wide Training Costs Total Cost: $ 1 568 915** Alberta Health approved $4.1 million Pre and post training time for trainers, health record staff, data analysts $270 per day Hourly wage: $33.75 Number of ICD-10-CA students: 1 968 Number of retraining days: 5 504 **Total Average cost of retraining: $797 per code**
Part-time Coders: 200 000 outpatient doctors' offices Number of coders: 200 000 coders in hospitals. Number of retraining hours: 4-8 Hourly wage: $50 Retraining Expenses: $500 Lost work-time: 8 hours ($400) **Total cost of retraining: $900 per coder**	**Calgary Health Region** Number of ICD-10-CA students: 468 Number of retraining days: 1077 Hourly wage: $33.75 Training facility costs: $5 850 Customized "Windows" training costs: $2 625 **Total average cost of retraining: $ 693.62 per coder**
Code Users (Employees of payer organizations) Number of coders: 150 000 Number of retraining hours: 4-8 Hourly wage: $50 Retraining Expenses: Lost work-time: 8 **Total cost of retraining: $900 per coder**	
Physicians **Total cost of retraining:**	
PRODUCTIVITY LOSSES (Considering break-in period of 1-3 months)	**PRODUCTIVITY LOSSES (short term productivity)**

Coders Time difference to code with ICD-10 from ICD-9: 1.7 minutes Number of hospital discharge patients/month: 1.8 million Wage per hour: $50 Extra coder hours required for ICD-10: 50 000 **Total of additional costs in first month: $2.5 million**	**Coding Services in Calgary Health Region** Cost of productivity on Coding Services: $392 320 Space Impact on Corporate Data: $ 69 040 **Total impact cost: $461 360**
Physicians -10% of patient records not routine and need clarification of physician -Need 5 minutes of physician time per record Number of discharges per month: 1.8 million Combined cost physicians and coders: $200/hour Additional physician hours: 20 000 **Total cost for extra work: $ 4 million**	
SYSTEM CHANGES	**SYSTEM CHANGES**
Providers Interviews from providers showed that little costs would be incurred from transition. Changeover estimated $1 per annual discharge.	
Software Vendors Information from 2 major vendors. Design to final testing is 3 years. 10% of software labour force Time of work: 1-year **Total cost of adapting software: $100 million**	
Payers Cost of adapting larger field size and alphanumeric characters, and reprogramming logic: **2/3 cost is expanding field size** **1/3 cost is adopting an updated code** Group 1 4% of market: Rely on software vendor to adapt Group 2 12% of market: Invest 40% per client member	**Calgary Health Region** **3210 person-days effort**

Group 3 3% of market: $1 per member investment **Averaged total cost $100 million-$250 million**	**Impacted IT Application Estimates: $3 735 406**
Medicare & Medicaid Services (CMS) Total contractors: 100 Cost of policy change for contractors: **Estimate: $25-$125 million**	

Source: Losev, T. (2017). Developing a Cost Projection for ICD-11 Implementation in Canada. *Unpublished manuscript.*

Appendix D: Cost estimation of ICD-9 to ICD-10 transition by Calgary Health Region

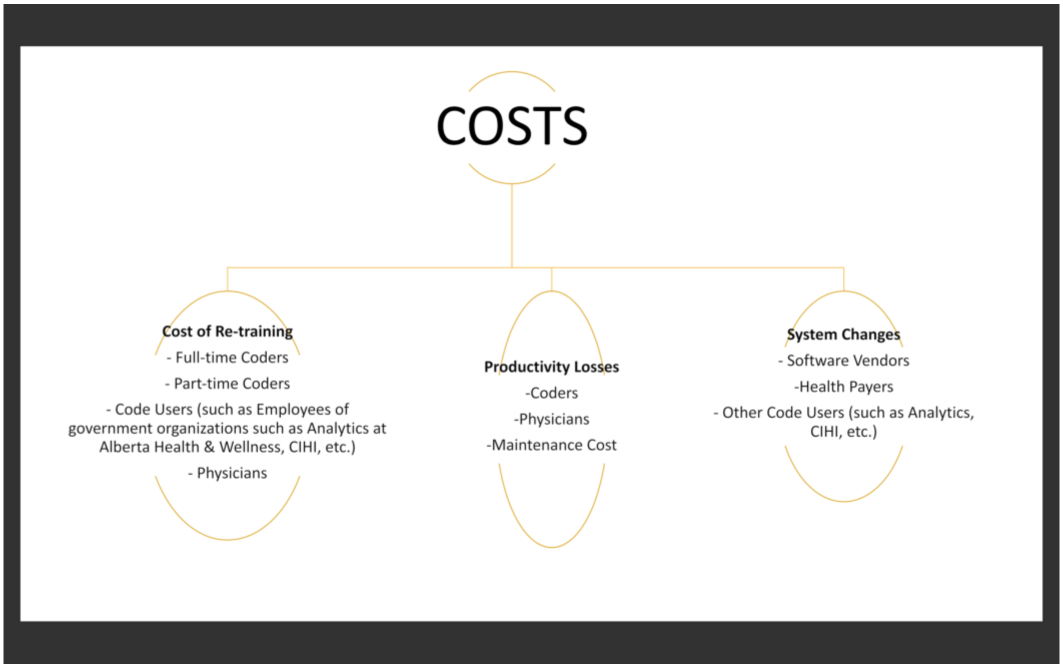

Figure 2: Calgary Health Region ICD-10 Cost Estimation Framework

Cost impact for human resource had been assumed to be a one-time cost for training health record staff, as well as the increased coding hours required to code. Training time for trainers, health record staff and data analysts were set at $270 per day, while costs for additional attendants were set at $150 for half day and $370 for two-day sessions. Costs for classroom, accommodation, travel and meals were added. (ICD Business Case, 2001).

Software vendors, Meditech, Phoenix, Cactus, Softworks and MediSolutions were contacted to provide bids for implementing the ICD 10 required software and maintenance.

The IT costs for implementing ICD-10-CA/CCI in Alberta are focused on changes to:

- Coding structure (code format and size)
- Presentation format
- Logic for report selection, summarization and management support systems
- Data storage and transmission format
- Abstracting rules

The following types of systems were assessed:

- Hospital and other health care facility abstract systems
- Encoding software
- Grouping software
- Community oriented systems, such as mental health, home care and ambulatory care
- Notifiable disease systems for sexually transmitted diseases, tuberculosis, etc.

Management support systems, including executive information systems and decision support systems are already in place or under development. (ICD Business Case, 2001).
The following tables are the cost estimation results:

Table 5: Summary of ICD-10-CA/CCI Training Requirements. Source: ICD-10-CA/CCI BUSINESS CASE. (2001, September 18)

	Student Count	**Days of Effort**	**Average Cost**	**Total Cost**
Health Record Trainers	4	60	$995	$14,500
Health Record Staff & Data Analysts	79	711	$810	$240,680
Other Users	306	306	$510	$45,900
Project Management and Administration Costs				$15,060
Training Facility Costs				$5,850
Customized "Windows" Training Costs	79	6		$2,625
GRAND TOTALS				$324,615

Table 6: Impact on Productivity and Space Requirement Cost Estimates. Source: ICD-10-CA/CCI BUSINESS CASE. (2001, September 18).

IMPACT	**COST**
Cost of productivity on Coding Services	$392,320
Space Impact on Corporate Data	$69,040
GRAND TOTAL $	$461,360

Table 7: Summary of Calgary Health Region Costs and Efforts Estimates. Source: ICD-10-CA/CCI BUSINESS CASE. (2001, September 18).

IMPACT	**COST**
Impacted IT Application Estimates	$3,735,406
Calgary Health Region Detailed Training Estimates	$324,615
Impact on Short Term Productivity & Space Impacts	$461,360
GRAND TOTAL	$4,521,381

Appendix E: Benefit Calculation Formulae Scenarios

The benefit calculation for this thesis is based in part on the benefit estimation methodology as outlined in the RAND report. For example, the benefits of disease management via ICD codes may be illustrated with a resource savings due to a reduction in the number of diabetic patients placed in disease management programs. If there is a reduction in the number of diabetic patients after disease management and control, it is expected that there would be a reduction in the amount of resource required to treat diabetic patients, and hence that amount may be used elsewhere in healthcare funding allocation. This benefit can be calculated as follows:

Formula 2: Simple formula for benefit estimation calculation for resources saved

Total Benefit (CAD $)	**Estimation Method**
$Z_{\text{patients ICD10}}$	Number of patients identified by ICD 10 coded charts and placed in disease management groups
$Z_{\text{patients ICD9}}$	Number of patients seeking diabetic treatment before ICD 10 was implemented
$Z_{\text{patient cured}}$	Number of patients showing measurable reduction in Hb1Ac levels due to disease management in one year
$Z_{\text{resource cost ICD9}}$	$(Z_{\text{patients ICD9}}\ X\ Z_{\text{resource use ICD-9 per patient}})$
$Z_{\text{resource cost ICD10}}$	$(Z_{\text{resource use ICD-10 per patient}}\ X\ (Z_{\text{patients ICD10}} - Z_{\text{patient cured}}))$
$Z_{\text{resource cost saved}}$	$(Z_{\text{patients ICD9}}\ X\ Z_{\text{resource use ICD-9 per patient}}) - (Z_{\text{resource use ICD-10 per patient}}\ X\ (Z_{\text{patients ICD10}} - Z_{\text{patient cured}}))$

X denotes the multiplication sign.

Here, Z denotes the resource cost of providing treatment to diabetic patients.

If the value for $Z_{\text{resource cost saved}}$ is positive, then we can conclude that there has been a reduction in the resources used to treat diabetic patients in the span of one year as the number of diabetic patients has fallen due to being placed in disease management groups. Unfortunately, in Canada, ICD codes are not used to aid the physicians in administering healthcare for improved patient outcomes by placing patients in disease management groups.

If physicians evaluate the ICD coded charts after a patient has left a hospital, they may identify which medicines and treatment work better for a certain patient with certain characteristics and ailments. This might result in a pattern which would make it faster and easier for physicians to provide treatment. Thus, resulting in earlier discharges of patients. This would also entail an efficient use of data analytics in healthcare.

Formula 3: Simple formula for benefit estimation calculation for amount saved in hospital resources

Total Benefit (CAD $)	Estimation Method
$D_{\text{patients ICD 9}}$	Number of patients discharged under system using ICD 9 codes
$D_{\text{patients ICD10}}$	Number of patients discharged under system using ICD 10 codes
$D_{\text{standard number days}}$	The average number of days a patient stays in a hospital
$D_{\text{days saved ICD 9}}$	$D_{\text{patients ICD 9}} - D_{\text{standard number days ICD9}}$
$D_{\text{days saved ICD 10}}$	$D_{\text{patients ICD 10}} - D_{\text{standard number days ICD10}}$
$D_{\text{increase in days saved}}$	$D_{\text{days saved ICD 10}} - D_{\text{days saved ICD 9}}$
Total amount saved in hospital resources	$D_{\text{increase in days saved}}$ X Cost of each night/patient

X denotes the multiplication sign.

Here, D denotes the number of patients discharged.

This calculation can be done using one-year discharge data under each coding system. If the total amount saved in hospital resources is positive, we can say that there has been a net benefit since patients have been discharged early due to the increased efficiency and faster treatment via reviewing of charts and codes by physicians.

An increase in ambulatory care due to faster service may be the result of better hospital management due to the use of ICD 10 codes. This is because it is expected that ICD 10 uses a better structure and includes the newest procedures, which means that paramedics can pull up information related to every procedure and provide the best treatment available based on data analytics. Hence, there would be a decrease in the backlog of cases. This may be measured using the following formula:

Formula 4: Simple formula for benefit estimation calculation for amount saved in wages due to time saved in providing treatment for each patient

Total Benefit (CAD $)	Estimation Method
$T_{\text{treatment for patient using ICD9}}$	Time in minutes required to provide treatment to patient under system using ICD 9 codes
$T_{\text{treatment for patient using ICD10}}$	Time in minutes required to provide treatment to patient under system using ICD 10 codes
T_{saved}	$T_{\text{treatment for patient using ICD9}} - T_{\text{treatment for patient using ICD10}}$
Total amount saved in remuneration	T_{saved} X paramedic's wages/hour

X denotes the multiplication sign.

Here, T denotes the time in minutes. This estimation calculation can be done over a one-year duration.

If T_{saved} is positive, and this number is multiplied with the wages paramedics receive per hour, we will end up with an estimate in terms of CAD dollars which will show the resources saved from a hospital's or the government's budget in wages. This amount may be utilized in other areas of healthcare.

Formula 5: Simple formula for benefit estimation showing the increase in number of patients treated per day

$N_{\text{number of patients treated using ICD9}}$	Number of patients treated per day under system using ICD 9 codes
$N_{\text{number of patients treated using ICD10}}$	Number of patients treated per day under system using ICD 10 codes
N_{increase}	$N_{\text{number of patients treated using ICD10}} - N_{\text{number of patients treated using ICD9}}$

X denotes the multiplication sign.

Here, N denotes the number of patients treated per day. If N_{increase} is positive, although this is not a dollar amount, we may still conclude that there is a benefit arising from the increased number of patients being given treatment on a particular day, due to better hospital management using ICD codes.

Challenges in Adopting the Benefit Estimation Methodology

One similar estimation scenario which may be adopted from the RAND model is the benefit arising from a reduction in diabetic patients due to better disease management, and hence, a saving of resources therein. The formula for calculating this benefit is explained in formula 2, above. Although this estimate is good to be thought of as benefit measure in theory, it may not be as easy to implement these measures in practice. The main impediment for estimation of gains from better disease management is the unfortunate practise of physicians not using the coded data to inform on patients' diagnoses.

Formulae 3, 4 and 5 provide benefit estimations for amount saved in hospital resources, amount saved in wages due to time saved in providing treatment for each patient and the benefit estimation showing the increase in number of patients treated per day. All three of these formulae have to be adjusted for inflation. Moreover, there could be various pitfalls related to these estimation methods; for formula 3, the number of patients discharged earlier could be due to faster recovery of patients instead of due to hospital management using ICD-10 coding system. Similarly, other factors may contribute to faster service in ambulatory care and thus cause the increase in the number of patients cared for in the time frame after ICD-10 implementation. In addition, the current Canadian healthcare system does not use ICD codes for hospital management and data analytics for healthcare delivery, which is a vital assumption for the estimation methods described above.

Hence, it is very difficult to exactly and accurately assign dollar values to benefits arising from ICD-10 implementation. Moreover, data for the variables such as number of patients treated in a day, time in minutes required to provide treatment to patient, number of patients discharged, cost of each night/patient, minutes required to code a chart using ICD-9 and ICD-10 codes, number of patients showing measurable reduction in Hb1Ac levels due to disease management, number of patients identified by ICD-10 coded charts and placed in disease management groups, are not available for the years 2002 and 2003. Hence, it is not possible to calculate the monetary benefits of ICD-10 implementation. The information for the above variables is also difficult to collect for

ICD-10 to ICD-11 transition. As a result, the only possible benefit calculation which may be completed is to look at whether there has been an increase in data quantity and quality after ICD-10 implementation.

Appendix F: Copyright Permissions

Re: Request for permission to cite report

Shahreen Khair

Tue 23-Oct-2018 8:18 AM

Sent Items

To: Kathleen Addison <[redacted];

Cc: mingshan01 <[redacted]; Mingshan Lu <[redacted]>; Christina Green [redacted]>;

Thank you.

Get Outlook for iOS

From: Kathleen Addison <[redacted]>
Sent: Tuesday, October 23, 2018 7:28:01 AM
To: Shahreen Khair
Cc: mingshan01; Mingshan Lu; Christina Green
Subject: RE: Request for permission to cite report

Good morning Shahreen,
I don't believe there is a problem in using this.
Thanks for asking.
ka

From: Shahreen Khair [[redacted]]
Sent: October 22, 2018 5:11 PM
To: Kathleen Addison <[redacted]>
Cc: mingshan01 <[redacted]>; Mingshan Lu [redacted]
Subject: Request for permission to cite report

Good Afternoon,
Hope you are well.

I am a masters student at the Department of Economics at the University of Calgary. Dr. Mingshan Lu has supervised my thesis.

I am writing to request your permission to cite the report: ICD-10-CA/CCI Business Case, in my Masters Thesis. My thesis is titled: The Economics of Big Data in Healthcare: A Case Study of the ICD-9 to ICD-10 Transition. The business case report is a very helpful resource in completing my paper and I would be grateful if I could use the cost estimates as a reference from my thesis.

I have cited the paper as follows in my thesis: *ICD-10-CA/CCI Business Case* (pp. 1-19, Rep.). (2001). Calgary, AB: Calgary Health Region.

Thank you very much.

Best regards,

Shahreen Khair

Re: Request for permission to cite papers

Shahreen Khair

Mon 22-Oct-2018 3:31 PM

Sent Items

To: Tania Losev [redacted];

Cc: mingshan01 [redacted]; Mingshan Lu <[redacted];

Thank you very much.

Get [Outlook for iOS]

From: Tania Losev
Sent: Monday, October 22, 2018 3:22:07 PM
To: Shahreen Khair
Cc: mingshan01; Mingshan Lu
Subject: Re: Request for permission to cite papers

Hi Shahreen,

It's great to hear from you! I'm so glad that my papers were helpful in your work – please go ahead and cite them. Thank you for letting me know.

Enjoy the rest of autumn.

All the best,

Tania Losev | [redacted]

I would like to acknowledge the people of the Treaty 7 region in Southern Alberta, which includes the traditional territories of the Blackfoot: Siksika, the Piikuni, the Kainai, and the Tsuut'ina, and the Stoney Nakoda: Bearspaw, Chiniki and Wesley First Nations. I would also like to note that the University of Calgary is situated on land adjacent to where the Bow River meets the Elbow River, and that the traditional Blackfoot name of this place is "Mohkinstsis" which we now call the City of Calgary. The City of Calgary is also home to Métis Nation of Alberta, Region III.

[redacted]

From: Shahreen Khair
Sent: Monday, October 22, 2018 3:17:26 PM
To: Tania Losev
Cc: mingshan01; Mingshan Lu
Subject: Request for permission to cite papers

Hello Tania,
Hope you are well.

I am writing to request your permission to cite the following two papers, in my Masters Thesis. Your papers have been very helpful resources in completing my paper.
I have cited them as follows in my thesis:

Losev, T. (2017). Developing a Cost Projection for ICD-11 Implementation in Canada. Unpublished manuscript.
Losev, T. (2017). Estimating the Dollar Value of Big Data Analytics in Healthcare: A scoping Study of Cost-Benefit Estimation of Big Data in Healthcare. Unpublished manuscript

Thank you very much.
Best Regards,
Shahreen Khair
[redacted]

[redacted]