2015-10-02

Quality of Hospital Discharge Abstract Databases over Time

Jiang, Shuonan Jason

UNIVERSITY OF CALGARY

Quality of Hospital Discharge Abstract Databases over Time

by

Shuonan Jason Jiang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

OCTOBER, 2015

Abstract

Objective: To assess the validity of hospital discharge abstract databases (DAD) in regards to patient comorbidities, with a focus on the trend of validity from 2002 to 2013.

Method: We compared patient comorbidity coding in DAD, against a reference standard of the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) database. We calculated sensitivity, specificity, positive predictive value and negative predictive value for nine patient comorbidities from 2002 to 2013.

Results: 63 483 patients were linked between DAD and APPROACH. DAD coding validity varied depending on condition. From 2002 to 2013, hypertension and diabetes had consistent levels of validity, whereas hyperlipidemia, peripheral vascular disease, cerebrovascular disease and pulmonary disease had declining levels of validity. Validity trends for malignancy, heart failure and liver disease were inconclusive.

Conclusion: The level of DAD coding validity and the trend of validity over time is strongly dependent on the condition tested.

Acknowledgements

I would like to thank my supervisor Dr. Hude Quan, for being a great mentor, and for his tireless support throughout my graduate studies. I would also like to thank my supervisory committee: Dr. Cindy Beck, Dr. Mingshan Lu, and Dr. Matthew James, for their guidance and encouragement during this research. Also, a big thank you to Ms. Danielle Southern, and Mr. Bing Li, for their help with data extraction, as well as Alberta Health Services and APPROACH for allowing me to use their data. I'd also like to express my gratitude to Mr. Mingkai Peng, and Dr. Guanmin Chen, who provided much support with programming and data analysis. I'd also like to thank Dr. Tolulope Sajobi, who provided much needed guidance for the statistical analysis. Lastly, I would like to thank the Western Regional Training Centre and the Ward of the 21st Century for their support that has aided my development as a health care researcher.

Dedication

To my loving parents, who uprooted their own lives so that I may have the opportunities I enjoy today.

Table of Contents

List of Tables

List of Figures

List of Symbols, Abbreviations and Nomenclature

Abbreviation	Definition
AHRQ	Agency for Health Care Research and Quality
APPROACH	Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease
BEEM	Best Evidence in Emergency Medicine
CEVD	Cerebrovascular Disease
CIHI	Canadian Institute for Health Information
COPD	Chronic Obstructive Pulmonary Disease
DAD	Discharge Abstract Database
EMR	Electronic Medical Record
GI	Gastrointestinal
HMDB	Hospital Morbidity Database
ICD	International Classification of Diseases
NPV	Negative Predictive Value
NSAPD	Nova Scotia Atlee Perinatal Database
OECD	Organisation for Economic Co-operation and Development
PHAC	Public Health Agency of Canada
PHN	Personal Health Number
PPV	Positive Predictive Value
PSI	Patient Safety Indicator
PVD	Peripheral Vascular Disease
WHO	World Health Organization

Chapter 1: Introduction

1.1 Overview

Chronic diseases present one of the greatest challenges faced by the Canadian health care system. It places a significant economic burden on the health care system and is the leading cause of mortality in Canada. For example, four types of chronic diseases: cancer, diabetes, cardiovascular and chronic obstructive pulmonary diseases (COPD), results in the death of an estimated 153 000 Canadians every year. This accounts for almost three-quarters of all mortality in Canada [1]. Chronic disease treatment and prevention is one of the primary concerns in the Canadian health care system [2]. Proper chronic disease surveillance is vital to the provision of health care in Canada.

A concerning trend in recent years is the increasing prevalence of many chronic diseases in Canada. For example according to data from the Public Health Agency of Canada (PHAC), between 1998 and 2008, hypertension prevalence has increased from 12.5% to 19.6% and diabetes prevalence has increased from 3.3% to 5.4% [3,4]. This type of upward trend can also be seen in a number of other chronic diseases [5]. PHAC calculated these numbers using several administrative databases, including the physician claims database and the hospital discharge abstract database (DAD) [6]. Therefore, the accuracy of these trends depend on the accuracy of these sources of data. Are these trends caused by real changes in the chronic disease prevalence of the Canadian population?

The answer lies in whether data quality has remained consistent during this period. Changes in data quality will affect the calculation of prevalence rates. This can be illustrated in a simple example. Suppose we want to calculate the prevalence of a particular condition over a

period of 2 years. Let us assume that data quality has improved from year one to year two. In year one, the data source captured 50% of cases, whereas in year two it captured 100% of cases. Now assume that the true prevalence is consistent over both years at 20%. Using this data source you would mistakenly calculate a prevalence of 10% in year one and 20% in year two. What seemed like an increase in disease prevalence was completely a result of improving data quality. Therefore, in order to accurately assess the trend of disease prevalence over time, the consistency of data quality over the study period must be known.

1.2 Study Purpose

The purpose of this study was to assess the validity and consistency of comorbidity coding in DAD between 2002 and 2013. This was done by comparing DAD against a reference standard: the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) database.

Chapter 2: Background

2.1 What are Administrative Data?

Administrative data are information that is collected through the administration of health care programs. These data are collected for three primary purposes: administration of health care services, enrollment into health insurance plans, and reimbursement for services. The main providers of administrative data are federal, provincial or state governments, and private health care insurers. Administrative databases were not originally designed for use in health research; however, they have become increasingly popular with health services researchers due to the fact that they are readily available, inexpensive to access and often contains large populations [7].

In Canada there are many administrative databases. Examples include DAD, emergency care, urgent care, drug data, etc. [7]. Despite the large number of administrative databases, they share some similar qualities. Canada has a publicly funded universal health insurance system, with the vast majority of Canadians enrolled [8]. Therefore, administrative data contains information on just about the entire population. Further, each patient has a unique personal health number (PHN), allowing patients to be easily identifiable in the database [9]. These databases also contain information over many years, allowing researchers to follow patients through time.

2.2 What is Discharge Abstract Databases (DAD)?

One important administrative database is DAD. It contains demographic and clinical information on separations from all acute care facilities. Separations include discharge, death, sign-out, and transfer. DAD was first developed in 1963 when it was first implemented in Ontario. It has since expanded to all other provinces except Québec. Québec reports data on its

inpatient separations to the Hospital Morbidity Database (HMDB), which while separate from DAD serves a very similar function. Currently, DAD contains information on approximately 75% of all inpatient separations in Canada, while the HMDB, accounts for the majority of the remaining 25% of separations in Canada [10].

DAD includes diagnostic, intervention, and patient demographic information. Information on patient diagnosis and interventions are entered by physicians into medical chart records. Following patient discharge, death, or transfer, the medical chart is reviewed by coders who translate this information into International Classification of disease (ICD) codes. This information is then sent to the Canadian Institute for Health Information (CIHI), where it is reviewed, edited, and corrected if necessary, before it is reported and published [11].

2.3 International Classification of Disease (ICD) Codes

The ICD coding system was developed by the World Health Organization (WHO). It allows the classification of diseases as well as various signs, symptoms, abnormal findings, cause of injury and/or health outcomes. The goal of the WHO in creating ICD codes was to promote international comparability in the classification and collection of health data. The first iteration of the ICD coding system was published in 1949. Since then new versions have been published periodically, with each iteration attempting to improve the accuracy and comprehensiveness of the previous version. The most current version is ICD-10, which was first published in 1994 [12,13].

Along with broad changes for each coding iteration, there are also minor regional changes or additions to the basic ICD coding system. For example the Canadian version of the

ICD-10 system is called ICD-10-CA. It includes enhancements that are relative to the Canadian health care system. For example, included in ICD-10-CA is a code for injury caused by being struck by a hockey puck [14]. ICD-10-CA was implemented in different provinces between 2001 and 2006. Implementation in Alberta occurred in 2002. Currently, all Canadian provinces have transitioned to ICD-10-CA [15].

ICD-10 codes use a six character code, containing both alphabetic and numeric characters, to represent various diseases, conditions and health outcomes. Generally speaking, the first three digits of an ICD-10 code represents a broader condition or related group of conditions, with the remaining digits representing more detailed information [16].

2.4 How is Discharge Abstract Data used?

Due to the fact that DAD is readily available, inexpensive to access and covers large populations, its use is increasingly popular among health care researchers. Further, administrative data as a whole are increasingly popular in research on health service outcomes, patient safety, cost of health care, and chronic disease surveillance [17].

DAD is very useful in health service outcome research. DAD allows researchers to investigate how the existence of certain comorbidities influence patient outcomes such as mortality, and length of stay. DAD also indicates which of the various diagnosed conditions is the primary condition or the condition most responsible for the resources use of the hospital stay. This allows researchers to investigate whether having certain comorbidities could be correlated with certain primary conditions or diagnoses [18]. DAD contains unique patient identifiers, and demographic and clinical data over many years, meaning that it is possible to follow patients

over a long period of time. DAD also contains information on what procedures a patient underwent during admission. This allows researchers to follow patients and see what happened after their procedure. It allows them to investigate outcomes to determine if they got better, worse, or stayed the same.

DAD is commonly used in patient safety research. The most important thing in patient safety research is to identify and report medical errors. Only when these errors are identifiable can the magnitude of the errors be assessed and possible solutions explored. Recently, institutions such as the Agency for Health Care Research and Quality (AHRQ) have developed tools that are able to identify cases of medical errors [19]. DAD has shown that it is quite useful for patient safety research, and its utilization should continue to increase in the future.

DAD is useful for research on the cost of health care services. In Canada, health care spending has become an increasingly large part of government budgets. In recent years health care spending has grown at a faster rate than GDP [20]. This suggests that the current trend of health care spending is unsustainable and it is important to improve the efficiency of health care services. DAD allows researchers to identify admissions that are more expensive than usual, using metrics such as length of stay and procedures done. Researchers can then examine what factors led to this higher than average cost of care. Regional differences in terms of spending can also be identified using DAD. Due to the fact that DAD covers a large area, researchers are able to see differences in spending across smaller geographical regions and attempt to identify reasons why such a difference exists. DAD allows researchers to identify potential inefficiencies in health care procedures and differences between regions [21]. Thus, it is a valuable tool for research into the cost of health care provision.

Due to the fact that DAD covers close to 100% of the total population, it is very useful in surveillance research. Diagnosed comorbidity information allows researchers to easily calculate prevalence and incidence rates at a population level. The large scope of DAD allows researchers to conduct surveillance research at a greater magnitude, and with better generalizability than would be possible in a clinical setting [22].

2.5 Impact of DAD Quality

The quality of DAD is an important issue. Poor data quality will have negative consequences for outcomes as well as quality of care research. This data phenomenon is often referred to as "Garbage in, garbage out" [23].

Outcomes research conducted with poor quality DAD has serious negative consequences on study results. For example in surveillance research, inaccurate information in DAD would lead to biased prevalence and incidence. This leads to studies making incorrect conclusions about the disease burden in a population. Further, poor and inconsistent data quality is problematic for time series analysis. If data quality is not consistent, it would be impossible to determine whether changes seen in time trend analysis are a result of real changes of disease occurrence or changes in data quality. Therefore, research into DAD quality is vital to disease surveillance. An example of this type of research can be seen in a 2009 paper by Quan et al. In this paper, researchers validated administrative data in regards to identification of hypertension across time and across regions. They performed this validation by comparing administrative data against chart review data. The authors reported measures of comparability such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and Kappa. This paper also conducted time trend analysis by comparing the validity of administrative data at two points in time: 2001

and 2004. After adjusting for possible confounding factors of age, sex, and comorbidity, the authors found no differences in the validity of administrative data in regards to hypertension between the two years [24].

DAD data quality is also vital to research into quality of care. Recently, there have been efforts to monitor and improve the quality of care by developing patient safety indicators (PSI). PSIs are a group of indicators that can be used to identify potential adverse events. PSIs were developed after literature review, analysis of ICD codes, and review by a clinical panel. PSIs can be used to identify potential causes of adverse events and where and how often they occur [25]. Thus, PSIs along with DAD are able to direct future research to improve the quality of care. Therefore, it is important that the quality of DAD be determined. PSIs are heavily reliant on ICD diagnostic codes in DAD. Errors in the ICD code may occur due to miscoding, miscommunication between physicians and coders, and/or missed diagnoses. Therefore, if a larger amount of coding error exists, PSIs will lead to incorrect conclusions as to what causes adverse events, and where they occur [19]. As a result, quality of care researchers may waste time and effort attempting to fix problems that may not exist, or are not of the scope and severity that was originally assumed.

A 2012 study by Drosler et al. determined the relationship between PSI and DAD quality. In this study the authors attempted to use PSIs to compare the health care performance of 15 Organisation for Economic Co-operation and Development (OECD) countries. They used a reduced set of U.S. AHRQ PSIs, and applied these definitions to administrative data from the various countries. One of this study's findings is that PSI rates are positively correlated with the intensity of diagnostic coding. PSI rates were higher for countries where the DAD contained a larger mean number of secondary diagnosis. The intensity of diagnostic coding varied greatly

among countries and can be considered a measure of coding thoroughness [25]. This study showed that the quality of DAD has a profound effect on PSIs.

2.6 What is APPROACH?

The APPROACH database was created in 1995. APPROACH was created due to the need for the assessment of the process and outcome of care for patients with coronary artery disease. APPROACH is designed to capture all patients undergoing cardiac catheterization in Alberta. This allows research to be conducted on the population level. APPROACH contains clinical information on 170 000 Albertans that have undergone cardiac catheterization and/or revascularisation procedures. Information is collected regarding the patient's procedure as well as his/her comorbidities. Patients are also asked to provide follow up information at 1, 3 and 5 years. This allows researchers to investigate the path and impact of patients receiving acute care [26].

A unique feature of the APPROACH database is the data collection process. After a patient is referred for a catheterization process, they are first registered into the system. They then need to provide consent to be included in the database, as well as consent for possible follow-up. Next, baseline information about the patient is collected. This is done directly by clinicians and other laboratory technicians. At this stage information such as patient comorbidities and past medical history will be collected. Following this, information regarding test and procedure results will also be directly recorded by clinicians and other laboratory personnel [27]. The data collection process in APPROACH provides two major advantages. First, data must be entered according to a consistent criteria, unlike DAD, where the documentation is variable and depends on the discretion of the physician or the coding practises

of a particular hospital. Second, the direct collection of data in the clinical setting reduces the chance of error due to data translation. Unlike DAD where coders record information from medical charts that were written by clinicians, clinicians are able to enter data directly. This removes a major potential cause of error, and is one of the main reasons for our use of APPROACH as the "reference standard".

2.7 What is Data Quality?

There are five components that constitute data quality: validity, timeliness, consistency, usability and relevance. Validity refers to the closeness of observations in data to the true values or the values generally accepted to be true. Timeliness refers to how up to date the data are at time of release. Consistency refers to the comparability of data validity for data produced at different times or in different geographical locations. Usability refers to the ease with which data can be understood and accessed. Relevance refers to how well data meets the needs of users [28]. The focus of this study will be on data validity and consistency.

2.8 What is Validity?

This study will attempt to validate DAD over time. Therefore, it is important to establish what validity means. We will consider several concepts of validity: construct validity, face validity, content validity and criterion validity [29].

2.8.1 Construct Validity

Construct validity can be defined as an experimental demonstration showing how well a test is measuring what it claims to be measuring. Thus construct validity is established by the

accumulation of evidence. In other words, construct validity is established by repeated tests or experiments that consistently deliver the same results [29]. For example, a physician may want to know whether a pain medication is effective. He/she could test this by coming up with a questionnaire asking his patients what pain level they are experiencing. He can then give this medication to his patients and compare responses to this questionnaire before and after administering this medication. If the results of the questionnaire show that patients routinely reported lower levels of pain after taking the medication, then it can be said that this test has construct validity.

2.8.2 Face Validity

Face validity refers to the extent that a test seems to be testing what it is supposed to test. In other words a test can be said to have face validity if it seems that it will likely measure what it is supposed to measure. Basically face validity is based on intuitive judgement [29]. For example if a researcher wanted to develop a checklist that would help diagnose alcoholics, they would simply assess whether the checklist seems to make sense. Basically, face validity assesses whether a test superficially appears to be valid. Face validity cannot be measured with statistical methods. It is a qualitative measure. Face validity is often considered the least scientific form of validity, due to the fact that it cannot be quantitatively measured and compared.

While face validity is often considered the weakest type of validity, it is still used in health care research. Face validity is quick and cheap to establish, and could be used as a preliminary screen for validity. If something does not have face validity, it is unlikely to be valid in other ways [30]. Face validity is commonly used as a first step to establishing practice and research guidelines. For example, in a study by Carpenter et al., the authors attempted to establish the validity of the Best Evidence in Emergency Medicine (BEEM) rater score [31]. The

BEEM rater scale is a tool that evaluates emergency medicine related studies in order to highlight studies that are most likely to influence practice guidelines. One part of this study involved establishing the face validity of the BEEM scale. Two investigators assessed related studies and compared their assessments to the result of the BEEM scale. Then based on their intuitive judgement, the researchers concluded that the BEEM scale had good face validity [31].

2.8.3 Content Validity

Content validity refers to how well a measure represents all the components of the variable to be measured. Judgment of whether all components are accounted for is based on the social construct [29]. Social construct refers to the accumulated knowledge and understanding of a current society [32]. For example based on social constructionism, we know that a car has 4 wheels, an engine, doors, gas tank, etc. This understanding of what a car is has been developed by years of technological development, and trial and error. In other words, content validity refers to whether something is correct based on what we as a society know about the subject matter. For example, suppose we are to develop a strategy for running a professional sports team. If this strategy only focuses on training, coaching, and player development, while completely ignoring the finances of running a sports franchise, we would conclude that such a strategy lacks content validity. We know this because the previous experience of successful sports franchises have shown that the financial side of sports is as important as the actual game itself. Thus content validity should be assessed by experts, whose knowledge represents what we as a society know. An example of content validity assessment in health care research can be found in a 2013 article by Patterson et al [33]. In this study researchers attempted to establish the validity of a framework for detecting adverse events in helicopter emergency medical services. Content

validity was established by a panel of 10 expert clinicians, who after careful analysis concluded that the framework had good content validity [33].

Content validity, while considered a stronger measure of validity than face validity is still a qualitative measure of validity. It is not commonly measurable using statistical methods [29].

2.8.4 Criterion Validity

Criterion validity refers to how well a measure or set of measures compares against another measure or set of measures. This type of validity is established by comparing the measure being tested against an external source assumed to be of high validity [29]. For example, assume a researcher wants to develop a new test for measuring intelligence. Let's say there exists an in-depth and comprehensive IQ test which is generally regarded as a good test of intelligence. This test, however, is costly and time consuming to administer. The new test is quicker and simpler to administer. The researcher could then compare the new test against the old test to assess criterion validity of the new test.

Criterion validity measures are quantifiable, and can be easily measured and compared. An example of criterion validity can be seen in a 2010 study by Jette et al [34]. In this study, researchers assessed the accuracy of ICD coding in DAD for epilepsy. They compared the coding data against data obtained from chart review. They found good comparability between chart review and DAD data, therefore establishing criterion validity [34]. There are a large number of statistical measures that assess Criterion validity. This is the type of validity that will be assessed in this study, where we will attempt to validate DAD by comparing it to an external source: APPROACH.

2.9 Statistical Measures for Validity

When calculating statistical measures of validity, there are a few commonly used statistics; Cohen's Kappa, and four statistical measure of diagnostic test accuracy: sensitivity, specificity, PPV and NPV.

2.9.1 Cohen's Kappa

Cohen's Kappa coefficient is a measure of agreement for two variables. Unlike a simple measure of agreement, Kappa takes into account the agreement that could simply occur by chance. In other words Kappa measures the degree of agreement beyond that of simple chance. Kappa is calculated by taking the proportion of observed agreement minus the proportion of agreement by chance divided by 1 minus the proportion of agreement by chance. This can be seen in the formula below.

Figure 1. Cohen's Kappa

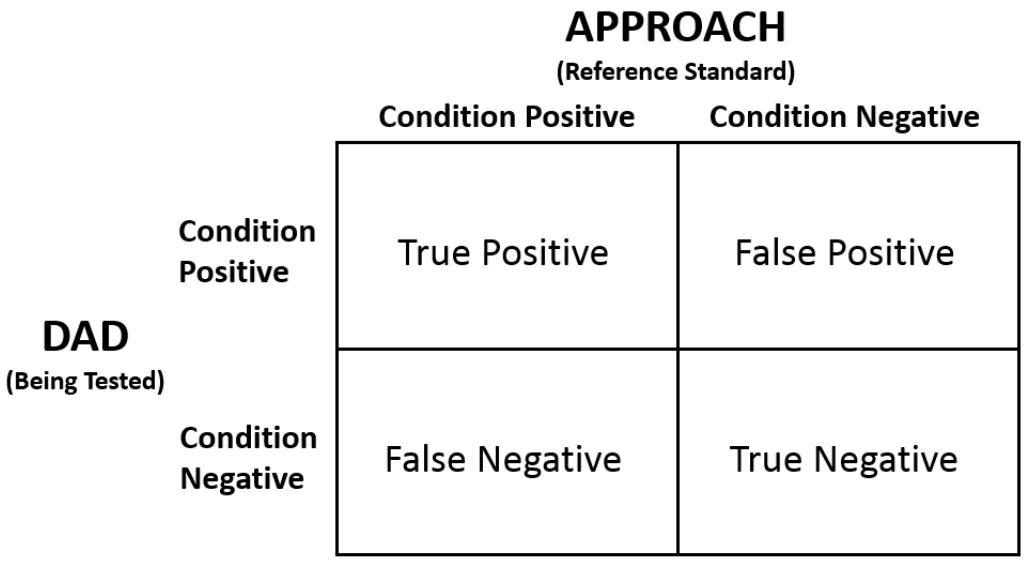
$$\kappa = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

Kappa values are measured on a scale of 0 to 1, with values close to 0 representing little or no agreement, and values close to 1 representing significant or perfect agreement [35].

2.9.2 Diagnostic Test Accuracy

When assessing the accuracy of a diagnostic test, or in our case, the validity of a source of data. It is necessary to first construct a 2x2 table. The 2x2 table compares the positive and negative values of your test data against the positive and negative values of your reference standard. Therefore, when comparing DAD against the reference standard of APPROACH, the 2x2 table will look like this.

Figure 2. Diagnostic test accuracy 2x2 table



Using the 2x2 table, we are able to calculate sensitivity, specificity, PPV and NPV

Figure 3. Sensitivity Formula

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Sensitivity refers to the ability of a test to identify positive results. It is the proportion of the number of true positives over the number of true positives plus the number of false negatives. For example if we are referring to the identification of a disease, true positives would represent cases where the test is positive given that person actually had the disease. False negatives would represent cases where the test is negative given the person actually had the disease. In other words sensitivity represents the probability of a positive test given that the patient really did have the disease [36].

Figure 4. Specificity Formula

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

Specificity refers to the ability of a test to identify negative results. It is the proportion of the number of true negatives over the number of true negatives plus the number of false positives. If we once again refer to the example of identification of disease in a patient population, true negatives would represent cases where the test is negative given the person is actually without disease. False positives would represent cases where the test is positive and the person is in reality without disease. Specificity represents the probability of a negative test given that the patient is without disease [36].

Figure 5. PPV Formula

$$\text{PPV} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

PPV is defined as the proportion of positive tests results that are true positive. It is calculated by taking the proportion of the number of true positives over the number of true positives plus the number of false positives. Referring to the example of identification of a disease, PPV would represent the probability that a test is correct at identifying cases of disease [36].

Figure 6. NPV Formula

$$\text{NPV} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}$$

NPV is defined as the proportion negative tests results that are true negatives. It is calculated by taking the proportion of the number of true negatives over the number of true negatives plus the number of false negatives. Once again using the example of disease identification, NPV would represent the probability that a test is correct at identifying cases that are absent of disease [36].

Out of these four measures, PPV and NPV are affected by the prevalence of the validated condition. Very high prevalence will artificially inflate PPV, whereas very low prevalence will artificially inflate NPV. Careful consideration of prevalence needs to be taken when interpreting PPV and NPV.

2.10 Factors that affect DAD validity

DAD validity is heavily dependent on the accuracy of the ICD diagnostic codes. There are several factors that may lead to errors in coding where the ICD codes do not reflect the information in medical charts or that of reality [37].

In order to explore the factors related to DAD validity, it is important to go over the DAD data collection process. When a patient is admitted to hospital, a doctor sees the patient and provides a diagnosis of his/her condition. This information is included in hospital charts. When the patient is discharged from hospital, this chart is sent to professional coders, where the information is transcribed to ICD codes and inputted into DAD. This data is than reviewed, and an attempt will be made to fix any missing or erroneous data based on data feature logic checks [37]. There are many points in this process where error may occur. When the patient comes in, the physician might make an incorrect or incomplete diagnosis. Next, errors may occur when

writing this information into medical charts. When the patient is discharged and the chart is sent to coders, there is the potential for miscommunication between the coder and the physician. Perhaps the coder is unable to understand the physician's handwriting, or perhaps the coder misunderstands what is diagnosed in the chart. Next the coder might miscode the diagnosis, due to improper or incomplete knowledge of the ICD coding system [38]. As we see there are many potential causes for error during the DAD collection process.

2.10.1 Misdiagnosis

In order for a patient's comorbidity information to be recorded, that patient first needs to see and be diagnosed by a physician. If the physician misdiagnose the patient, the information collected in DAD will be inaccurate. This is the first requirement for DAD validity, a misdiagnosis if left uncorrected ensures that information collected in all subsequent steps are inaccurate. A great deal of research has investigated how common misdiagnoses occurs. For example, Arthur Elstein, a cognitive psychologist, researched clinical decision making and concluded that diagnostic errors occur 10-15% of the time [39]. However, a review study by Graber found that the rate of diagnostic error differs depending on the estimation method used [40]. For example, autopsy studies found that major diagnoses discrepancies occurred in 10 – 20% of cases. On the other hand, a study that looked at voluntary reports of diagnostic error found that only 0.5% of cases were reported for diagnostic error [40]. Therefore, there is still no definitive answer as to the true rate of misdiagnosis. However, assuming that Elstein is correct and that misdiagnosis occurs in 10 – 15% of cases, then comorbidity information will contain at a minimum that same amount of error, even before other potential causes of error are considered.

2.10.2 Physician Documentation

The next step in the data collection process is the recording of information to medical

records. These records are also sometimes referred to as physician documentation. This documentation is used to code the information in DAD. Thus the quality of physician documentation is important to the quality of DAD. Physician documentation quality is affected by inaccurate or incomplete recording of information as well as the legibility of physician handwriting. Inaccurate or incomplete recording reduces the quality of information in medical charts, while illegible handwriting reduces the ability of others to use or translate that information. There are several causes of poor physician documentation. First, little to no education is provided to physicians regarding the importance or proper method to document. Second, physicians are often under time limitations, and documentation is regarded as low priority. Third, physicians are often required to complete multiple documents containing similar information [41]. For example, a 2006 report from Ontario described the current state of physician documentation quality. It stated that some hospitals require physicians to document medications on both face sheets and discharge summaries. As a result physicians often are not trained to document properly, do not have enough time to do it, and deal with an overly complex documentation system [42]. This leads to poor documentation. Poor documentation also means improper coding in DAD, reducing DAD accuracy.

2.10.3 Coder Training

Once physician diagnosis is recorded on medical charts, these charts are translated into ICD codes by coders. The ability of these coders to understand the conditions that are diagnosed and translate them into the appropriate ICD codes are vital to the accuracy. If a coder misinterprets what diagnosis is recorded in medical charts, or if the coder assigns the wrong ICD code, the validity of DAD will be diminished. Research has shown that the amount of coder training as well as coder workload has a direct effect on DAD accuracy [37]. Coders in regions

that receive more intensive training perform better than coders in regions that receive less intensive training. Further, heavier coder workload requirements results in greater amounts of coding error [37].

2.10.4 Patient Condition/Severity

The condition that a patient presents and the severity of such conditions affects the accuracy of comorbidity information in DAD. Research has shown that primary conditions are coded better than secondary conditions [43]. Primary condition in this case refers to either the main reason for the patient's admission, or the condition responsible for the largest amount of resource use. The fact that primary conditions were coded better makes sense, because treatment of the primary condition is the most important part of a hospital admission, and is central to the delivery of care. Research has also shown that the more severe a condition is the better coded it is [43]. This is caused by the same phenomenon that results in primary conditions being coded better. Conditions with greater severity requires more intensive care, has more attention paid to it, and as a result are coded better.

2.11 Data Consistency

Consistency refers to whether an entity conforms to certain standards or is similar to another entity. In regards to data quality, consistency refers to how well data compares between regions and across different points in time.

2.11.1 Geographic/Region/Hospital Location

An important factor in the consistency of DAD validity is geographical location. The accuracy of DAD varies greatly between locations. For example, ICD coding procedures often

vary from country to country. A study by Quan et al. showed that the criteria for coding the primary condition differ from country to country [44]. These differences can have significant implications for the comparability of research conducted using such data. In addition, there is also a high possibility that differences exist in coder education, coder experience, patient conditions, clinician workload etc. In fact, some of the earliest research using administrative data found regional differences in the provision of medical care [17,45]. These types of regional differences are likely to result in differences in DAD validity. Therefore, the consistency of DAD validity must be considered when comparing DAD research across geographical regions, or hospital locations.

2.11.2 Time

DAD accuracy also seems to be affected by time. We theorize that since the wide spread use of DAD in the 1980s, DAD coding accuracy has been improving. This is a result of greater emphasis in coding training, and experience gained over time. We also theorize that there will be a dip in DAD accuracy whenever there is a transition from one version of the ICD coding system to the next. The transitioning to the newer coding version requires a learning curve for clinicians and coders. Studies that examined this phenomenon have not found conclusive results. Some research shows a dip in validity for some conditions but not others [46,47]. The focus of this study is the consistency of DAD validity from 2002 to 2013. Therefore, a literature review was conducted to find research studying administrative data accuracy over time.

Chapter 3: Literature Review

3.1 Literature Review

A systematic review was conducted to find studies which examined the consistency of DAD quality over time. A previously validated search strategy for the identification of studies using administrative data was used [48]. This search strategy has high validity when identifying research using administrative data from any country. These results were then combined with keywords and MESH headings related to the following categories:

- Quality/Accuracy
- Time/Trend
- ICD coding

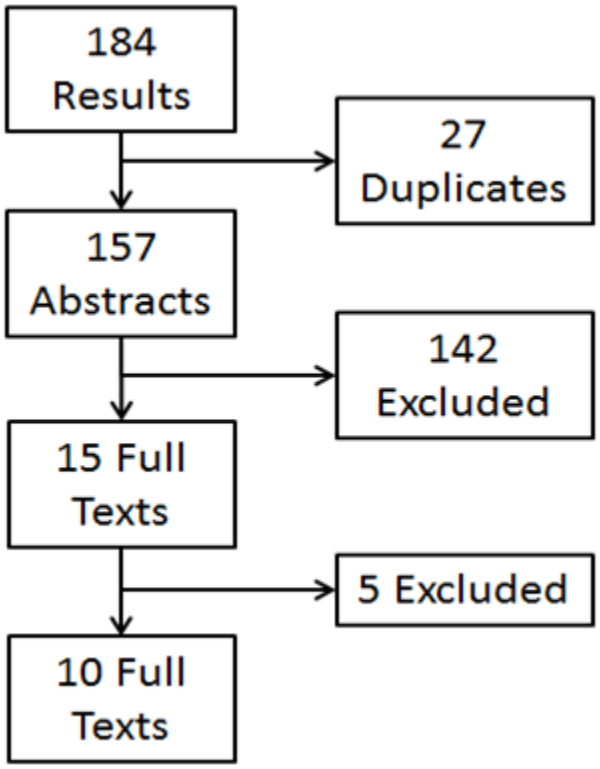
Keywords and MESH headings were combined in the search using the "OR" operator within each category. The search results from each category was then combined using the "AND" operator. Search results were screened at both the abstract and full text stage. The inclusion criteria was:

- Studies assessing the validity of DAD in regards to ICD coding
- Studies looking at consistency of DAD validity over time

3.2 Literature Review Results

A flow chart of the search process is included below.

Figure 7. Literature Review Flowchart



A summary of the ten studies found in the literature review is shown below.

Table 1. Literature Review Results

Author	Year	Country	Data Source (n)	Gold Standard	Comparison type	Comparison groups	Sensitivity	Specificity	PPV	NPV
Allen	2012	Canada	NSAPD, DAD, Phycisian services data (41 533)	NSAPD	Time	before Apr 1, 1997	85.7 (76.0-92.1)	99.9 (99.8-99.99)	71.3 (61.3-79.6)	99.9 (99.9-100)
						Apr 1, 1997 or later	87.9 (80.3-93.0)	99.7 (99.7-99.8)	63.4 (55.4-70.7)	99.9 (99.9-100)
Chen	2010	Canada	DAD (3362)	Chart review	Time	2001	90.5	96.9	78.8	99.3
						2004	91.7	98.2	83.2	99.2
Daneshvar	2012	Canada	DAD (637)	Chart review	Time	Please see chart				
De Coster	2008	Canada	DAD (4008)	Chart review	ICD	ICD-10/CCI	Please see chart			
						ICD-9 CM				
Hagen	2009	Norway	Discharge data (1080)	Chart review	ICD	ICD-8	1	0	0.775	
						ICD-9	0.906	0.319	0.347	
						ICD-10	0.838	0.972	0.88	
Januel	2011	Switzerland	Discharge data (3499)	Chart review	Time	1999	Please see chart			
						2001				
						2003				
Jette	2010	Canada	DAD (486)	Chart review	ICD	ICD-9	Please see chart			
						ICD-10				
Quan	2008	Canada	DAD (4008)	Chart review	ICD	ICD-9 CM	Please see chart			
						ICD-10				
Quan	2009	Canada	DAD (3362)	Chart review	Time	2001	77	95	79	94
						2004	74	94	83	92
Reid	2012	Canada	DAD, ER, physician claims (720)	Chart review	Time	2003	Please see chart			
						2006				

Of the ten studies, eight assessed DAD quality in terms of comorbidities, while two assessed quality in terms of procedures. The focus of this present study is DAD quality in regards to comorbidities. Therefore, the studies by Daneshvar et al. and De Coster et al., which assessed procedures will be excluded from further analysis [49,50]. The remaining eight studies were all "gold standard" comparisons. A "gold standard" comparison is a type of study where the accuracy of a source of information is compared to another source that is deemed to be true. Of the eight studies, seven used medical chart review as the "gold standard". This is where researchers pull information from the medical charts of the participants and use it to assess the accuracy of DAD. The exception was the Allen et al. study which used a clinical registry as the "gold standard" [51].

Most studies assessed consistency of DAD accuracy over a small number of time periods for a single condition. For example, a study by Chen et al. assessed the accuracy of DAD coding for Diabetes comparing 2001 to 2004 [52]. Two studies: Quan et al. and Januel et al., also contained a small number of time comparisons, but both included multiple conditions [46,53].

A 2012 study by Allen et al. validated a set of administrative data case definitions for diabetes mellitus in a population of pregnant women. The Nova Scotia Atlee Perinatal Database (NSAPD) was used as a "gold standard". The NSAPD was used as a "gold standard" because it is a clinical database containing information on a population level. Part of the results of this study compared the validity of the administrative data case definitions in two time periods: before April 1, 1997 and April 1, 1997 or later. Comparison of these two time periods show that sensitivity increased slightly from 85.7% in the earlier period to 87.9% in the later period, while PPV decreased from 71.3% in the early period to 63.4% in the later period. However, neither of these differences were statistically significant. Further specificity dropped slightly from 99.9% to

99.7% and NPV was consistent at 99.9%. Specificity and NPV also did not display any statistically significant differences. This study provides little evidence that the accuracy of administrative data case definitions differed between the two periods [51].

A 2010 study by Chen et al. validated administrative data coding algorithms for diabetes mellitus. Patients from 48 urban and 16 rural general practitioners' clinics in Alberta and British Columbia were included in the study. The coding algorithm was validated using a "gold standard" methodology, where chart review data were used as the "gold standard". The validity of this coding algorithm was compared for two time periods: 2001 and 2004. After adjusting for age, sex, and comorbid conditions, the authors concluded that validity of this administrative data coding algorithm was not statistically significantly different between 2001 and 2004 [52].

The 2009 Hagen et al. study validated DAD in regards to traumatic spinal cord injuries. Chart review was once again used as the "gold standard". This study compared DAD accuracy in regards to three ICD versions: ICD-8, ICD-9, and ICD-10. This was not a direct time comparison, but given the fact that different ICD iterations were used at different times, we can treat this comparison as a de facto time comparison. Study results show that accuracy was overall poor for all three ICD versions. Sensitivity decreased from ICD-8 to ICD-10, while specificity increased. PPV on the other hand fluctuated during the period. Therefore, it seems that the accuracy of DAD did not follow a clear trend between ICD-8 and ICD-10 [54].

A 2011 study by Januel et al. investigated the accuracy of DAD comorbidity coding in regards to multiple conditions. Chart review was used as the "gold standard". This study assessed DAD accuracy at three points in time: 1999, 2001, and 2003. It found that the trend of accuracy over time is heavily dependent on the condition. Accuracy improved for some conditions, and declined or stayed the same for others. Overall, however, the authors concluded that the average

sensitivity of all comorbidities increased slightly. Further sensitivity increased for 30 co-morbidities, while sensitivity decreased for 6. However of the 30 comorbid conditions where sensitivity increased, this increase was only significant for 6. Of the 6 comorbid conditions where sensitivity decreased, only 1 was statistically significant. Thus these results suggests that DAD accuracy increased slightly from 1999 to 2003 [53].

A 2010 study by Jette et al. looked at the accuracy of DAD for the identification of epilepsy. Chart review was used as the "gold standard". This study compared validity between ICD-9 and ICD-10. This study found that coding accuracy for epilepsy was similar in ICD-9 and ICD-10, with PPVs of 98.9% for ICD-9 and 98.6% for ICD-10. This suggests that there was not a significant difference in DAD accuracy between the two ICD coding versions [34].

A 2008 study by Quan et al. assessed the accuracy of DAD for multiple comorbidities. This study compared DAD accuracy between ICD-9 and ICD-10, and used chart review data as the "gold standard". Overall the authors concluded that DAD accuracy was similar between the two versions. For example, sensitivity was found to be significantly higher in ICD-10 for one condition, and significantly lower for seven conditions relative to ICD-9. For the remaining 24 conditions sensitivity was similar between ICD-9 and ICD-10. This suggests that for the majority of conditions studied, there is no statistically significant difference between the accuracy of DAD using ICD-9 and ICD-10 [46].

The next study found in our review is a 2009 study by Quan et al. This study investigated the accuracy of case definitions for hypertension. This study contains the same population as the 2010 study by Chen et al. However, in this case the accuracy of coding was assessed for hypertension rather than diabetes mellitus. Two time periods were analyzed: 2001 and 2004. Chart review was again selected as the "gold standard". The authors concluded that after

adjustment for age, sex and comorbid conditions, the accuracy of DAD coding in regards to hypertension is similar between 2001 and 2004 [24].

The last study found in this review is a 2012 study by Reid et al. It examined the accuracy of DAD coding definitions in regards to epilepsy. Chart review was used as the "gold standard", and it compared the fiscal year of 2003 and the fiscal year of 2006. Overall the authors reported that accuracy of coding definitions were high in both time periods, and that accuracy was not statistically significantly different between the two periods [55].

Overall, the results of this review are inconclusive. There does not appear to be a clear trend in regards to DAD accuracy in either the direct time comparisons or the ICD version comparisons. However, it does seem that DAD coding accuracy is heavily dependent on condition. Studies that investigated multiple comorbidities often found that DAD accuracy increased for some conditions, while the accuracy of other conditions decreased or stayed the same [53,46].

Several authors listed ambiguity regarding the accuracy of the "gold standard" as a limitation of their studies. The most common "gold standard" used was chart review data. In theory data obtained from medical charts should be very accurate. However, errors in the medical chart or errors made during the chart review process could potentially make chart review data an improper "gold standard". Further, there is no standardized definition or criteria for the assessment of a comorbidity via chart review. The problem is that it is simply not possible to know the actual amount of error in medical charts, even though logically error in medical charts should be minimal. Therefore, it was necessary for researchers to assume that medical charts are accurate and appropriate for use as a "gold standard". Other limitations that authors included are

small sample size, and poor generalizability in studies where the study population was obtained from a specific region or group of patients.

The majority of studies compared the accuracy of DAD over a small number of time periods. There is a lack of research comparing accuracy of DAD over longer time scales. Further most of these studies dealt with a single or small number of conditions or procedures.

After a review of literature, there seems to be some gaps in currently completed research:

1. Lack of studies comparing a larger number of comorbidities or procedures: most studies dealt with a single condition.
2. Lack of studies containing a long-term larger range and frequency of comparisons: most studies compared two or three points or periods of time, or ICD versions.

Chapter 4: Objectives and Methods

4.1 Objectives

4.1.1 Objective 1

To assess the trend of DAD validity in regards to patient comorbidities from 2002 to 2013, in Alberta.

4.1.2 Objective 2

To assess factors that may influence the trend of DAD validity over time.

4.2 Methods Overview

The validity of DAD in regards to patient comorbidities was assessed using a "gold standard" comparison. Data from APPROACH was used as the "gold standard". The two databases was linked using PHN, and agreement between the two databases was assessed. Validity was measured for each comorbidity recorded in APPROACH. DAD validity for each comorbidity was assessed on a year to year basis, in order to investigate the trend of validity over time. Next, external factors: age, sex, and hospital location, were examined to determine whether they confound these trends. Confounding occurs when an external factor is correlated with both the dependent and independent variables in a statistical model. An emphasis of this study was to include a larger number of comorbidities compared over a longer time frame.

4.3 Study Population

This study includes data on all inpatients in the APPROACH database. The population was limited to inpatients only, because DAD only includes data on patients that have been admitted to acute care institutions. Further, all patients in this study have known PHN. This makes it possible to link these patients between APPROACH and DAD.

4.4 Data Sources

4.4.1 APPROACH

APPROACH contains information on all patients receiving cardiac catheterization in Alberta. APPROACH contain information on 14 comorbidities. Comorbidities are recorded during baseline evaluation by clinicians. Comorbid information are recorded as simple yes/no variables in the database. These comorbidities are:

- Hypertension
- Hyperlipidemia
- Diabetes type 1
- Diabetes type 2
- Heart failure
- Peripheral vascular disease
- Cerebrovascular disease
- Pulmonary disease
- Malignancy
- Liver disease
- Gastrointestinal disease
- Renal insufficiency
- Chronic renal failure
- Acute renal failure

This study includes APPROACH data from 2002 to 2013. Data prior to 2002 was not used for this study as a result of Alberta's transition from ICD-9 to ICD-10 in 2001. Thus by using 2002 as a cut-off we are able to deal exclusively with ICD-10 codes.

A potential issue with APPROACH data is that missing comorbid data are at times entered from DAD during the verification process. This presents a problem during our validation

process, because in such cases we would effectively be attempting to validate DAD against itself. We obtained the raw APPROACH data that did not contain enhancement from DAD. Therefore, we were able to eliminate missing APPROACH comorbid data from our analysis.

4.4.2 DAD

DAD contains information on all patient hospitalisations in Alberta. When patients are discharged, information is transcribed from hospital charts to DAD, by professional coders.

DAD captures information on patient demographics, comorbidities and procedures. Comorbidities are recorded using ICD-10 codes. There are 25 variables for each admission under which ICD codes may be entered. This allows up to 25 comorbidities to be entered for each admission. These variables are labeled DIAG1 to DIAG25. A previously established coding definition will be used to translate the ICD codes into the various conditions [56]. The coding definition used is shown below.

Table 2. ICD-10 Coding Definition

CONDITION	ICD-10 CODE
CEREBROVASCULAR DISEASE	G45.x, G46.x, H34.0, I60.x-I69.x
PULMONARY (COPD)	I27.8, I27.9, J40.x-J47.x, J60.x-J67,x, J68.4, J70.1, J70.3
HEART FAILURE	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0
PVD	I70.x-I71.x, I73.1-I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
LIVER DISEASE	B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4, I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5-K76.7
MALIGNANCY	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C77.x-C80.x, C81.x-C85.x, C88.x, C90.x-C97.x
HYPERTENSION	I10.x, I11.x-I13.x, I15.x
HYPERLIPIDEMIA	E78.0-E78.5

DIABETES	E10.x, E13.10, E13.12, E14.10, E14.12, E11.x, E13.0, E14.0

In addition to these comorbidities, DAD contains other information that are vital to this study:

- Age
- Sex
- Admission/Discharge date
- Hospital location

The admission and discharge dates allow DAD admissions to link to the correct APPROACH visit. The other three variables will be used when assessing possible confounders on the trend of DAD validity

4.5 Variables

4.5.1 Comorbidities

The most important variables for analysis are information on patient comorbidities. These are recorded in the variables DIAG1 – DIAG25 in DAD. These variables contain diagnostic codes of comorbidities in ICD-10 format. A previously established coding definition was used to translate these codes into a simple yes/no format for each patient comorbidity. This allows comorbidity information to be comparable to that in APPROACH.

APPROACH contains information on 14 patient comorbidities. Each comorbidity variable in APPROACH is coded as follows:

- 1 – the patient has the condition
- 0 – the patient does not have the condition

- 2 – the condition status for this patient is unknown
- . – No value was entered in the data

All values that were missing or entered as 2 were recoded as 0. This is based on the assumption that if a patient had a particular comorbidity, the clinician would have entered it as such, and in most cases missing values and values of 2 are actually indications of absence of the condition. A check was performed by comparing data, where missing values were coded as 0, against data where missing values were excluded. The results showed that the difference in the prevalence of the conditions and the trend of validity were insignificant. This provides supporting evidence that missing values and values of 2 for the most part could be classified as 0.

4.5.2 Excluded Comorbidities

Three comorbidities: renal insufficiency, chronic renal failure, and acute renal failure had very high rates of missing data in APPROACH, where approximately 50% of data were missing. Therefore, these three conditions were excluded from analysis. For the remaining 11 variables, type 1 and type 2 diabetes were combined into a single diabetes variable. Further, gastrointestinal (GI) disease did not have a direct match in the ICD coding definition. The coding definition included ICD codes for peptic ulcer disease, which is only one form of GI disease. However, the APPROACH variable represented all GI diseases. Thus, we did not have a validated coding definition for GI disease that could be used to translate ICD codes in DAD. Therefore we decided to exclude GI disease from further analysis. This left us with 9 comorbidities suitable for validation:

- Hypertension
- Hyperlipidemia
- Diabetes
- Heart failure
- Peripheral vascular disease
- Cerebrovascular disease
- Pulmonary disease
- Malignancy
- Liver disease

4.5.3 Demographic Information/Confounding factors

Basic patient demographic information such as PHN allow the databases to be linked. Dates of admission/discharge and dates of catheterization allows analysis of time trends. Further, variables such as age, sex, and hospital location were assessed as possible confounding factors in regards to the time trend relationship.

4.6 Analytic Plan

4.6.1 Linkage

Data from patients' APPROACH visit were linked to their DAD admissions using their PHN. All observations with missing PHNs were excluded from analysis. Given that this study included a large amount of data from 2002 to 2013, occasionally one APPROACH visit would match with multiple DAD admissions. Therefore, it was necessary to link an APPROACH visit with only one of those admissions. This was done based on the following criteria:

1. The date of the APPROACH visit falls within the admission period. In other words, the date of the APPROACH visit is between the admission and discharge date. Or, the APPROACH visit is on either admission or discharge date.
2. Occasionally, one APPROACH visit will still be matched with two admissions. This occurs when the discharge date of the earlier admission falls on the admission date of the second admission, and both dates also fall on the date of the APPROACH visit. In this situation, we linked the APPROACH visit with the earlier admission instead of the later admission.

This linked each inpatient APPROACH visit with a single admission. The data is now ready for comparison.

4.6.2 Data Analysis

All data analysis was done using SAS 9.3. SAS is an analytical software that is well suited to working with large databases.

The first step in data analysis was to translate the ICD codes from DAD into yes/no variables representing the existence of comorbidity. This was done using a previously established ICD coding definition [56]. Using SAS software, the program searched the 25 diagnostic variables for ICD codes that are in the coding definition. If a relevant ICD code was found in any of the 25 diagnostic variables, then that DAD admission was marked down as diagnosed with that condition. This translation process allowed the comorbidity information in DAD to be comparable with the comorbidity information in APPROACH.

Next, DAD quality and its consistency over time was assessed. Each linked APPROACH visit and DAD admission pair was placed into clusters of individual years. This was based on the date of the APPROACH visits. Therefore, we have clusters for all APPROACH visits in 2002, 2003, 2004 … etc. Then we compared the comorbidity information in DAD against the comorbidity information in APPROACH. This gave us measures of comparability such as sensitivity, specificity, PPV and NPV for each year. These statistical measures represent how well the information in DAD matches up with the information in APPROACH. This process was repeated for all 9 comorbidities, for each year in the study period.

Time trend analysis was performed using ordinary least squares (OLS) regression with an auto correlated error. This allowed us to assess whether sensitivity, specificity, PPV, and NPV

were changing from year to year for all 9 conditions. We chose this method because OLS regression is the cleanest and simplest way to assess change of a single univariate variable. We needed to use an auto correlated error, because normal OLS regression assumes that errors are independent. This assumption does not hold up when dealing with time trend analysis. Using this type of regression model, we tested whether the slope for sensitivity, specificity, PPV and NPV was statistically significantly different from zero. A slope estimate that is statistically significantly different from zero would indicate that the outcome measure is changing over time. For this statistical test we chose a significance level of 0.05. This method was repeated for all 9 conditions.

The next step was to assess possible confounding factors such as age, sex, and hospital location. The idea was to see if these factors are significantly different between each cluster, and assess whether these factors are significantly affecting outcomes. Stratified analysis on the time trend of data validation was performed for all 9 conditions.

4.7 Ethics Approval

Ethics approval was obtained from the Conjoint Health Research Ethics Board at the University of Calgary.

Chapter 5: Results

5.1 Study Population Characteristics

Overall, 147 854 APPROACH visits were recorded between 2002 and 2013. Of these, 63 483 patients were linked between DAD and APPROACH. The average age of patients in the total study population was 63.3 and 70.2% were male. These age and sex distributions remained mostly consistent over time. In 2002, the average age was 63.8 and 69.9% were male. In 2013, the average age was 62.6 and 70.5% were male. In terms of hospital location, 45.2% of patients came from Calgary area hospitals, 44.7% came from Edmonton area hospitals and 10.1% came from elsewhere in Alberta. While the ratio of patients from elsewhere in Alberta remained mostly constant over the study period, there was a bit of a shift in the ratio of patients from Calgary and Edmonton. In 2002, 48.1% of patients came from Calgary, whereas 42.6% came from Edmonton. In 2013, 41.2% of patients came from Calgary, whereas 47.2% of patients came from Edmonton. We are unsure what caused this split. We theorize that it could be just a result of the linkage. Changes in the proportion of missing PHN could be what is causing this difference.

5.2 Validation Results

Overall, data validity differed greatly depending on the comorbidity being tested. Hypertension and diabetes displayed high validity, while heart failure displayed moderate validity. The rest of the comorbidities analyzed showed poor to very poor validity. The time trend of validity differed significantly depending on the condition. We will discuss the trend of validity for each of the 9 comorbidities included in this study.

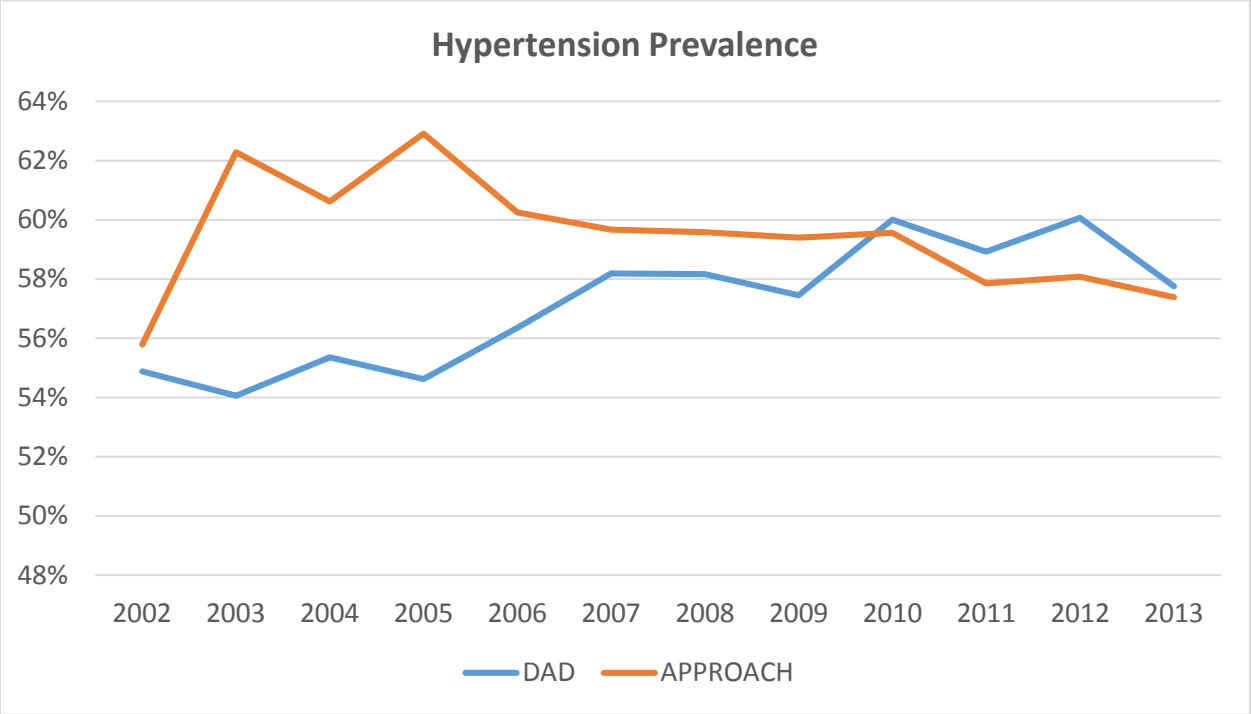
5.2.1 Hypertension

Hypertension prevalence ranged from 55.8% to 62.9% in APPROACH, and ranged from 54.1% to 60.1% in DAD. For the most part there were no significant differences between prevalence in DAD and APPROACH. However up until 2010, DAD prevalence was consistently lower than APPROACH. From 2010 to 2013, prevalence in the two databases was very similar, with minor fluctuations from year to year.

Table 3. Hypertension Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	55.8	54.9
2003	62.3	54.1
2004	60.6	55.4
2005	62.9	54.6
2006	60.3	56.3
2007	59.7	58.2
2008	59.6	58.2
2009	59.4	57.5
2010	59.6	60.0
2011	57.8	58.9
2012	58.1	60.1
2013	57.4	57.8

Figure 8. Hypertension Prevalence



DAD coding for hypertension showed high validity in most years. Sensitivity, specificity, PPV and NPV all ranged from 65.5% to 87.0%. When looking at time trends, we find no statistically significant trends for sensitivity and specificity. We found statistically significant trends for specificity with an estimate of -0.0094 per year, and for PPV with an estimate of -0.0061. This means that specificity is decreasing by 0.94% a year, while PPV is decreasing by 0.61% a year.

Table 4. Hypertension Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	82.2, (79.9-84.5)	79.6, (76.9-82.3)	83.6, (81.3-85.8)	78.0, (75.3-80.7)
2003	74.6, (72.5-76.7)	79.8, (77.3-82.3)	85.9, (84.1-87.7)	65.5, (62.9-68.2)
2004	77.1, (75.1-79.1)	78.1, (75.6-80.6)	84.4, (82.6-86.3)	68.9, (66.3-71.5)
2005	75.5, (73.5-77.6)	80.9, (78.4-83.3)	87.0, (85.3-88.7)	66.1, (63.4-68.7)
2006	79.5, (77.5-81.5)	78.8, (76.3-81.3)	85.0, (83.2-86.9)	71.7, (69.1-74.4)
2007	78.5, (76.4-80.6)	71.9, (69.1-74.6)	80.5, (78.4-82.5)	69.3, (66.5-72.1)
2008	80.0, (77.9-82.0)	74.0, (71.3-76.7)	81.9, (79.9-83.9)	71.5, (68.7-74.3)
2009	79.3, (77.2-81.4)	74.5, (71.7-77.2)	81.9, (79.9-84.0)	71.1, (68.3-73.8)
2010	81.8, (79.8-83.7)	72.1, (69.3-74.8)	81.2, (79.2-83.1)	72.9, (70.1-75.6)
2011	80.8, (78.8-82.9)	71.2, (68.4-73.9)	79.4, (77.3-81.4)	73.0, (70.3-75.7)
2012	82.5, (80.6-84.4)	71.0, (68.3-73.7)	79.8, (77.8-81.7)	74.5, (71.9-77.1)
2013	79.8, (77.8-81.8)	71.9, (69.3-74.5)	79.3, (77.3-81.3)	72.5, (70.0-75.1)

Figure 9. Hypertension Validity

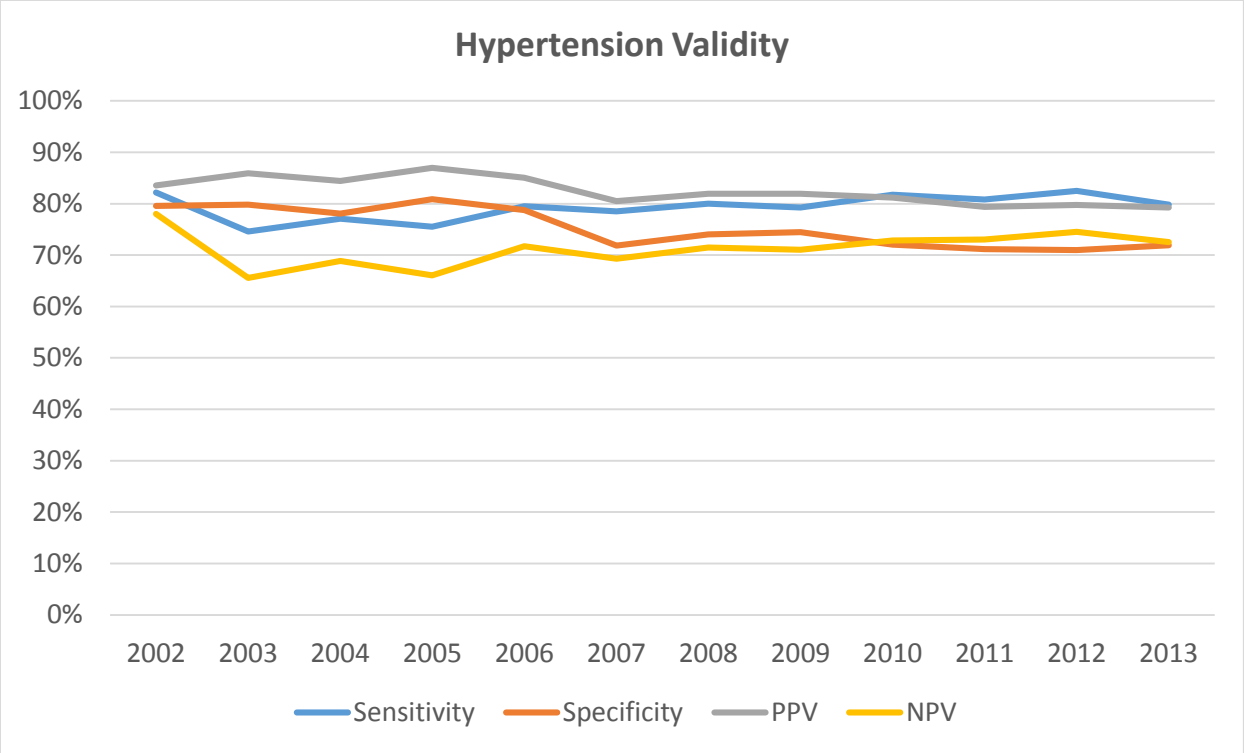


Table 5. Linear regression for DAD validity trend in hypertension coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
Hypertension	Sensitivity	0.0036	0.0019	1.89	0.0874
	Specificity	-0.0094	0.0016	-5.84	0.0002
	PPV	-0.0061	0.0013	-4.74	0.0008
	NPV	0.0029	0.0029	1.01	0.3363

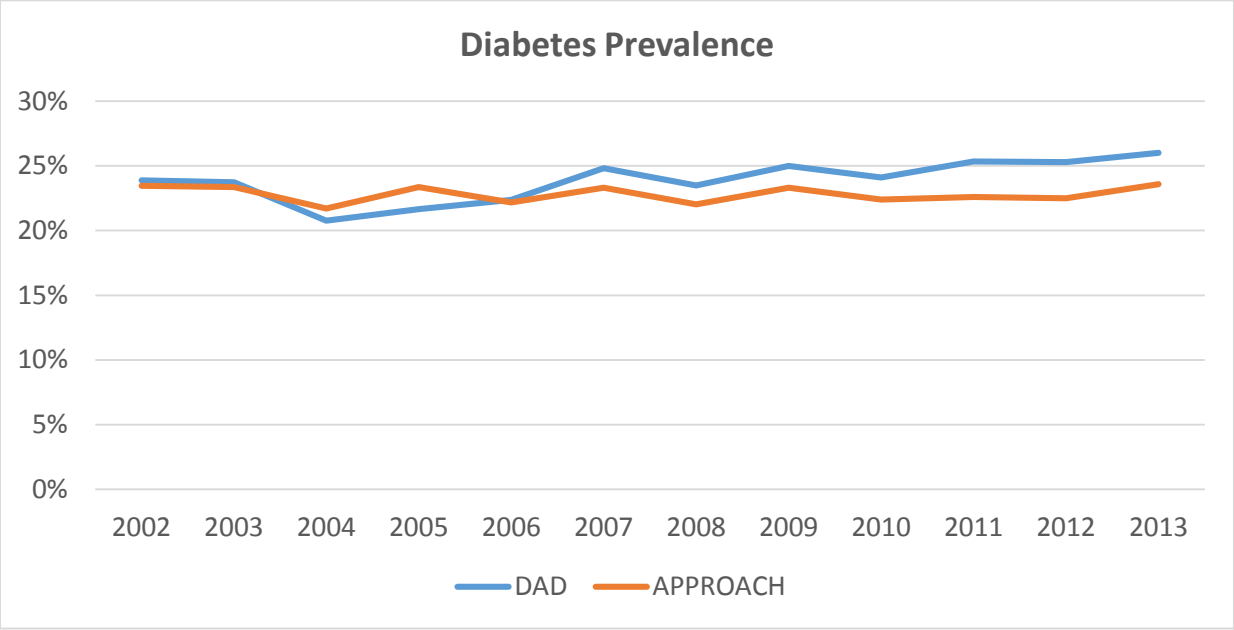
5.2.2 Diabetes

Prevalence of diabetes is very similar between APPROACH and DAD throughout the study period. Prevalence of diabetes in APPROACH is virtually constant, while prevalence of diabetes in DAD shows a slight upward trend from 2004 onwards.

Table 6. Diabetes Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	23.5	23.9
2003	23.4	23.7
2004	21.7	20.8
2005	23.4	21.7
2006	22.2	22.4
2007	23.3	24.8
2008	22.0	23.5
2009	23.3	25.0
2010	22.4	24.1
2011	22.6	25.4
2012	22.5	25.3
2013	23.6	26.0

Figure 10. Diabetes Prevalence



The coding for diabetes in DAD also has high validity. Sensitivity, specificity, PPV and NPV all ranged from 81.5% to 97.6%. This indicates that diabetes coding in DAD is accurate. Time trend analysis shows that there are no statistically significant changes in sensitivity and NPV over time. However, we do see statistically significant trends for specificity and PPV. Specificity had an estimate of -0.0030 and PPV had an estimate of -0.0076. This shows that specificity is decreasing by 0.3% a year, while PPV is decreasing by 0.76% a year.

Table 7. Diabetes Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	92.1, (89.7-94.6)	97.0, (96.2-97.9)	90.5, (87.8-93.1)	97.6, (96.8-98.4)
2003	89.6, (87.2-92.0)	96.3, (95.5-97.2)	88.2, (85.6-90.7)	96.8, (96.0-97.6)
2004	83.8, (80.8-86.7)	96.7, (95.9-97.4)	87.5, (84.7-90.2)	95.5, (94.7-96.4)
2005	81.5, (78.5-84.5)	96.6, (95.8-97.4)	87.9, (85.3-90.6)	94.5, (93.5-95.4)
2006	88.9, (86.4-91.5)	96.6, (95.8-97.4)	88.1, (85.5-90.7)	96.8, (96.1-97.6)
2007	91.1, (88.8-93.5)	95.4, (94.4-96.3)	85.6, (82.9-88.4)	97.3, (96.5-98.0)
2008	90.8, (88.4-93.3)	95.5, (94.6-96.5)	85.2, (82.3-88.1)	97.4, (96.6-98.1)
2009	90.7, (88.3-93.1)	95.0, (94.0-96.0)	84.6, (81.8-87.5)	97.1, (96.3-97.9)
2010	91.8, (89.5-94.0)	95.4, (94.5-96.4)	85.3, (82.5-88.1)	97.6, (96.9-98.3)
2011	91.4, (89.1-93.7)	93.9, (92.9-95.0)	81.5, (78.5-84.5)	97.4, (96.7-98.1)
2012	91.6, (89.4-93.8)	93.9, (92.9-95.0)	81.4, (78.5-84.3)	97.5, (96.8-98.2)
2013	90.7, (88.5-93.0)	94.0, (93.0-95.0)	82.3, (79.5-85.1)	97.0, (96.3-97.8)

Figure 11. Diabetes Validity

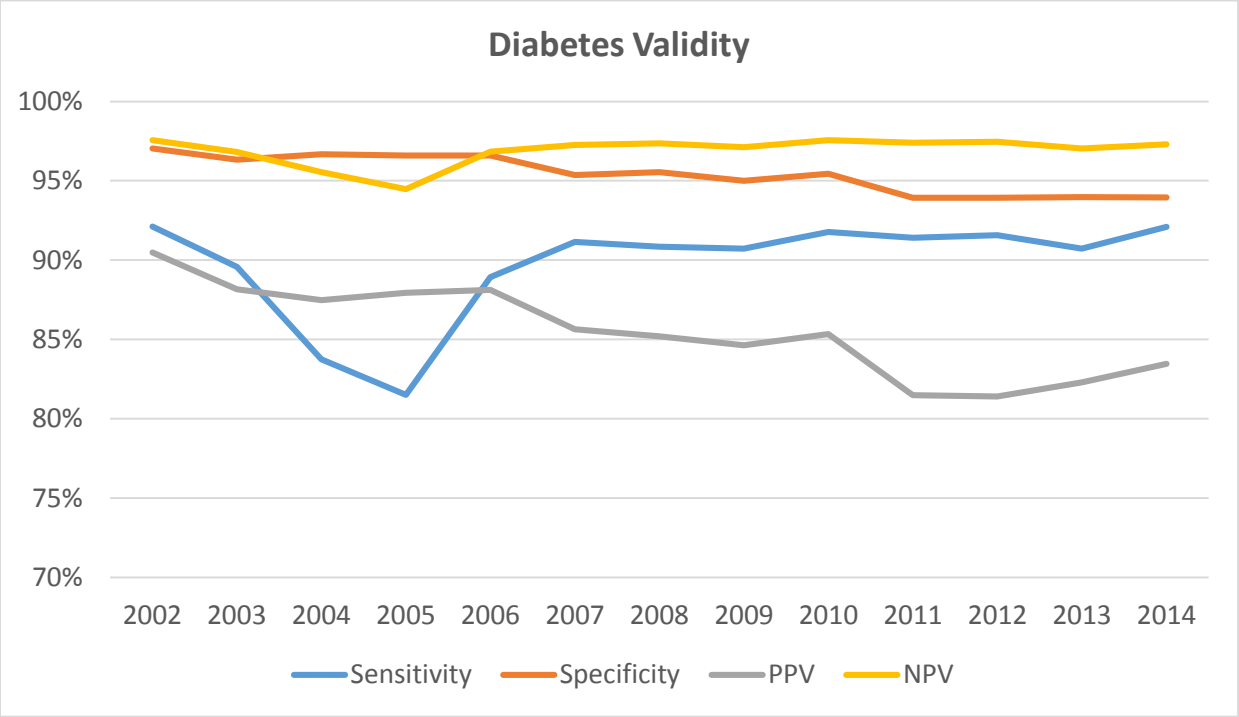


Table 8. Linear regression for DAD validity trend in diabetes coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
Diabetes	Sensitivity	0.0039	0.0027	1.47	0.1720
	Specificity	-0.0030	0.0003	-8.51	<.0001
	PPV	-0.0076	0.0008	-9.01	<.0001
	NPV	0.0010	0.0008	1.37	0.1994

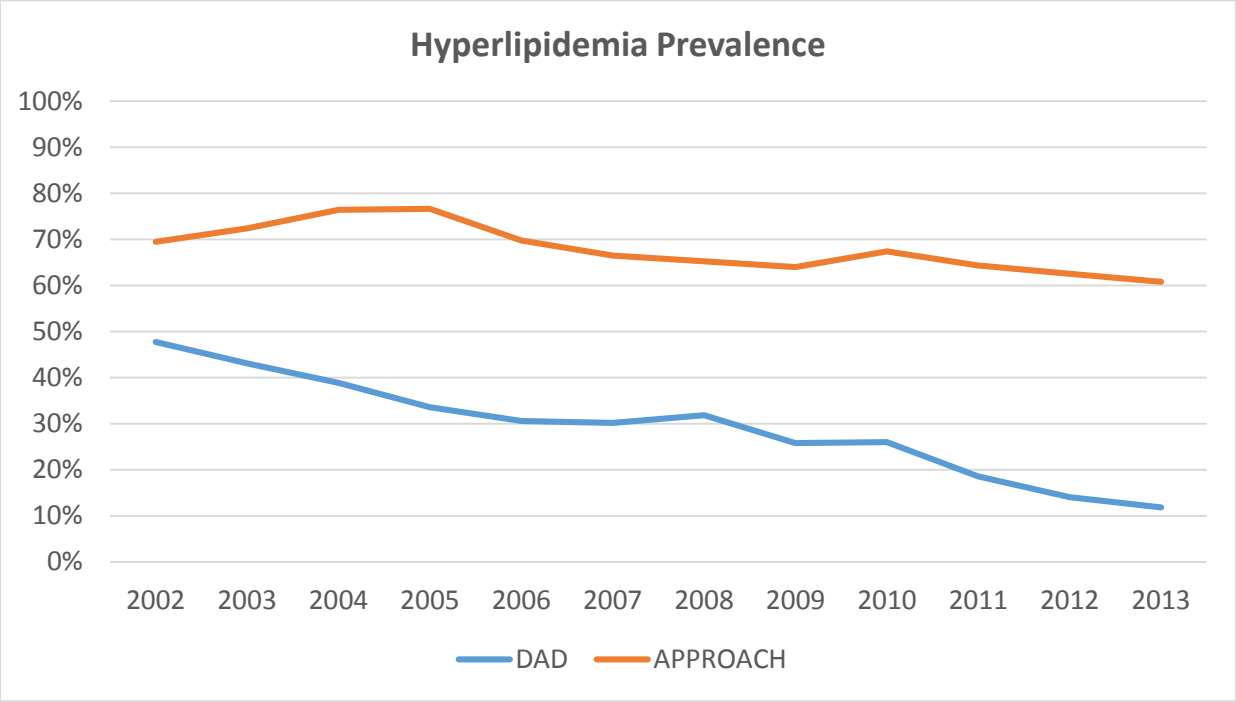
5.2.3 Hyperlipidemia

Hyperlipidemia prevalence differed greatly between DAD and APPROACH. DAD prevalence was consistently lower than APPROACH prevalence during the study period. Hyperlipidemia prevalence in APPROACH slightly decreased from 69.5% to 60.8% during the study period, whereas hyperlipidemia prevalence in DAD decreased dramatically from 47.8% to 11.8%. If we assume the prevalence from APPROACH is accurate, the large difference in hyperlipidemia prevalence between DAD and APPROACH suggests the possibility that the condition was under coded in DAD.

Table 9. Hyperlipidemia Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	69.5	47.8
2003	72.4	43.1
2004	76.5	38.9
2005	76.6	33.6
2006	69.8	30.6
2007	66.5	30.2
2008	65.3	31.8
2009	64.0	25.8
2010	67.4	26.0
2011	64.4	18.6
2012	62.5	14.1
2013	60.8	11.8

Figure 12. Hyperlipidemia Prevalence



The results from the validation analysis confirm that hyperlipidemia was under coded in DAD. While specificity and PPV were fairly high, ranging from 75.8% to 93.5%. Sensitivity and NPV were both very low. Sensitivity ranged from 15.3% to 58.1%, while NPV ranged from 29.5% to 44.3%. These results show that hyperlipidemia was severely under coded in DAD. When we looked at the trend of validity, we find no statistically significant difference over time for NPV. However, we see that sensitivity, specificity and PPV all show statistically significant change over time. The estimates are -0.034, 0.014, and -0.009 for sensitivity, specificity and PPV respectively. These estimates indicate that sensitivity is decreasing by 3.4% a year, specificity is increasing by 1.4% a year and PPV is decreasing by 0.9% a year.

Table 10. Hyperlipidemia Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	58.1, (55.5-60.8)	75.8, (72.3-79.2)	84.5, (82.2-86.8)	44.3, (41.2-47.3)
2003	51.5, (49.3-53.8)	79.0, (76.1-82.0)	86.6, (84.6-88.6)	38.3, (35.9-40.8)
2004	43.7, (41.5-45.8)	76.6, (73.3-79.9)	85.8, (83.7-87.9)	29.5, (27.3-31.7)
2005	39.3, (37.2-41.4)	85.1, (82.3-87.9)	89.6, (87.7-91.6)	29.9, (27.8-32.0)
2006	37.2, (34.9-39.4)	84.5, (81.9-87.0)	84.7, (82.2-87.2)	36.8, (34.6-39.0)
2007	35.8, (33.5-38.1)	80.9, (78.2-83.6)	78.8, (75.9-81.7)	38.8, (36.6-41.1)
2008	38.1, (35.8-40.5)	80.0, (77.3-82.7)	78.2, (75.3-81.1)	40.7, (38.4-43.1)
2009	31.7, (29.4-34.0)	84.6, (82.2-87.0)	78.6, (75.4-81.8)	41.0, (38.7-43.3)
2010	31.3, (29.1-33.5)	84.8, (82.3-87.3)	81.0, (78.0-84.0)	37.4, (35.2-39.6)
2011	23.1, (21.0-25.1)	89.5, (87.5-91.5)	79.8, (76.2-83.4)	39.2, (37.1-41.3)
2012	17.3, (15.5-19.1)	91.3, (89.6-93.1)	76.9, (72.7-81.2)	39.8, (37.8-41.8)
2013	15.3, (13.5-17.0)	93.5, (92.0-95.0)	78.5, (74.0-83.0)	41.6, (39.6-43.5)

Figure 13. Hyperlipidemia Validity

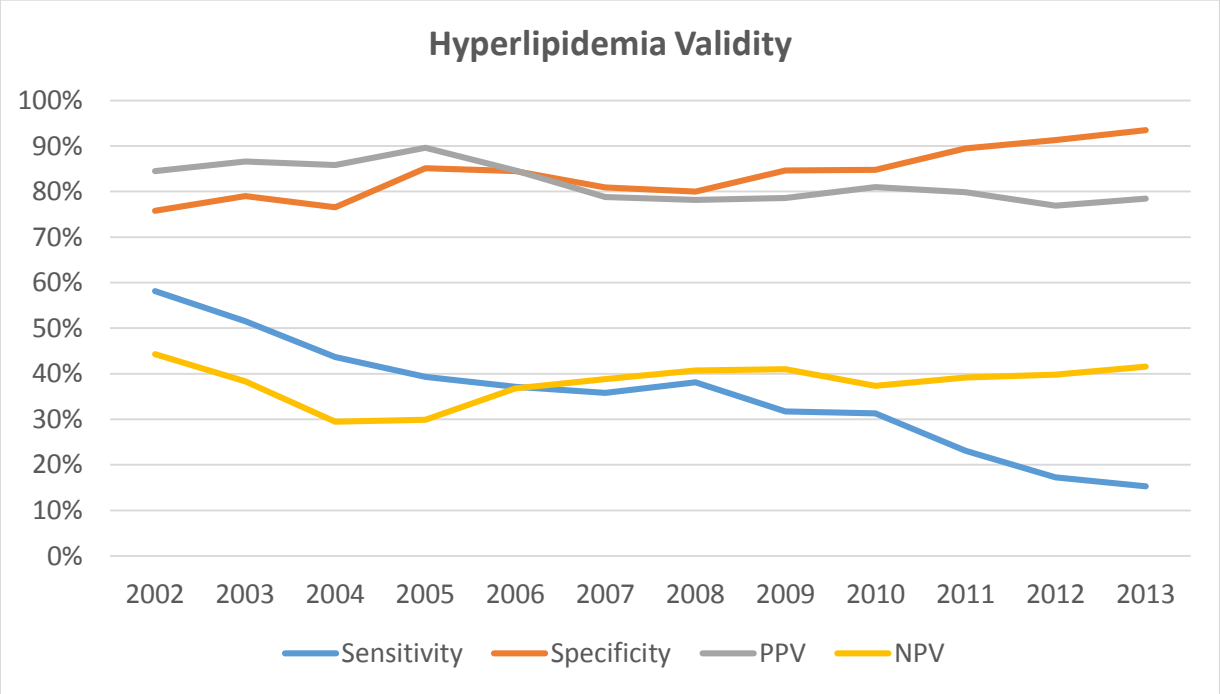


Table 11. Linear regression for DAD validity trend in hyperlipidemia coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
Hyperlipidemia	Sensitivity	-0.0342	0.0028	-12.07	<.0001
	Specificity	0.0138	0.0023	5.90	0.0002
	PPV	-0.0090	0.0022	-4.02	0.0024
	NPV	0.0036	0.0037	0.97	0.3526

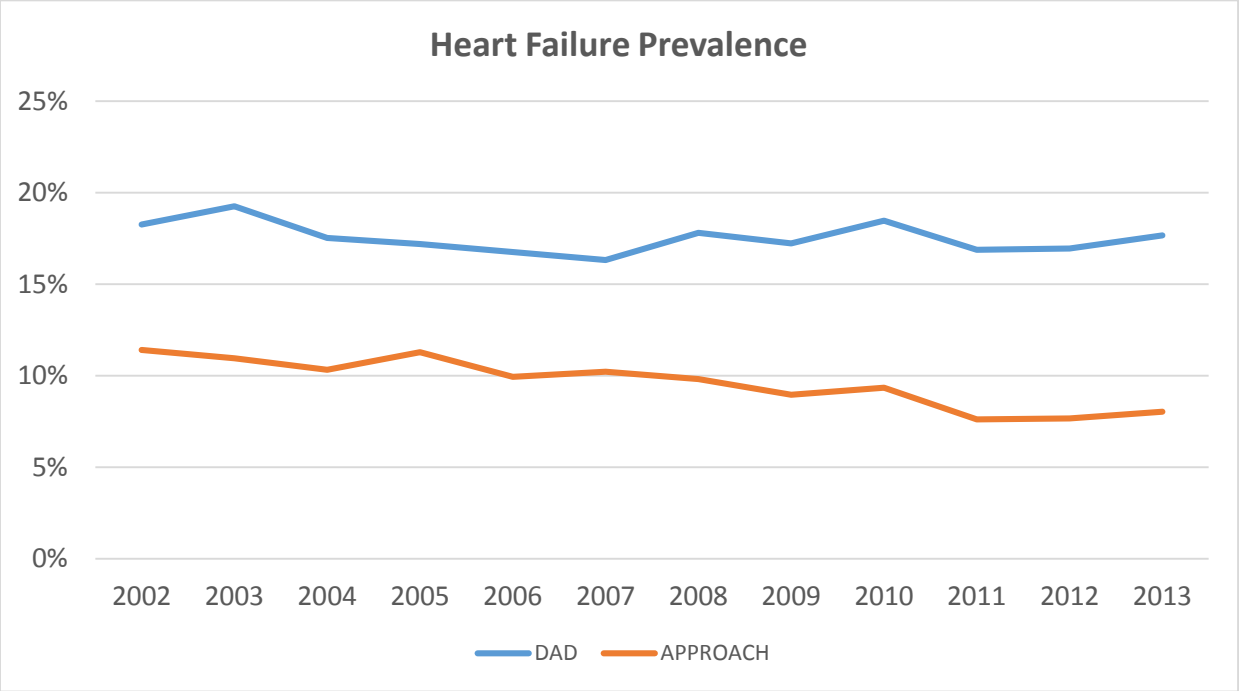
5.2.4 Heart Failure

Heart failure prevalence was consistently lower in APPROACH compared with DAD. Prevalence in APPROACH ranged from 8.0% to 11.4%, while prevalence in DAD ranged from 16.3% to 19.3%. Prevalence in DAD remained relatively consistent during the study period, while prevalence in APPROACH decreased slightly from 11.4% in 2002 to 8.0% in 2013.

Table 12. Heart Failure Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	11.4	18.3
2003	11.0	19.3
2004	10.3	17.5
2005	11.3	17.2
2006	9.9	16.8
2007	10.2	16.3
2008	9.8	17.8
2009	9.0	17.2
2010	9.4	18.5
2011	7.6	16.9
2012	7.7	16.9
2013	8.0	17.7

Figure 14. Heart Failure Prevalence



DAD validity for heart failure coding was poor. While sensitivity, specificity and NPV were all fairly good; ranging from 67.9% to 97.7%; PPV was low, only ranging from 33.4% to 47.7%. This means that out of all the patients that were coded with heart failure in DAD, only 47.7% or less of them actually had heart failure, based on APPROACH as the reference standard. Time trend analysis found no statistically significant change for sensitivity over time. However specificity, PPV and NPV all displayed statistically significant time trends. Estimates were -0.0015, -0.012, and 0.0012 for specificity, PPV and NPV respectively. This means that specificity is decreasing by 0.15% each year, PPV is decreasing by 1.2%, and NPV is increasing by 0.12% each year.

Table 13. Heart Failure Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	74.6, (68.9-80.3)	89.0, (87.5-90.5)	46.6, (41.5-51.8)	96.5, (95.5-97.4)
2003	77.9, (73.1-82.7)	87.9, (86.6-89.3)	44.3, (40.0-48.6)	97.0, (96.3-97.7)
2004	71.3, (66.0-76.6)	88.6, (87.4-89.9)	42.0, (37.5-46.4)	96.4, (95.6-97.2)
2005	72.6, (67.6-77.6)	89.9, (88.7-91.1)	47.7, (43.2-52.2)	96.3, (95.5-97.0)
2006	75.7, (70.5-81.0)	89.7, (88.5-91.0)	44.9, (40.2-49.6)	97.1, (96.4-97.8)
2007	67.9, (62.1-73.6)	89.5, (88.3-90.8)	42.5, (37.7-47.3)	96.1, (95.2-96.9)
2008	74.6, (69.1-80.1)	88.4, (87.0-89.7)	41.1, (36.5-45.8)	97.0, (96.2-97.7)
2009	74.0, (68.2-79.9)	88.4, (87.0-89.7)	38.5, (33.8-43.2)	97.2, (96.5-97.9)
2010	78.8, (73.6-84.1)	87.7, (86.4-89.1)	39.9, (35.5-44.4)	97.6, (96.9-98.2)
2011	74.0, (67.8-80.2)	87.8, (86.5-89.1)	33.4, (28.9-37.8)	97.6, (97.0-98.3)
2012	72.0, (65.8-78.1)	87.6, (86.3-88.9)	32.6, (28.3-36.9)	97.4, (96.8-98.1)
2013	76.4, (70.8-82.0)	87.5, (86.2-88.8)	34.8, (30.5-39.0)	97.7, (97.1-98.3)

Figure 15. Heart Failure Validity

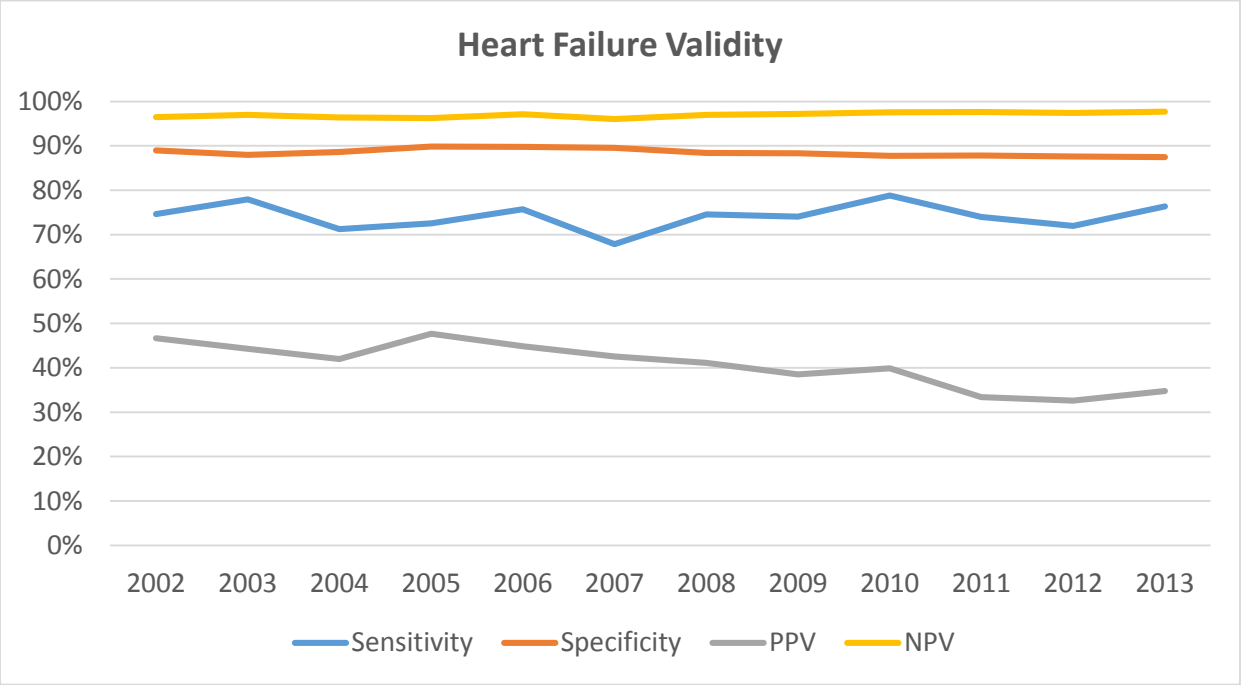


Table 14. Linear regression for DAD validity trend in heart failure coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
Heart Failure	Sensitivity	0.0006	0.0026	0.24	0.8161
	Specificity	-0.0014	0.0006	-2.46	0.0336
	PPV	-0.0124	0.0020	-6.12	0.0001
	NPV	0.0012	0.0003	3.62	0.0047

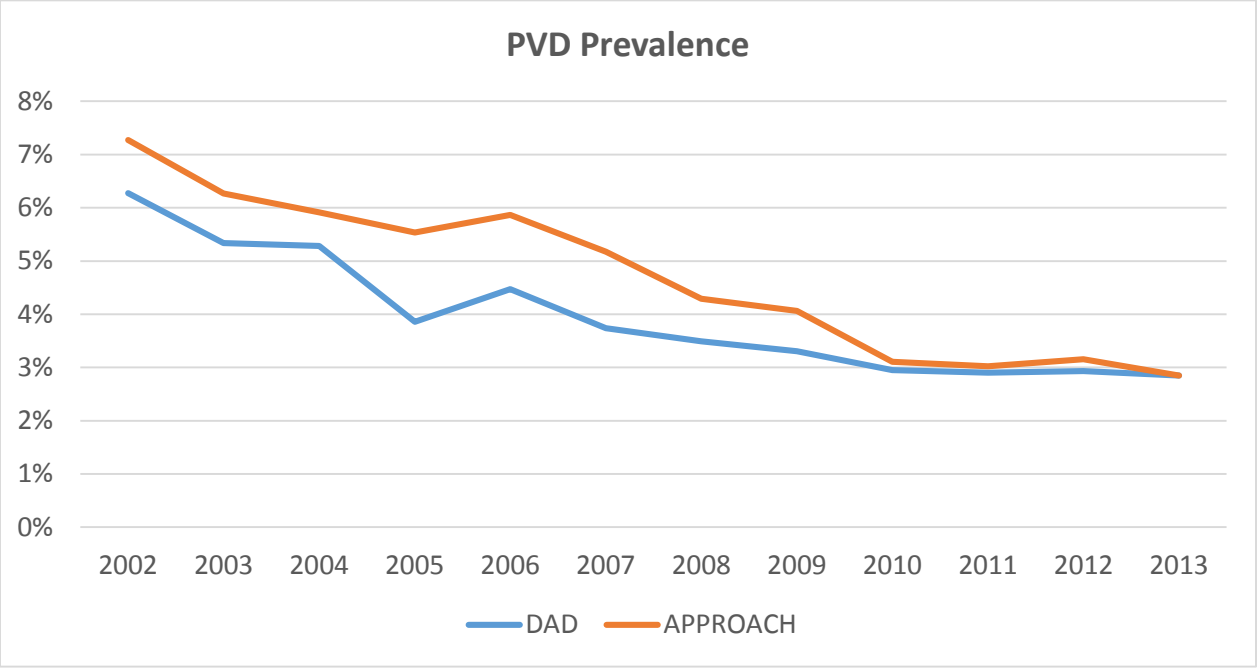
5.2.5 Peripheral Vascular Disease (PVD)

The prevalence of PVD in APPROACH was slightly higher than the prevalence of PVD in DAD. However, the prevalence in both databases decreased over time.

Table 15. PVD Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	7.3	6.3
2003	6.3	5.3
2004	5.9	5.3
2005	5.5	3.9
2006	5.9	4.5
2007	5.2	3.7
2008	4.3	3.5
2009	4.1	3.3
2010	3.1	3.0
2011	3.0	2.9
2012	3.2	2.9
2013	2.8	2.8

Figure 16. PVD Prevalence



While the prevalence of the two databases were fairly close, the validation was poor. Specificity and NPV were fairly high, however, both sensitivity and PPV were low. This means that there were a lot of cases of PVD that were missed by DAD, and out of those that were coded as PVD, a large proportion were coded incorrectly. We saw statistically significant time trends for all four measures of validity. The estimates were -0.019, 0.001, -0.026, 0.0021 for sensitivity, specificity, PPV and NPV respectively. These estimates show that sensitivity is decreasing by 1.9% a year, specificity was increasing by 0.1%, PPV was decreasing by 2.6% and NPV was increasing by 0.21% each year.

Table 16. PVD Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	41.2, (33.1-49.4)	96.5, (95.6-97.3)	47.8, (39.0-56.6)	95.4, (94.5-96.4)
2003	37.7, (30.3-45.1)	96.8, (96.1-97.5)	44.3, (36.0-52.5)	95.9, (95.1-96.7)
2004	38.2, (30.7-45.8)	96.8, (96.1-97.5)	42.8, (34.7-50.9)	96.1, (95.4-96.9)
2005	28.7, (21.4-35.9)	97.6, (97.0-98.2)	41.1, (31.7-50.5)	95.9, (95.1-96.7)
2006	33.4, (25.9-41.0)	97.3, (96.7-98.0)	43.9, (34.8-52.9)	95.9, (95.1-96.7)
2007	27.8, (20.0-35.5)	97.6, (97.0-98.2)	38.4, (28.5-48.3)	96.1, (95.3-96.9)
2008	26.0, (17.6-34.5)	97.5, (96.9-98.1)	32.0, (22.1-41.9)	96.7, (96.0-97.4)
2009	24.1, (15.6-32.7)	97.6, (96.9-98.2)	29.6, (19.6-39.7)	96.8, (96.1-97.5)
2010	24.4, (14.8-33.9)	97.7, (97.1-98.3)	25.7, (15.7-35.6)	97.6, (97.0-98.2)
2011	21.0, (11.8-30.2)	97.7, (97.1-98.3)	21.9, (12.4-31.3)	97.5, (96.9-98.2)
2012	19.7, (11.1-28.2)	97.6, (97.0-98.2)	21.1, (12.1-30.1)	97.4, (96.8-98.0)
2013	23.3, (13.8-32.7)	97.7, (97.2-98.3)	23.3, (13.9-32.7)	97.7, (97.2-98.3)

Figure 17. PVD Validity

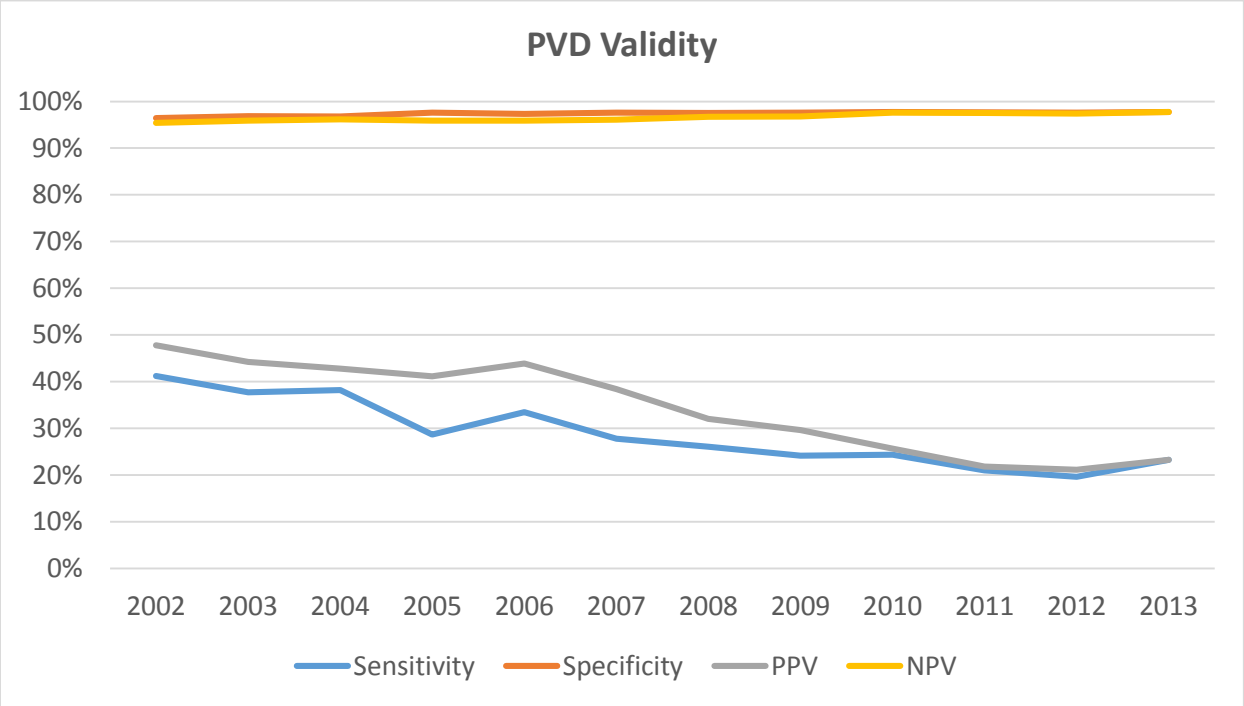


Table 17. Linear regression for DAD validity trend in PVD coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
PVD	Sensitivity	-0.0186	0.0023	-8.11	<.0001
	Specificity	0.0010	0.0002	4.92	0.0006
	PPV	-0.0263	0.0023	-11.51	<.0001
	NPV	0.0021	0.0002	9.41	<.0001

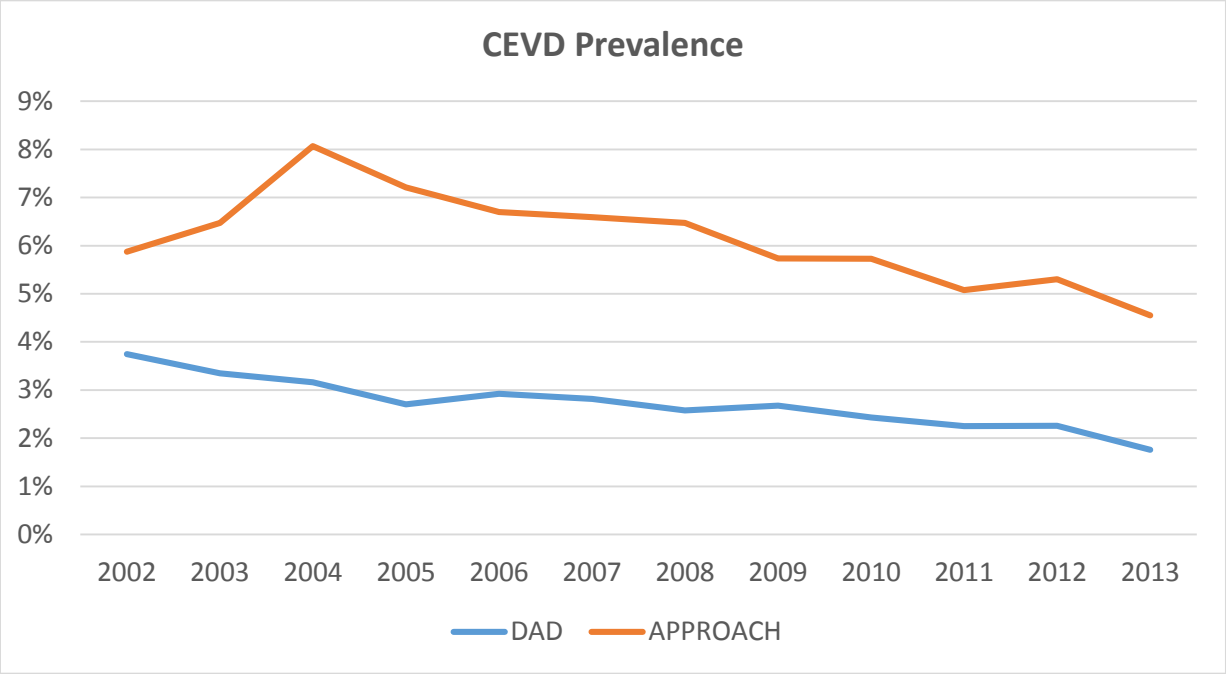
5.2.6 Cerebrovascular Disease (CEVD)

The prevalence of CEVD was noticeably lower in DAD than in APPROACH. In DAD, CEVD prevalence ranged from 1.8% to 3.7%. In APPROACH, CEVD prevalence ranged from 4.6% to 8.1%. Further, it appears that the prevalence in both databases were decreasing over time.

Table 18. CEVD Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	5.9	3.7
2003	6.5	3.3
2004	8.1	3.2
2005	7.2	2.7
2006	6.7	2.9
2007	6.6	2.8
2008	6.5	2.6
2009	5.7	2.7
2010	5.7	2.4
2011	5.1	2.3
2012	5.3	2.3
2013	4.6	1.8

Figure 18. CEVD Prevalence



DAD had poor validity for the coding of CEVD. While specificity and NPV were high, sensitivity and PPV were very low. This means that CEVD was under coded in DAD. Further, out of those that were coded a large proportion of them were coded incorrectly. Looking at time trends, we found no statistically significant trend for specificity and NPV. However, sensitivity and PPV both saw statistically significant declines, with estimates of -0.012 for sensitivity and -0.018 for PPV. This means that sensitivity is declining by 1.2% a year, while PPV is declining by 1.8% a year.

Table 19. CEVD Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	29.8, (21.4-38.2)	97.9, (97.2-98.5)	46.7, (35.3-58.1)	95.7, (94.8-96.6)
2003	23.9, (17.4-30.3)	98.1, (97.5-98.6)	46.1, (35.7-56.5)	94.9, (94.0-95.8)
2004	24.0, (18.3-29.7)	98.7, (98.2-99.1)	61.1, (50.8-71.5)	93.7, (92.7-94.6)
2005	17.3, (11.9-22.6)	98.4, (97.9-98.9)	46.0, (34.6-57.4)	93.9, (93.0-94.8)
2006	16.1, (10.6-21.5)	98.0, (97.5-98.6)	36.8, (25.9-47.6)	94.2, (93.3-95.1)
2007	18.5, (12.5-24.5)	98.3, (97.8-98.8)	43.4, (31.7-55.0)	94.5, (93.6-95.4)
2008	19.4, (13.3-25.6)	98.6, (98.1-99.1)	48.8, (36.5-61.2)	94.7, (93.7-95.6)
2009	18.9, (12.3-25.4)	98.3, (97.8-98.8)	40.5, (28.4-52.5)	95.2, (94.4-96.1)
2010	15.9, (9.9-21.9)	98.4, (97.9-98.9)	37.6, (25.5-49.7)	95.1, (94.2-95.9)
2011	15.5, (9.3-21.8)	98.5, (98.0-99.0)	35.0, (22.7-47.4)	95.6, (94.8-96.4)
2012	13.7, (8.1-19.4)	98.4, (97.9-98.9)	32.3, (20.5-44.0)	95.3, (94.5-96.1)
2013	11.8, (6.1-17.5)	98.7, (98.3-99.2)	30.6, (17.6-43.7)	95.9, (95.2-96.7)

Figure 19. CEVD Validity

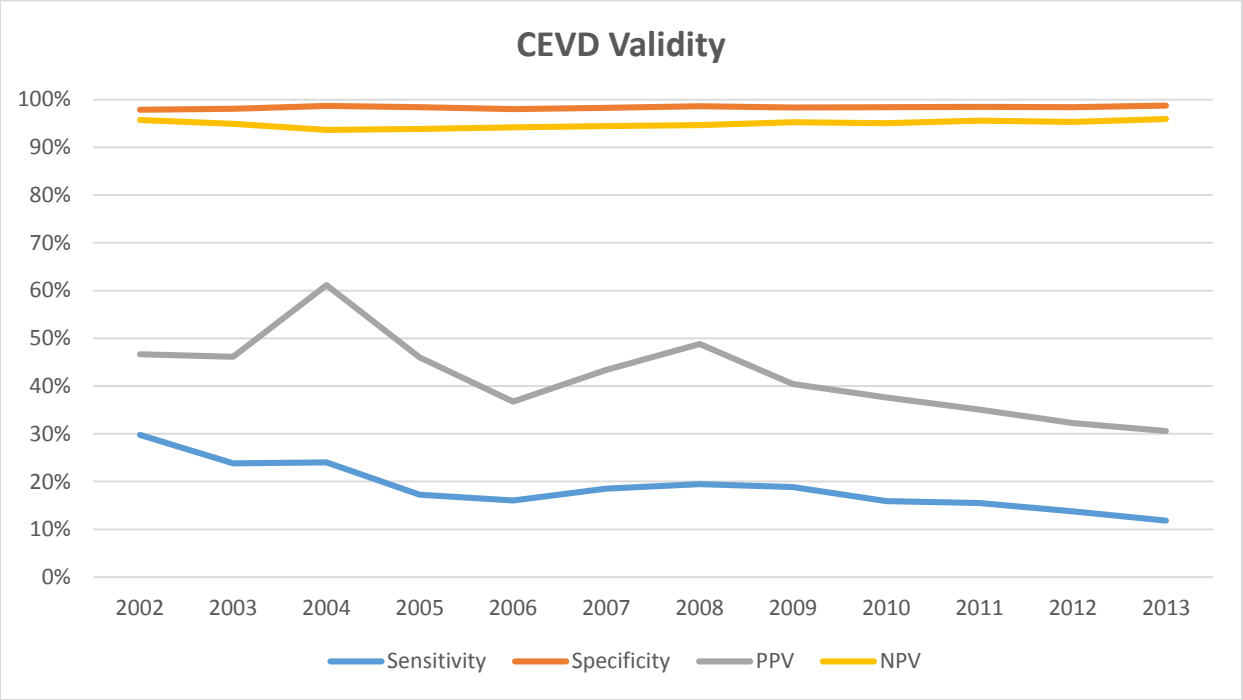


Table 20. Linear regression for DAD validity trend in CEVD coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
CEVD	Sensitivity	-0.0121	0.0022	-5.51	0.0003
	Specificity	0.0004	0.0002	2.21	0.0519
	PPV	-0.0178	0.0049	-3.66	0.0044
	NPV	0.0010	0.0006	1.82	0.0988

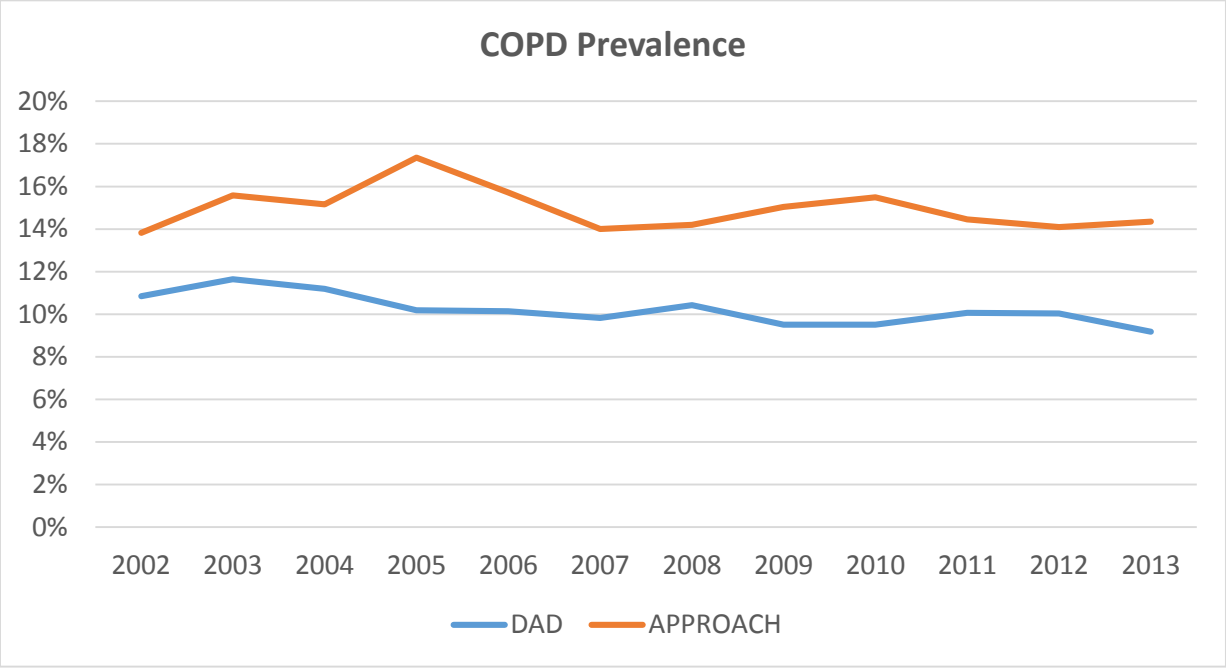
5.2.7 Pulmonary Disease (COPD)

The prevalence of COPD in DAD was consistently lower than that in APPROACH. This type of pattern indicates the possibility that COPD was under coded in DAD. The trend of prevalence over time in both databases seem fairly stable.

Table 21. COPD Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	13.8	10.8
2003	15.6	11.6
2004	15.2	11.2
2005	17.4	10.2
2006	15.7	10.1
2007	14.0	9.8
2008	14.2	10.4
2009	15.0	9.5
2010	15.5	9.5
2011	14.4	10.1
2012	14.1	10.0
2013	14.4	9.2

Figure 20. COPD Prevalence



COPD had high specificity and NPV, but low sensitivity and PPV. This means that the COPD was under coded in DAD, and coded inaccurately. No statistically significant time trend was found for specificity and NPV, however both sensitivity and PPV showed statistically significant declines. Sensitivity had an estimate of -0.013 and PPV had an estimate of -0.013. This means that sensitivity and PPV were both declining by approximately 1.3% each year.

Table 22. COPD Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	49.4, (43.4-55.3)	95.3, (94.3-96.3)	62.9, (56.4-69.4)	92.2, (90.9-93.4)
2003	45.9, (41.1-50.8)	94.7, (93.8-95.6)	61.5, (56.1-67.0)	90.5, (89.3-91.7)
2004	44.1, (39.3-49.0)	94.7, (93.8-95.6)	59.8, (54.2-65.4)	90.5, (89.3-91.7)
2005	38.0, (33.6-42.4)	95.7, (94.8-96.5)	64.8, (59.1-70.4)	88.0, (86.7-89.3)
2006	38.7, (33.9-43.4)	95.2, (94.3-96.1)	59.9, (53.9-65.8)	89.3, (88.0-90.5)
2007	38.3, (33.2-43.5)	94.8, (93.9-95.8)	54.7, (48.5-61.0)	90.4, (89.2-91.6)
2008	41.5, (36.3-46.7)	94.7, (93.8-95.7)	56.5, (50.4-62.6)	90.7, (89.5-91.9)
2009	34.6, (29.7-39.5)	94.9, (94.0-95.9)	54.7, (48.3-61.2)	89.1, (87.8-90.4)
2010	34.3, (29.6-39.1)	95.0, (94.1-96.0)	55.9, (49.6-62.2)	88.8, (87.5-90.1)
2011	34.1, (29.2-38.9)	94.0, (93.0-95.0)	48.9, (42.8-55.1)	89.4, (88.2-90.7)
2012	35.6, (30.7-40.4)	94.1, (93.2-95.1)	49.9, (43.9-55.9)	89.9, (88.7-91.1)
2013	33.0, (28.3-37.6)	94.8, (93.9-95.7)	51.6, (45.4-57.8)	89.4, (88.2-90.6)

Figure 21. COPD Validity

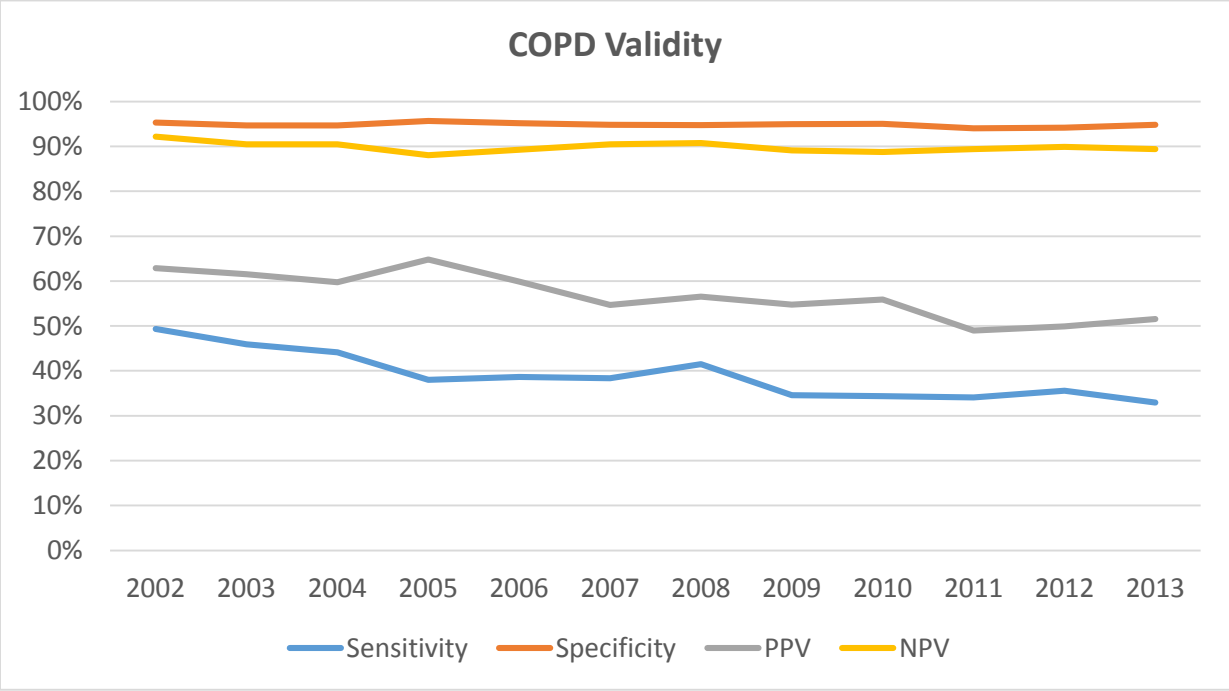


Table 23. Linear regression for DAD validity trend in COPD coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
COPD	Sensitivity	-0.0130	0.0021	-6.24	<.0001
	Specificity	-0.0007	0.0003	-1.99	0.0743
	PPV	-0.0127	0.0020	-6.34	<.0001
	NPV	-0.0014	0.0008	-1.61	0.1384

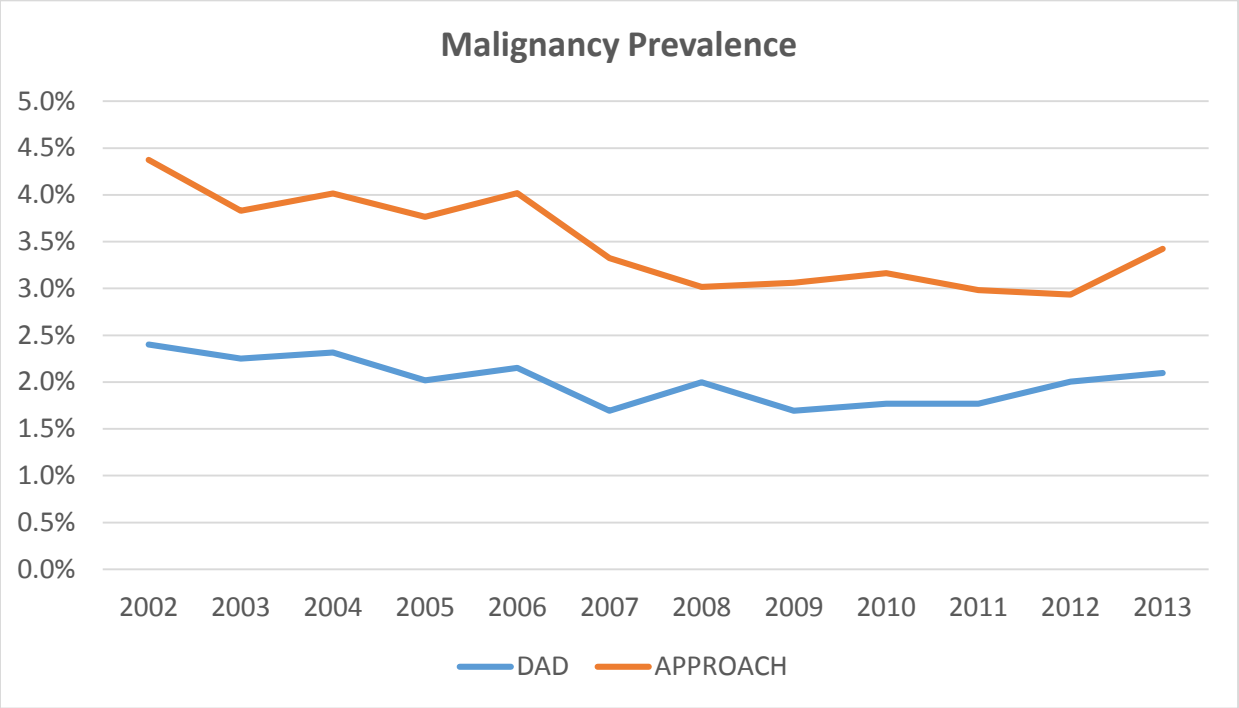
5.2.8 Malignancy

We saw a familiar pattern when looking at the prevalence of malignancy in the two databases. Prevalence of malignancy in DAD was consistently lower compared with APPROACH in the study period. This again suggests the possibility that DAD was under coding the condition. Prevalence of malignancy appears to decline in APPROACH, and is mostly similar in DAD over time.

Table 24. Malignancy Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	4.4	2.4
2003	3.8	2.3
2004	4.0	2.3
2005	3.8	2.0
2006	4.0	2.2
2007	3.3	1.7
2008	3.0	2.0
2009	3.1	1.7
2010	3.2	1.8
2011	3.0	1.8
2012	2.9	2.0
2013	3.4	2.1

Figure 22. Malignancy Prevalence



DAD had high specificity and NPV, but low sensitivity and PPV. This means that malignancy was under coded in DAD and when coded it was coded poorly. When looking at time trends, we see no statistically significant trend for sensitivity and specificity. However PPV and NPV does show statistically significant trends with estimates of -0.02 for PPV and 0.00054 for NPV. This means that PPV is declining by approximately 2% a year, while NPV is increasing by approximately 0.05% a year.

Table 25. Malignancy Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	33.7, (23.6-43.8)	99.0, (98.6-99.5)	61.5, (47.5-75.4)	97.0, (96.3-97.8)
2003	29.1, (20.2-38.0)	98.8, (98.4-99.2)	49.6, (36.9-62.3)	97.2, (96.6-97.9)
2004	27.5, (19.0-35.9)	98.7, (98.3-99.2)	47.7, (35.3-60.0)	97.0, (96.4-97.7)
2005	21.5, (13.5-29.5)	98.7, (98.3-99.2)	40.2, (27.2-53.2)	97.0, (96.3-97.6)
2006	21.1, (13.3-29.0)	98.6, (98.2-99.1)	39.5, (26.6-52.3)	96.8, (96.1-97.4)
2007	15.4, (7.6-23.2)	98.8, (98.3-99.2)	30.2, (16.4-44.1)	97.1, (96.5-97.8)
2008	20.5, (11.3-29.8)	98.6, (98.1-99.1)	31.0, (18.0-44.0)	97.6, (96.9-98.2)
2009	15.3, (7.0-23.6)	98.7, (98.3-99.2)	27.7, (13.9-41.5)	97.4, (96.7-98.0)
2010	21.5, (12.4-30.5)	98.9, (98.5-99.3)	38.5, (24.2-52.8)	97.5, (96.9-98.1)
2011	19.4, (10.4-28.3)	98.8, (98.3-99.2)	32.6, (18.9-46.3)	97.6, (96.9-98.2)
2012	26.1, (16.3-35.8)	98.7, (98.3-99.2)	38.2, (25.2-51.2)	97.8, (97.2-98.3)
2013	19.9, (11.8-28.0)	98.5, (98.1-99.0)	32.5, (20.3-44.6)	97.2, (96.6-97.8)

Figure 23. Malignancy Validity

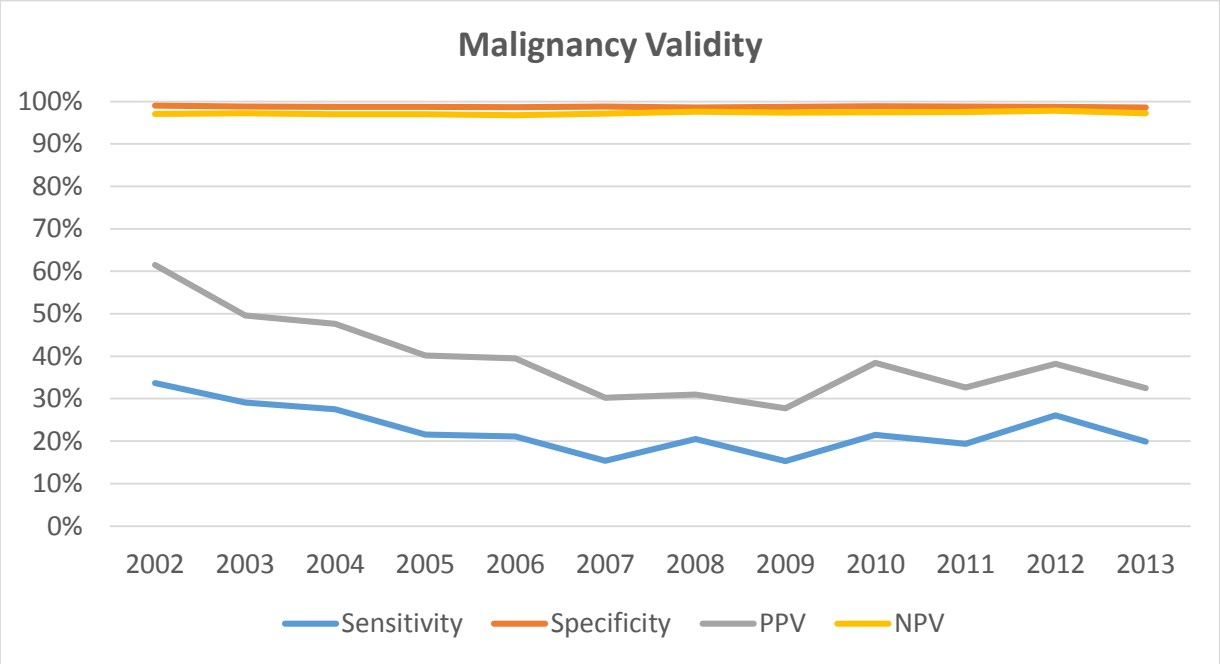


Table 26. Linear regression for DAD validity trend in malignancy coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
Malignancy	Sensitivity	-0.0087	0.0040	-2.19	0.0533
	Specificity	-0.0002	0.0001	-1.91	0.0853
	PPV	-0.0199	0.0058	-3.45	0.0063
	NPV	0.0005	0.0002	2.73	0.0211

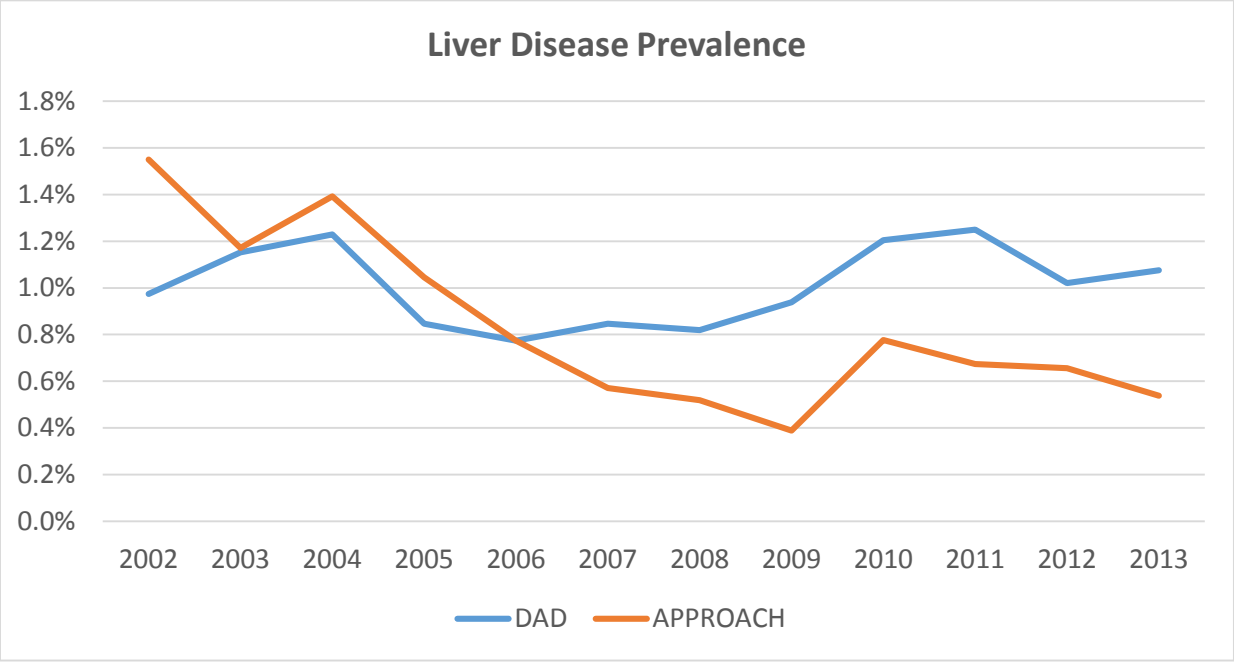
5.2.9 Liver Disease

The prevalence of liver disease in APPROACH and DAD was similar. However the prevalence in APPROACH appears to decline slightly, whereas the prevalence in DAD was mostly constant. The prevalence in both databases was very low, meaning that this was a rare condition.

Table 27. Liver Disease Prevalence

YEAR	PREVALENCE (%)	
	APPROACH	DAD
2002	1.5	1.0
2003	1.2	1.2
2004	1.4	1.2
2005	1.0	0.8
2006	0.8	0.8
2007	0.6	0.8
2008	0.5	0.8
2009	0.4	0.9
2010	0.8	1.2
2011	0.7	1.3
2012	0.7	1.0
2013	0.5	1.1

Figure 24. Liver Disease Prevalence



While specificity and NPV were both high, this could be caused by the fact that the prevalence was very low. Both sensitivity and PPV were very low, meaning that liver disease was under coded in DAD, and it was not coded well. When looking at time trends, we find no statistically significant trend for sensitivity. However we do see statistically significant trends for specificity, PPV and NPV. The estimates are -0.00033, -0.025 and 0.00057 for specificity, PPV, and NPV respectively. These estimates suggests that specificity will decrease by 0.033% a year, PPV will decrease by 2.5% a year and NPV will increase by 0.057% a year.

Table 28. Liver Disease Validity

YEAR	VALIDITY			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
2002	24.2, (8.7-39.7)	99.4, (99.0-99.7)	38.5, (16.6-60.3)	98.8, (98.3-99.3)
2003	31.7, (15.0-48.4)	99.2, (98.9-99.6)	32.3, (15.6-48.9)	99.2, (98.8-99.5)
2004	27.3, (12.9-41.7)	99.1, (98.8-99.5)	30.9, (15.2-46.6)	99.0, (98.6-99.4)
2005	19.0, (4.3-33.6)	99.3, (99.0-99.6)	23.4, (6.1-40.7)	99.1, (98.8-99.5)
2006	26.8, (7.0-46.7)	99.4, (99.1-99.7)	26.8, (7.4-46.2)	99.4, (99.1-99.7)
2007	31.0, (6.1-56.0)	99.3, (99.0-99.6)	20.9, (3.5-38.3)	99.6, (99.4-99.9)
2008	26.9, (1.6-52.3)	99.3, (99.0-99.6)	17.1, (.6-33.5)	99.6, (99.4-99.9)
2009	42.1, (8.6-75.6)	99.2, (98.9-99.6)	17.4, (1.7-33.1)	99.8, (99.6-100.0)
2010	22.5, (3.5-41.5)	99.0, (98.6-99.4)	14.5, (2.0-27.1)	99.4, (99.1-99.7)
2011	28.6, (6.5-50.6)	98.9, (98.5-99.3)	15.4, (2.8-27.9)	99.5, (99.2-99.8)
2012	19.4, (.4-38.5)	99.1, (98.7-99.5)	12.5, (.1-24.9)	99.5, (99.2-99.7)
2013	13.3, (-4.7-31.3)	99.0, (98.6-99.4)	6.7, (-2.4-15.7)	99.5, (99.3-99.8)

Figure 25. Liver Disease Validity

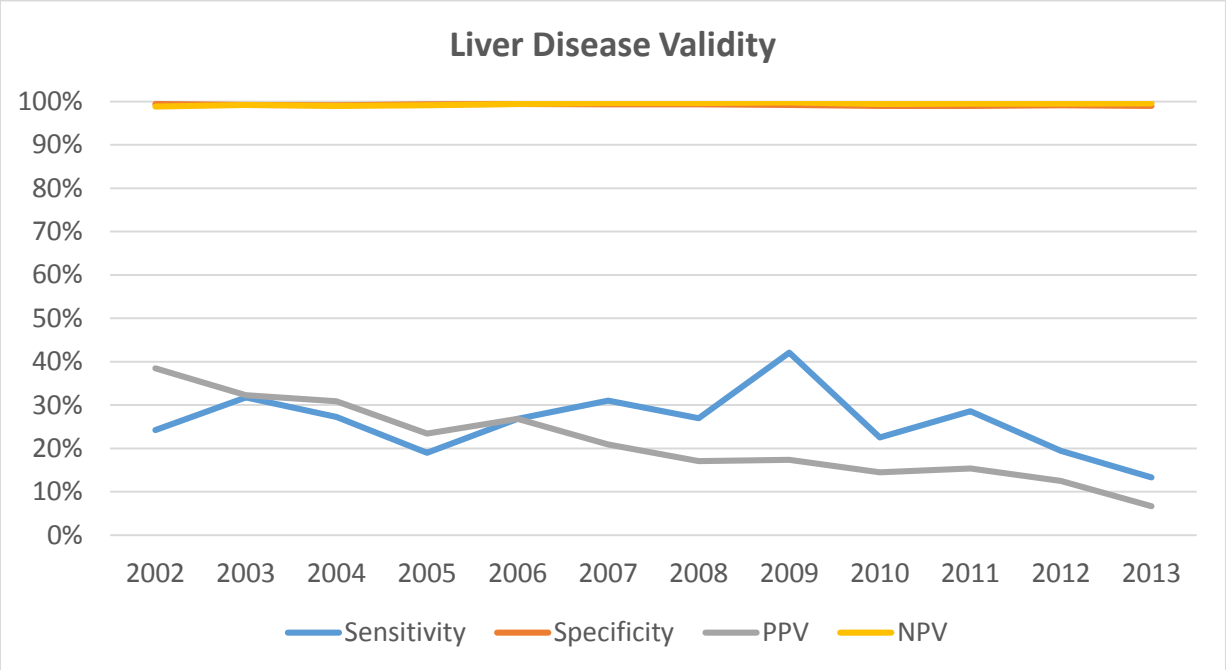


Table 29. Linear regression for DAD validity trend in liver disease coding, 2002-2013

		Coefficient	Standard Error	t-value	Pr>t
Liver Disease	**Sensitivity**	-0.0057	0.0062	-0.91	0.3835
	Specificity	-0.0003	0.0001	-2.98	0.0138
	PPV	-0.0249	0.0020	-12.23	<.0001
	NPV	0.0006	0.0002	3.36	0.0073

5.3 Stratification Results

The validation results were stratified by age, sex and hospital location for all 9 comorbidities. We calculated estimated slopes as well as 95% confidences intervals for the trend of sensitivity, specificity, PPV and NPV for all 9 conditions. These results are shown in the tables below.

Table 30. Linear slope and 95% confidence interval for DAD validity, 2002-2013 by Age

Condition		Coefficient, 95% Confidence Interval (%)	
		Age	
		18-64	>65
Hypertension	Sensitivity	0.39, (0.05, 0.73)	0.35, (-0.15, 0.84)
	Specificity	-1.39, (-1.77, -1.01)	-0.80, (-1.14, -0.46)
	PPV	-0.48, (-0.70, -0.25)	-0.74, (-1.07, -0.41)
	NPV	0.03, (-0.70, 0.76)	0.30, (-0.27, 0.86)
Diabetes	Sensitivity	0.49, (-0.05, 1.03)	0.28, (-0.28, 0.84)
	Specificity	-0.34, (-0.43, -0.24)	-0.27, (-0.37, -0.18)
	PPV	-0.64, (-0.87, -0.40)	-0.88, (-1.18, -0.58)
	NPV	0.13, (-0.03, 0.29)	0.07, (-0.07, 0.22)
Hyperlipidemia	Sensitivity	-3.53, (-4.90, -2.16)	-3.58, (-4.16, -3.00)
	Specificity	1.12, (0.67, 1.57)	1.63, (1.14, 2.12)
	PPV	-0.71, (-1.07, -0.35)	-1.05, (-1.68, -0.42)
	NPV	-0.20, (-0.91, 0.52)	0.82, (0.08, 1.56)
Heart Failure	Sensitivity	-0.12, (-0.55, 0.32)	0.37, (-0.47, 1.21)
	Specificity	-0.21, (-0.39, -0.02)	-0.14, (-0.22, -0.06)
	PPV	-1.27, (-1.73, -0.81)	-1.08, (-1.60, -0.56)
	NPV	0.13, (0.04, 0.23)	0.09, (0.02, 0.15)
PVD	Sensitivity	-2.10, (-2.69, -1.51)	-1.36, (-2.39, -0.33)
	Specificity	0.13, (0.06, 0.21)	0.06, (0.03, 0.09)
	PPV	-2.68, (-3.37, -1.99)	-2.53, (-3.18, -1.88)
	NPV	0.24, (0.15, 0.34)	0.17, (0.11, 0.22)
CEVD	Sensitivity	-1.40, (-1.95, -0.85)	-0.76, (-1.26, -0.26)
	Specificity	0.07, (0.00, 0.14)	0.00, (-0.03, 0.04)
	PPV	-1.67, (-2.57, -0.77)	-1.77, (-3.03, -0.51)
	NPV	0.09, (-0.07, 0.26)	0.07, (0.00, 0.15)
COPD	Sensitivity	-2.04, (-3.17, -0.91)	-0.98, (-1.63, -0.33)
	Specificity	-0.08, (-0.17, 0.02)	-0.11, (-0.17, -0.05)
	PPV	-1.15, (-1.60, -0.70)	-1.52, (-1.97, -1.07)
	NPV	-0.14, (-0.34, 0.07)	-0.09, (-0.26, 0.09)
Cancer	Sensitivity	-1.26, (-2.20, -0.32)	0.05, (-0.92, 1.01)
	Specificity	-0.04, (-0.07, 0.00)	-0.02, (-0.04, 0.01)
	PPV	-2.16, (-3.42, -0.90)	-1.65, (-3.42, 0.12)
	NPV	0.02, (-0.04, 0.07)	0.06, (0.00, 0.12)
Liver Disease	Sensitivity	0.09, (-2.44, 2.61)	-1.44, (-2.85, -0.03)
	Specificity	-0.01, (-0.04, 0.02)	-0.05, (-0.09, -0.01)
	PPV	-1.66, (-3.84, 0.52)	-2.77, (-3.36, -2.18)
	NPV	0.08, (0.04, 0.11)	0.04, (0.00, 0.08)

Table 31. Linear slope and 95% confidence interval for DAD validity, 2002-2013 by Sex

Condition		Coefficient, 95% Confidence Interval (%)	
		Sex	
		Male	Female
Hypertension	Sensitivity	0.40, (-0.05, 0.85)	0.29, (-0.01, 0.58)
	Specificity	-0.98, (-1.31, -0.66)	-0.80, (-1.13, -0.46)
	PPV	-0.62, (-0.91, -0.34)	-0.59, (-0.84, -0.34)
	NPV	0.23, (-0.35, 0.81)	0.47, (-0.12, 1.06)
Diabetes	Sensitivity	0.44, (-0.14, 1.01)	0.30, (-0.17, 0.78)
	Specificity	-0.32, (-0.41, -0.22)	-0.24, (-0.34, -0.14)
	PPV	-0.78, (-1.05, -0.52)	-0.67, (-0.93, -0.41)
	NPV	0.10, (-0.06, 0.27)	0.10, (-0.03, 0.24)
Hyperlipidemia	Sensitivity	-3.48, (-4.09, -2.87)	-3.30, (-3.85, -2.75)
	Specificity	1.49, (0.98, 2.00)	1.16, (0.78, 1.54)
	PPV	-0.91, (-1.39, -0.43)	-0.89, (-1.37, -0.40)
	NPV	0.37, (-0.42, 1.17)	0.34, (-0.32, 1.01)
Heart Failure	Sensitivity	0.24, (-0.29, 0.77)	-0.28, (-0.95, 0.39)
	Specificity	-0.08, (-0.21, 0.05)	-0.29, (-0.43, -0.15)
	PPV	-1.03, (-1.51, -0.55)	-1.62, (-1.99, -1.25)
	NPV	0.12, (0.06, 0.18)	0.11, (0.01, 0.21)
PVD	Sensitivity	-1.84, (-2.55, -1.13)	-1.91, (-2.58, -1.24)
	Specificity	0.10, (0.05, 0.15)	0.10, (0.04, 0.15)
	PPV	-2.55, (-3.06, -2.04)	-2.74, (-3.55, -1.93)
	NPV	0.20, (0.15, 0.25)	0.24, (0.15, 0.33)
CEVD	Sensitivity	-1.05, (-1.53, -0.57)	-1.43, (-2.15, -0.71)
	Specificity	0.05, (0.02, 0.09)	0.01, (-0.05, 0.08)
	PPV	-1.30, (-2.09, -0.51)	-2.68, (-4.37, -0.99)
	NPV	0.09, (-0.02, 0.19)	0.13, (-0.02, 0.28)
COPD	Sensitivity	-1.50, (-2.00, -1.00)	-0.56, (-1.13, 0.01)
	Specificity	-0.05, (-0.13, 0.02)	-0.16, (-0.27, -0.05)
	PPV	-1.29, (-1.71, -0.87)	-1.42, (-2.13, -0.71)
	NPV	-0.17, (-0.31, -0.03)	0.09, (-0.21, 0.39)
Cancer	Sensitivity	-1.08, (-2.05, -0.11)	-0.47, (-1.33, 0.39)
	Specificity	-0.03, (-0.05, -0.01)	0.00, (-0.05, 0.05)
	PPV	-2.20, (-3.30, -1.10)	-1.39, (-3.03, 0.25)
	NPV	0.04, (-0.02, 0.10)	0.09, (0.01, 0.17)
Liver Disease	Sensitivity	-0.97, (-2.34, 0.41)	-0.58, (-3.46, 2.30)
	Specificity	-0.03, (-0.06, 0.00)	-0.04, (-0.06, -0.02)
	PPV	-2.15, (-2.82, -1.48)	-3.31, (-4.41, -2.21)
	NPV	0.03, (0.00, 0.07)	0.11, (0.05, 0.16)

Table 32. Linear slope and 95% confidence interval for DAD validity, 2002-2013 by hospital location

Condition		Coefficient, 95% Confidence Interval (%)	
		Hospital Location	
		Calgary	Edmonton
Hypertension	Sensitivity	0.13, (-0.45, 0.71)	0.12, (-0.07, 0.31)
	Specificity	-0.97, (-1.41, -0.54)	-0.94, (-1.28, -0.60)
	PPV	-0.91, (-1.32, -0.51)	-0.57, (-0.84, -0.30)
	NPV	0.28, (-0.51, 1.08)	0.02, (-0.39, 0.42)
Diabetes	Sensitivity	0.78, (-0.12, 1.68)	0.04, (-0.22, 0.30)
	Specificity	-0.35, (-0.42, -0.28)	-0.27, (-0.39, -0.14)
	PPV	-0.92, (-1.12, -0.72)	-0.74, (-1.07, -0.41)
	NPV	0.21, (-0.01, 0.42)	0.01, (-0.07, 0.10)
Hyperlipidemia	Sensitivity	-0.53, (-3.47, 2.41)	-4.87, (-5.65, -4.09)
	Specificity	0.91, (0.44, 1.38)	2.21, (1.29, 3.13)
	PPV	-1.11, (-1.53, -0.69)	-0.82, (-1.37, -0.27)
	NPV	0.60, (-0.08, 1.29)	-0.18, (-1.14, 0.78)
Heart Failure	Sensitivity	0.04, (-0.57, 0.64)	0.02, (-0.61, 0.65)
	Specificity	-0.05, (-0.26, 0.16)	-0.21, (-0.40, -0.03)
	PPV	-0.95, (-1.46, -0.43)	-1.66, (-2.31, -1.01)
	NPV	0.12, (0.05, 0.19)	0.12, (0.02, 0.23)
PVD	Sensitivity	-1.83, (-2.49, -1.17)	-2.81, (-3.59, -2.03)
	Specificity	0.14, (0.09, 0.20)	0.06, (-0.01, 0.12)
	PPV	-2.82, (-3.94, -1.70)	-3.08, (-3.89, -2.27)
	NPV	0.31, (0.26, 0.37)	0.10, (0.03, 0.17)
CEVD	Sensitivity	-1.26, (-1.72, -0.80)	-1.63, (-2.38, -0.88)
	Specificity	0.04, (0.00, 0.08)	0.06, (0.00, 0.11)
	PPV	-2.09, (-3.28, -0.90)	-2.04, (-3.15, -0.93)
	NPV	0.09, (0.00, 0.17)	0.11, (-0.01, 0.23)
COPD	Sensitivity	-1.84, (-2.56, -1.12)	-0.18, (-0.80, 0.43)
	Specificity	0.16, (0.04, 0.28)	-0.23, (-0.37, -0.08)
	PPV	-0.02, (-0.65, 0.61)	-1.61, (-2.14, -1.08)
	NPV	-0.39, (-0.65, -0.12)	0.11, (-0.01, 0.23)
Cancer	Sensitivity	-0.70, (-1.52, 0.13)	-1.10, (-2.80, 0.60)
	Specificity	-0.03, (-0.06, 0.01)	0.00, (-0.04, 0.04)
	PPV	-1.77, (-3.28, -0.26)	-2.51, (-4.22, -0.80)
	NPV	-0.01, (-0.07, 0.05)	0.08, (0.03, 0.14)
Liver Disease	Sensitivity	-0.29, (-1.27, 0.70)	0.62, (-1.85, 3.09)
	Specificity	0.01, (-0.01, 0.03)	-0.06, (-0.11, -0.02)
	PPV	-1.22, (-2.46, 0.02)	-2.71, (-3.83, -1.59)
	NPV	0.07, (0.03, 0.12)	0.04, (0.02, 0.07)

The results in Table 21 and 22 show that the 95% confidence intervals for all the slope coefficients of sensitivity, specificity, PPV and NPV overlap with those aged 18-64 compared with those aged 65 and over, and with males compared with females. This means that time trends of sensitivity, specificity, PPV and NPV are not statistically affected by age or sex for all 9 conditions.

Hospital location does not affect sensitivity, specificity, PPV and NPV time trends for eight of the nine conditions (see Table 23). However, COPD contains a statistical difference in sensitivity and NPV trends between Calgary area and Edmonton area hospitals. Our results suggest that sensitivity for coding of COPD is declining at a faster rate in Calgary area hospitals compared with Edmonton area hospitals. Further it suggests that while NPV is decreasing in Calgary area hospitals, it is increasing in Edmonton area hospitals. With that said, both differences are just barely statistically significant. Further, while we observe this difference in the trend of sensitivity and NPV for COPD, the majority of outcome measures showed no statistically significant difference between Calgary and Edmonton area hospitals.

Overall, even with a few minor statistically significant differences, these stratification results show that there is no clinically significant effect on the time trends of the study outcomes caused by age, sex, or hospital location.

Chapter 6: Discussion

6.1 Discussion of Preliminary Literature Review

Prior to conducting the literature review in chapter 3, we conducted a preliminary literature search. We discovered that the results of this preliminary review did not adequately address our research question. However, the experience of this preliminary search helped adjust our search strategy. This section will discuss this preliminary search and how it led to a new refined search strategy.

Three databases were searched for the preliminary literature review: Medline, PubMed and Embase. The search strategy combined the search results of keywords related to administrative data and discharge abstract data, with the search results of keywords and MESH headings related to quality or validation. All articles related to the validation of discharge data were included in the search. Articles that were not written in English were excluded from consideration.

6.1.1 Preliminary Review Results

The search strategy yielded 4595 results. There were 2589 duplicates, and 2006 abstracts to be screened. 1703 were excluded at this stage, leaving 213 full text articles for review. After full text review 154 were excluded for not meeting the inclusion criteria, and 5 were excluded for being written in a language other than English. Overall the preliminary review yielded 54 articles.

The vast majority of studies used a "gold standard" validation. "Gold standard" validation is where researchers attempt to assess the accuracy of a source of data by comparing it to another source of data (the gold standard) that is deemed accurate. There were three common

types of validation conducted: patient identification, comorbidity validation, and procedure validation. Patient identification involved researchers attempting to identify a cohort of patients with a particular condition or disease from a larger population. Researchers would do this using both the test data and the "gold standard", and compare how well the patients identified in test data matched with the information in the "gold standard". Comorbidity validation studies, followed a similar blueprint, however they focused on how accurately patient comorbidity information was collected. Procedure validation, also used a similar technique, only researchers looked at how accurate procedures were recorded. The most common type of study was identification studies (37 of 54), followed by comorbidity validation (9 of 54) and procedure validation (4 of 54).

Looking at when these studies were published, a clear trend can be seen where the number of studies is increasing from year to year. For example, the earliest study was published in 1994, and out of the 54 articles found, it was the only one published in that year. However 11 of the 54 studies were published in 2013. A clear increasing trend can be seen in regards to the publication of administrative data validation studies.

Another surprising trend that was seen in the search results was that 57% of studies were published in Canada. 20% of studies were published in the United States. The rest of the studies were mostly published in various European countries. After consultation with research committee members, it was unusual that there was not more studies published in the United States. Further, the overall number of search results seemed low to several committee members.

A possible cause of the overall low number of search results, as well as the unusual distribution of publication location is the keywords used in the search. The term "Discharge

Abstract Database" is most commonly used in Canada. The search was likely missing many administrative data validation studies published in other countries.

After further consultation, it was decided that a better search strategy, which is better able to capture administrative data from any country was needed. Further, given that such a search would likely yield a greater number of results, it was also decided to further focus and narrow the search inclusion criteria. Therefore, a more specific literature search was conducted focusing specifically on DAD validation studies that looked at comorbidity accuracy, and the trend of that accuracy over time. This literature search was described in chapter 3.

6.2 Discussion of Validation Results

6.2.1 Interpretation Issue

One issue we have with the results of our time trend analysis deals with statistical significance and clinical or practical significance. When looking at changes of sensitivity, specificity, PPV and NPV over time, we found many trends to be statistically significant. However, the magnitude of change for some of these trends are very small and may not actually be practically significant. There is no established guideline in regards to how large a change in data validity should be considered practically and clinically significant. Practical significance will vary depending on the specific research conducted and the time span analyzed. This can be illustrated with a simple example. Suppose we observe a change in sensitivity of 2% a year, and we want to consider whether this change is practically significant. The answer will depend on the time span of our analysis. Suppose we are simply considering data validity between this year and the previous year, and suppose our sensitivity in the previous year was 50%. This means that the sensitivity for the current year is now 52%. We can argue that this is a very minor difference and

for all practical purposes data validity is the same for these two years. Now imagine we want to do a comparison 15 years in the future. Assuming all else has remained constant, we would find that our sensitivity of 50% is now 80%. This is definitely a practically significant difference. Therefore, practical significance of changes in data validity must be considered on a case to case basis.

With that said, we decided it would be helpful to use some classification of trend changes in order to help frame our discussion. Considering our research time span of approximately ten years, we will discuss practical significance of trend changes using the following:

- Under 1%: Consistent over time
- 1% to 2%: Small change over time
- Over 2%: Large change over time

Again, we stress that these should not be used as a guideline to establish practical significance, rather it is simply a way to aid discussion of our results.

6.2.2 Hypertension

Our results show that DAD coding for hypertension was fairly well done over the study period. Sensitivity ranged from 74.6% to 82.5%, specificity ranged from 71.0% to 80.9%, PPV ranged from 79.3% to 87.0%, and NPV ranged from 65.5% to 78.0%. These results are similar to previous studies done on DAD hypertension coding validity [57,58,59,60]. For example, a study by Quan et al. found hypertension coding validity to have sensitivity of 75%, specificity of 94%, PPV of 81%, and NPV of 92% [24]. It should be noted that this study validated more than just DAD. The hypertension definition used in this study included 2 physician claims or 1 hospital

admission within 2 years. So while the data validation wasn't exactly the same, we find that the validation results are still somewhat similar to our results for this condition.

While our time trend analysis found that specificity and PPV for hypertension coding were decreasing at 0.9% and 0.6%, respectively. We felt the magnitude of the decrease was very small. Therefore, it seems that the data validity for hypertension coding is mostly stable from 2002 to 2013. There is little research published dealing with the trend of hypertension coding in DAD over time. However, an exception is a study by Januel et al. This study assessed the accuracy of hypertension coding at 3 time periods: 1999, 2001, and 2003. They found that the validity of DAD coding for hypertension actually increased slightly over that period due to introduction of ICD-10 [53]. There are two possible reasons why the results of this study differ with ours. First, the Januel et al. study was conducted over a slightly different time period (1999-2003) compared to our study (2002-2013). Second this study was conducted in Switzerland. It is possible that physician and coder practises in Switzerland and/or during that time period differed with the physician and coder habits in Alberta during 2002 to 2013. In Switzerland, there are few professional training programs for coders [53], whereas, coders in Alberta have to complete a two year college program [61].

6.2.3 Diabetes

Our study found high validity for the coding of diabetes in DAD. Sensitivity ranged from 83.8% to 92.1%, specificity ranged from 93.9% to 97.0%, PPV ranged from 81.5% to 90.5%, and NPV ranged from 94.5% to 97.6%. These results are similar to previously studies [51,52]. For example, a 2013 study by Leong et al. conducted a systematic review of diabetes coding using administrative data. They found that sensitivity was 82.3% and specificity was 97.9% [62]. However, this study validated a slightly different coding definition. The coding definition they

used included both physician claims and DAD information. Regardless, these findings correlate well with our results.

Similar to hypertension, our time trend analysis show that specificity and PPV are decreasing. However, the decrease is very minor, with specificity declining by 0.3% a year, and PPV declining by 0.8% a year. This equates to a decrease of 3% and 8% over a ten year period. Therefore it seems that data validity for diabetes is also fairly consistent. This type of trend has been seen in other research. For example, a 2010 study by Chen et al looked at diabetes coding validity in 2001 and 2004. However, this study validated a case definition based on "2 physician claims within 2 years, or 1 hospital admission". So, while the data source undergoing validation is not an exact match, this study also found no significant differences in the validity of administrative data recording diabetes, over the two time periods analyzed [52].

6.2.4 Hyperlipidemia

The findings of hyperlipidemia coding validity in our study range from moderate to poor. Sensitivity ranged from 15.3% to 58.1%, specificity ranged from 75.8% to 93.5%, PPV ranged from 76.9% to 89.6% and NPV ranged from 29.5% to 44.3%. These values are significantly lower than results seen in other studies. For example a study by Marrie et al. found that the coding for hyperlipidemia had sensitivity of 86.4%, specificity of 92.7%, PPV of 67.1% and NPV of 97.5% [63]. Again this is not the cleanest comparison because the Marrie et al. study used hospital admission, physician claims, and prescription data in its hyperlipidemia definition.

Our time trend analysis provides some evidence that coding validity for hyperlipidemia is decreasing. While specificity is increasing by 1.4% a year, sensitivity is decreasing by 3.4% a year and PPV is decreasing by 0.9% a year. The magnitude of the decrease in sensitivity is more than double the magnitude of the increase in specificity. Thus, we would conclude that data

validity for hyperlipidemia is decreasing over our study period. Few studies have considered the trend of hyperlipidemia validity over time. However, a study by Kokotailo et al. did compare ICD-9 and ICD-10 coding validity for hyperlipidemia. While this is not a direct time comparison, the fact that ICD-9 and ICD-10 were used at different times, means that it can be considered a de facto time comparison. This study found no significant differences between ICD-9 and ICD-10 coding [64]. A possible explanation as to why this differs from our study is that it is actually measuring a different time period. The authors of this study are basically comparing data from 2000 to 2001 against data from 2002 to 2003, whereas our study looked at time trends from 2002 to 2013.

6.2.5 Heart Failure

Validation results for heart failure coding in DAD is moderate to poor. Sensitivity ranged from 67.9% to 78.8%, specificity ranged from 87.5% to 89.9%, PPV ranged from 32.6% to 47.7%, and NPV ranged from 96.1% to 97.7%. These results are consistent with existing literature. Validation of heart failure in DAD tends to have large variations from study to study. Quach et al. conducted a systematic review for studies validating DAD coding for heart failure. They found that sensitivity ranged from 29% to 89%, PPV ranged from 12% to 100%, while specificity and NPV were all greater than 70% [65].

We found decreasing specificity of 0.15% a year, decreasing PPV of 1.2% a year and increasing NPV of 0.12% a year for heart failure coding. The decrease in PPV is much greater than the increase in NPV, suggesting that data validity may be decreasing overall. However, because the largest change was seen in PPV, we must also consider prevalence changes before we can interpret PPV. The prevalence of heart failure showed significant decline in APPROACH from 11.4% in 2002 to 8.0% in 2013. Therefore, we cannot conclude whether data validity for

heart failure is increasing, decreasing, or consistent over time. We can compare this once again to the study by Januel et al., where they found that validity of DAD coding for heart failure remained mostly constant from 1999 to 2001 to 2003 [53]. These trends for heart failure coding is consistent with our results.

6.2.6 Peripheral Vascular Disease (PVD)

The results of this study show that the coding for PVD in DAD is poor. Sensitivity ranged from 19.7% to 41.2%, specificity ranged from 96.5% to 97.7%, PPV ranged from 21.1% to 47.8%, and NPV ranged from 95.4% to 97.7%. The high specificity and NPV are similar to the rates found in other studies. However, sensitivity and PPV are slightly lower. For example a study by Quan et al. found that PVD coding in DAD had a sensitivity of 43.3% and a PPV of 65.5% [46]. Generally speaking sensitivity and PPV on the higher end of our range compared well with previous literature, while those years with sensitivity and PPV on the lower end did not.

We found statistically significant time trends for all four measures of validity. Sensitivity and PPV are decreasing yearly by 1.9% and 2.6% respectively. Specificity and NPV are increasing yearly by 0.1% and 0.21% respectively. The magnitude of the decrease in sensitivity and PPV far exceeds the increase in specificity and NPV. This provides evidence that the data validity for PVD coding is decreasing over time. A study conducted by Januel et al. also investigated trends for PVD coding. However, unlike our results they found that sensitivity and PPV increased for PVD coding from 1999 to 2003 [53]. However this study did validate DAD at a different time period and in a different country (Switzerland).

6.2.7 Cerebrovascular Disease (CEVD)

We found that the validity of CEVD coding in DAD was poor. Sensitivity ranged from 11.8% to 29.8%, specificity ranged from 97.9% to 98.7%, PPV ranged from 30.6% to 61.1%, and NPV ranged from 93.7% to 95.9%. Once again, specificity and NPV are on par with what has been seen in literature; however sensitivity and PPV are much lower than other research. For example, Quan et al. reported a sensitivity of 46.3% and a PPV of 83% [46].

When looking at trends of validation, we found no trend for specificity and NPV over time. However, sensitivity and PPV are decreasing by 1.2% a year and 1.8% a year, respectively. Therefore, we find that CEVD coding validity has slightly declined over the study period. Once again, we can compare this against the Januel et al. study. Their result conflicts with ours, as they found an increase in sensitivity and stable PPV from 1999 to 2003 [53].

6.2.8 Pulmonary Disease (COPD)

Validity of COPD coding in DAD is poor. Sensitivity ranged from 33.0% to 49.4%, specificity ranged from 94.0% to 95.7%, PPV ranged from 48.9% to 64.8%, and NPV ranged from 88.0% to 92.2%. Specificity and NPV, again compared well to previous research; however sensitivity and PPV were found to be lower than expected. For example, Quan et al. found a sensitivity of 52.8% and a PPV of 90.8% [46], while a study by Thomsen et al. found a PPV of 92% [66].

Similar to CEVD, we see no trend for specificity and NPV over time. However, sensitivity and PPV are both decreasing by 1.3% a year. This shows that data validity for CEVD is slightly decreasing. Again, we compared our results against the results from Januel et al. Unlike our results, Januel et al. actually found an increase in sensitivity and PPV of COPD

coding in DAD [53]. Perhaps this is once again caused by the time and region differences of the Januel et al. study compared to our study.

6.2.9 Malignancy

Malignancy coding in DAD is very poor. Sensitivity ranged from 15.3% to 33.7%, specificity ranged from 98.5% to 99.0%, PPV ranged from 27.7% to 61.5%, and NPV ranged from 96.8% to 97.8%. Similar to many of the previous conditions, specificity and NPV compared well against previous research. However sensitivity and PPV appears to be lower than previous research. For example, a study by Quan et al. found a sensitivity of 80.8% and a specificity of 86.7% [46].

We found no significant trend for sensitivity and specificity over time. However, PPV is declining by approximately 2% a year, while NPV is increasing by approximately 0.05% a year. The magnitude of PPV decrease far exceeds the magnitude of NPV increase. Similar to heart disease, the major change we see occurs in PPV. Thus, prevalence changes must also be considered. We see significant decrease in prevalence of malignancy from 4.4% to 3.4% over our study period. Therefore we are unable to conclude decisively on the trend of validity for malignancy. A small number of other studies looked at the trend of DAD validity for malignancy coding. A study by Januel et al. found increasing sensitivity and PPV for malignancy coding [53]. A study from Quan et al. compared malignancy coding in ICD-9 against ICD-10. Again while this is not a direct time comparison, it can be considered a pseudo time comparison. However unlike the Januel et al. study, they found that there were no differences in malignancy coding in DAD between ICD-9 and ICD-10 [46].

6.2.10 Liver Disease

Validation for Liver disease coded in DAD is very low. Sensitivity ranged from 19.0% to 42.1%, specificity ranged from 98.9% to 99.4%, PPV ranged from 12.5% to 38.5%, and NPV ranged from 98.8% to 99.8%. Once again, specificity and NPV compared well against previous studies. However, sensitivity was slightly lower and PPV was significantly lower than previous research. For example, Quan et al. reported a sensitivity of 85.4% and a PPV of 99.6% for the coding of liver disease in DAD [46].

We found no trend for change in sensitivity over time. We do see decreasing specificity of 0.033% a year, decreasing PPV of 2.5% a year and increasing NPV of 0.057% a year. Once again, the major change in validity is seen in PPV, and we must consider changes in prevalence before interpreting PPV. We see that prevalence of liver disease decreased significantly from 1.5% to 0.5% over our study period. Therefore we are unable to form a conclusion for the trend of validity for liver disease. There also isn't a consensus regarding the trend of liver disease coding validity in previous literature. Quan et al. found that sensitivity and PPV both remained mostly constant [46], whereas Januel et al. found that the sensitivity and PPV had both increased [53].

6.3 Potential Causes of Low and Inconsistent Validity

The results of this study show that DAD coding validity varies greatly depending on condition. Out of the nine conditions assessed in this study, only two displayed high validity. Further, when considering time trends of validity, four conditions show declining validity over time, two show consistent validity trends and validation trends for three conditions are

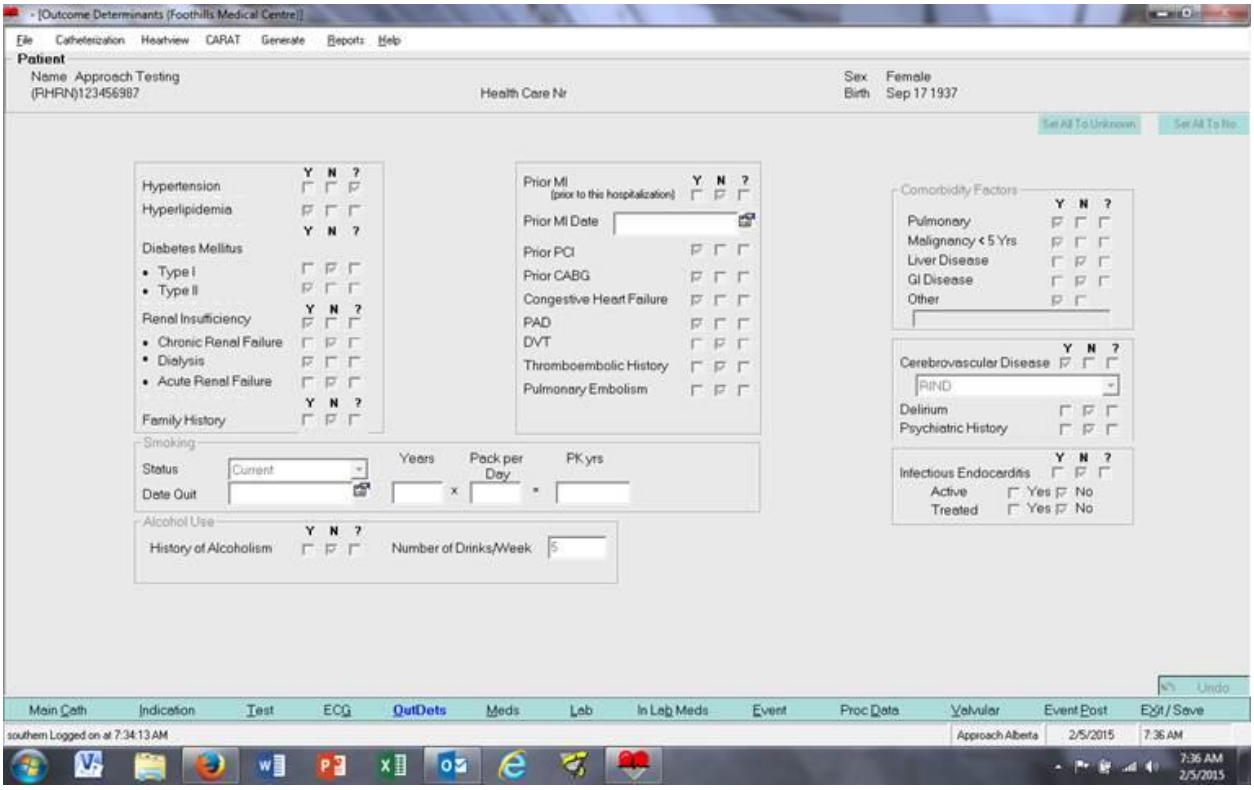
inconclusive. What causes some conditions to have higher and more consistent validity than others?

6.3.1 Differences between DAD and APPROACH in terms of data sources

There are a few fundamental differences with the way data is collected that may cause some of the differences in coding between the two data sources. First is the focus of the database. The physician that sees the patient during their APPROACH visit may not be the same physician that sees the patient during their hospital stay. These two physicians may have different priorities when assessing the patient. During the APPROACH visit, the physician is most interested in the catheterization procedure as well as accurately determining patient comorbidities. These items are a priority for the APPROACH physician, because he/she knows that the data collected in APPROACH could potentially be used in research at some point down the road. The priority of the physician that sees the patient during their inpatient stay, is simply treating the patient. This means that the physician will focus on making diagnosis that are the most important to the admission, and/or has the largest contribution to the patient's length of stay. The inpatient physician is likely to focus on diagnosis of the conditions that are the most important to that hospital stay. Therefore, the APPROACH physician is likely to capture a greater number of conditions.

Another difference between APPROACH and DAD is the data requirements for each patient visit/admission. Comorbidity data in APPROACH is entered using a required form. A screenshot of what this form looks like is included below.

Figure 26. APPROACH Screen Shot



As you can see the form lists all comorbidities that it wishes the clinician to collect. The clinician simply has to enter yes, no, or unsure for each comorbidity. DAD is collected from information in medical charts, and comorbidity information in medical charts do not have the same requirements. The number of comorbidities recorded on medical charts is simply left up to the discretion of the physician [67].

These two differences in the data collection methods of the two databases, means that more comorbidities are likely to be entered in APPROACH, and the comorbidity information in APPROACH is likely to be more complete. This is one of the reasons why DAD validity for some conditions is low when compared to APPROACH.

6.3.2 Physician Documentation Patterns

Having established that the data collection differences between APPROACH and DAD leads to more complete comorbidity information in APPROACH, what causes some DAD conditions to compare well with APPROACH data, while other conditions do not? One of the reasons is how physician diagnosis is entered in DAD.

As mentioned in the previous section, inpatient physicians have different priorities compared to physicians in APPROACH. Therefore, comorbidity information in medical charts is generally comorbidities that have the most impact on the admission, treatment, or the patient's length of stay. Further, inpatient physicians generally face time constraints due to heavy workloads, meaning that they have limited time to document clinical information on each chart. Therefore, they often do not have time to include all comorbidities and will prioritize based on relevancy to the current admission [68]. Also, even when a large number of comorbidities are recorded, conditions which are less common or less critical to the admission are not recorded on the front page or the summary page. For example, diabetes is a condition that usually has a strong correlation with treatment and complications and will often make it onto the front or summary page of a medical chart. However hyperlipidemia is a condition that usually doesn't correlate with significant length of stay and will often be left off the front or summary page. Therefore it is not surprising that DAD coding for diabetes is done well, whereas DAD coding for hyperlipidemia is done poorly. Generally, conditions that are included on the front and summary pages will be coded more accurately [43,69]. This is caused by the habits of ICD coders when reading medical charts.

6.3.3 ICD Coder Practice

The reason that conditions which are included on the front and/or summary page of a chart are more likely to be coded well in DAD is simply because they are more likely to be picked up by ICD coders. Similar to inpatient physicians, coders are trained to place priority on comorbidities that have the most significant impact on patient length of stay. ICD coders are also under time constraints due to high workload volume. As a result they have limited time to spend on each chart, and often will not be able to look too deeply into every individual chart. This means that conditions listed on the front or summary pages are more likely to be picked up by coders [43]. This time constraint also means that coders are likely to develop coding habits. This means that more common conditions which are seen often, such as hypertension or diabetes become more familiar and ends up being coded frequently. Further, they become less familiar with rarer conditions such as liver disease, and end up coding them infrequently [70]. The way physicians write medical charts coupled with the way ICD coders read these charts partially explains why some conditions are coded well in DAD, while some are not.

6.4 Is DAD suitable for use in surveillance research?

Now that we have obtained the validation results, we must consider whether DAD can be used in chronic disease surveillance. The answer depends on what type of surveillance you are doing, and which conditions you will be doing them for.

6.4.1 Magnitude of Prevalence

When the goal of conducting surveillance research is to calculate an accurate prevalence of a certain condition, then the level of DAD validity for that condition is vital. DAD is only useful for this type of research if it's coding for those conditions have a high level of validity.

Therefore out of the nine conditions assessed in this study, only hypertension and diabetes have a suitably high level of validity. When the magnitude of prevalence is the focus, DAD would only be suitable for this type of study if the condition studied has a high degree of validity.

6.4.2 Change in Prevalence

If the focus of the research is on changes only, it is the consistency of DAD validity that is important. Inconsistent data validity will affect the trend of prevalence changes over time. In our study, only hypertension and diabetes were found to have stable validity over time. Thus DAD would be usable when conducting surveillance research looking at the change of prevalence for those conditions.

6.4.3 Magnitude and change of prevalence

When you would like to conduct research that accurately calculates the prevalence of a condition and assesses how that prevalence changes over time, then it is necessary for DAD to have a high level of validity and consistency. In our study, only hypertension and diabetes met these criteria. Therefore, when conducting this type of research, DAD will only be suitable when researching hypertension and/or diabetes.

6.5 Study Strengths

This is the first DAD validation study with a specific focus on the trend of validity over a long time period. While, a large number of studies have validated DAD coding for various conditions, few have used a consistent methodology to investigate the trend of DAD validity over time. In fact, we believe that this is the first study that uses a consistent methodology to investigate the validity of DAD coding for 9 comorbidities over a period of 11 years.

This type of large scale, long term study is made possible by the availability of a large population based clinical database such as APPROACH. The most common reference standard used in previous studies is medical chart review. The issue with chart review is that as of now, the majority of medical charts in Canada still exist in paper form [71]. This means that conducting a chart review is a costly and time consuming process. In fact, it may not be realistically feasible to conduct a chart review on a similar scale, as would've been necessary for this study. The large population available in the APPROACH database, makes this kind of long term validation study possible.

There are a few features of the Canadian health care system that makes this type of study possible. Canada has universal health care coverage, meaning that a database such as DAD is able to contain information on the entire inpatient population [8]. In the Canadian health care system, patients are also assigned unique PHNs. This allows us to link patients across databases with a high degree of accuracy [9]. These two features of the Canadian health care system are essential in our conduction of a large scale data validation study.

6.6 Study Limitations

A limitation with this study is whether the APPROACH database is suitable as a "reference standard". Like many "reference standards", it is not possible to know whether it is completely accurate or true. Some things simply can't be measured with 100% accuracy; patient comorbidity is one of them. Therefore like many other forms of scientific research it is necessary to make assumptions. The main reason that APPROACH should be accurate is that from a methodological standpoint, errors should be minimized. Data is collected in the clinical setting

directly from clinicians. This removes a major source of error that could occur as a result of time lag and translation.

Another limitation of this study is how we dealt with missing data in APPROACH. As mentioned previously, comorbidities in APPROACH can be entered as 0, 1, 2 or the value could be missing. 0 indicates absence of the condition, 1 indicates presence of the condition and 2 indicates that the information is unavailable. If we were to exclude any patient record where one or more comorbidities were missing or entered as 2, our sample size would've dropped from 62 161 to 39837. In order to preserve the sample size, we decided to make the assumption that a missing value or a value of 2 actually represented absence of the condition. We assumed that it would be unlikely for clinicians to not enter a value when the condition was present, and that the vast majority of missing values should've actually been 0. This isn't the cleanest way to conduct this analysis, however we have reason to believe our assumption was correct. We compared the data where all missing values were removed from analysis, versus the data where the missing values were set to 0. We found no significant differences in the prevalence of conditions or on trends of validity over time.

This study contained only patients undergoing cardiac catheterization in Alberta. This presents two potential problems. First, this population naturally has a high prevalence of cardiac related comorbidities, such as hypertension and hyperlipidemia. Therefore, when conducting validation analysis on conditions that have very high prevalence, the PPV will be inflated [72,73]. Second, the characteristics of this population hurts the generalizability of this study. The fact that the cohort of patients in this study is all undergoing the same procedure, means that the study results may not be generalizable to the general population or perhaps even the broader inpatient population. Further, because the study population only includes patients from Alberta,

the study results may not be generalizable elsewhere [74]. This is especially the cause in other countries, where coder guidelines, physician habits, and even the structure of how discharge data is entered, may differ from that in Canada [75].

6.7 Future Research

While this study presents a good start towards the study of time trends of data validity, there are a few areas that would be useful for future study. First, it would be interesting to see how data validity is for a different patient population. Possibilities include, a more general inpatient population rather than just those undergoing cardiac catheterization, or perhaps an even broader population such as the Canadian general public population. It would also be interesting to assess the data validity in other countries or regions. The limiting factor for what population can be researched is the reference standard. In this study we are limited to cardiac patients receiving catheterization, because we are using APPROACH as our reference standard. Therefore in order to explore other populations we must find different reference standards. Possibilities include other clinical databases, or perhaps electronic medical records (EMRs).

As mentioned previously, DAD is only one of the administrative databases that are commonly used when calculating prevalence for chronic disease surveillance. This study only validated DAD. Thus, it only contains an inpatient population. Future research should look at data validity for outpatient populations as well. One possibility is to explore the validity of the physician claims database in Canada. In most Canadian provinces, physician submit claims using ICD diagnosis codes [76]. Further, a recent study by Cunningham et al. evaluated the quality of physician claims data and concluded that the data had high face validity [77]. Therefore, this data source could be a valuable tool for research towards the chronic disease prevalence of the

outpatient population in Canada. Research into the time trend of its validity would be extremely valuable to chronic disease surveillance research.

In this study we have analyzed validity of an inpatient database that is used for chronic disease surveillance among patients who underwent cardiac catheterization. Future research on other sources of data and on other populations will allow us to comprehensively assess chronic disease surveillance and its trends over time.

Chapter 7: Conclusions

To the best of our knowledge this is the first study looking at DAD coding validity in Canada, with a focus on long term trends. There were a number of factors that made conducting this study possible. First, Canada has a universal health care system. This means that the entire population is enrolled into the public health plan. Therefore DAD is able to capture all inpatient hospitalizations. Second, patients are identifiable using unique PHNs. This means that linkage can be conducted with a high degree of accuracy. Lastly, APPROACH is a large powerful clinical population-based database, which allows us to assess DAD validity over a long time period.

The key findings of this study are:

1. DAD validity varied depending on the condition. Conditions such as hypertension and diabetes were coded well and heart failure's coding was mediocre, while hyperlipidemia, PVD, CEVD, COPD, malignancy, and liver disease were all coded poorly. There is a trend where common conditions are coded well, while relatively rare conditions are coded poorly.
2. The trend of validity over time also varied depending on condition. None of the conditions displayed obvious increasing validity in the study period. Validation trends for hypertension and diabetes were mostly stable. Hyperlipidemia, PVD, CEVD and COPD all had declining validity over time. Our findings regarding the validation trends for malignancy, heart failure and liver disease were inconclusive.

These results have several implications for surveillance research that uses DAD to assess prevalence trends over time. DAD validity of the nine comorbidities analyzed in this study can

be classified into three groups: consistent strong validity, inconsistent weak validity, and inconclusive weak validity

Conditions that have consistent strong validity such as hypertension and diabetes, make it possible to use DAD when calculating the prevalence of these conditions. A high level of validity means that the number of cases identified will be fairly accurate. DAD can also be used to assess the time trend of the prevalence of these conditions. Changes in the prevalence of these conditions will be representative of real changes, since data quality has remained consistent over time.

For hyperlipidemia, PVD, CEVD, and COPD where the coding validity is weak, and the trend of validity is inconsistent, DAD is not useful for calculating prevalence, or changes in prevalence. Low coding validity means that calculated prevalence would be inaccurate. Further because the trend of coding validity for these conditions is inconsistent, it is likely that inconsistent data quality is causing prevalence changes that are not representative of real changes in disease status.

We found malignancy, heart disease and liver disease to have moderate to low coding validity and inconclusive trend of validity. This means that DAD would not be appropriate for calculating the prevalence of these conditions. Due to the low validity, any calculated prevalence will be incorrect. Further, because we cannot determine whether data quality is consistent for these two conditions, DAD should also not be used to assess prevalence changes over time.

References

1. Mirolla M. The cost of chronic disease in Canada. Glen Haven, NS:, The Chronic Disease Prevention Alliance of Canada; 2004.

2. PHAC. Preventing Chronic Disease Strategic Plan 2013-2016. [Online].; 2013 [cited 2015 May 10. Available from: http://www.phac-aspc.gc.ca/cd-mc/diabetes-diabete/strategy_plan-plan_strategique-eng.php.

3. PHAC. Report from the Chronic Disease Surveillance System: Hypertension in Canada, 2010. Ottawa: Public Health Agency of Canada, Chronic Disease Surveillance Division; 2010.

4. PHAC. Diabetes in Canada: Facts and figures from a public health perspective. Ottawa: Public Health Agency of Canada, Chronic Disease Surveillance and Monitoring Division; 2011.

5. Elmslie K. Against the Growing Burden of Disease. Ottawa: Public Health Agency of Canada, Centre for Chronic Disease Prevention; 2012.

6. PHAC. Report from the Canadian Chronic Disease Surveillance System. [Online].; 2010 [cited 2015 April 27. Available from: http://www.phac-aspc.gc.ca/cd-mc/cvd-mcv/ccdss-snsmc-2010/2-1-eng.php.

7. University of Manitoba. Population Health Research Data Repository: Administrative Health Databases. [Online].; 2011 [cited 2015 4 29. Available from: http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/repository/health_admin.html.

8. Health Canada. Canada's Health Care System (Medicare). [Online].; 2010 [cited 2015 May 7. Available from: http://www.hc-sc.gc.ca/hcs-sss/medi-assur/index-eng.php.

9. Rotermann M, Sanmartin C, Carriere G, Trudeau R, St-Jean H, Adelnasser S, et al. Two approaches to linking census and hospital data. Health Reports. 2014 Oct; 25(10): p. 3-14.

10. CIHI. Data Quality Documentation for External Users: Discharge Abstract Database, 2010-2011. Ottawa, Ontario:, Canadian Institute for Health Information; 2011.

11. CIHI. Data Quality Documentation, Discharge Abstract Database - Multi-Year Information. Ottawa, ON:, Canadian Institute of Health Information; 2012.

12. Moriyama I, Loy R, Robb-Smith A. History of the statistical classification of diseases and cuases of death. Hyattsville, MD:, National Center for Health Statistics; 2011.

13. WHO. International Classification of Diseases (ICD). [Online]. [cited 2015 May 7. Available from: http://www.who.int/classifications/icd/en/.

14. CIHI. Final Report The Canadian Enhancement of ICD-10. Ottawa, Ontario:, Canadian Institute for Health Information; 2001.

15. CIHI. ICD-10-CA/CCI Implementation Schedule. [Online]. [cited 2015 May 7. Available from: http://www.cihi.ca/CIHI-ext-portal/internet/en/document/standards+and+data+submission/standards/classification+and+coding/codingclass_icd10status.

16. University of Manitoba. Term: International Classification of Diseases, 10th Revision, with Canadian Enhancements (ICD-10-CA). [Online].; 2015 [cited 2015 May 7. Available from: http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?definitionID=102937.

17. Iezzoni L. Assessing Quality Using Administrative Data. Annals of Internal Medicine. 1997; 127(8): p. 666-674.

18. Kane R, Radosevich D. Conducting Health Outcomes Research. In.: Jones & Bartlett Learning; 2011. p. 291-294.

19. Quan H, Eastwood C, Cunningham C, Liu M, Flemons W, De Coster C, et al. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). BMJ Open. 2013; 3(10): p. 003716.

20. Skinner B, Rovere M. Canada's Medicare Bubble: Is Government Health Spending Sustainable without User-based Funding? , Fraser Institute; 2011.

21. CIHI. Measuring the Level and Determinants of Health System Effciency in Canada. Ottawa, ON:, Canadian Institute for Health Information; 2014.

22. Schoenman J, Sutton J, Kintala S, Love D, Maw R. The Value of Hospital Discharge Databases. University of Chicago, NORC; 2005.

23. Rouse M. garbage in, garbage out (GIGO). [Online].; 2008 [cited 2015 May 10. Available from: http://searchsoftwarequality.techtarget.com/definition/garbage-in-garbage-out.

24. Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, Hill MD, Ghali WA, McAlister FA; Hypertension Outcome and Surveillance Team of the Canadian Hypertension Education Programs. Validation of a case definition to define hypertension using administrative data. Hypertension. 2009 Dec; 54(6): p. 1423-8.

25. Drosler S, Romano P, Tancredi D, Klazinga N. International Comparability of Patient Safety Indicators in 15 OECD Member Countries: A Methodological Approach of Adjustment by Secondary Diagnoses. Health Services Research. 2012 Feb; 47(1): p. 275-292.

26. APPROACH. More About APPROACH. [Online].; 2014 [cited 2015 April 27. Available from: http://www.approach.org/about_pages/more_about.html.

27. Knudtson M, Ghali W. Overview of the Alberta Provincial Project for Outcome assessment in Coronary Heart Disease. Canadian Journal of Cardiology. 2000; 16(10): p. 1225-30.

28. CIHI. The CIHI Data Quality Framework. Ottawa, ON:, Canadian Institute for Health Information; 2009.

29. White T, McBurney D. Reliability and Validity of Measurements. In Research Methods. Belmont, CA: Cengage Learning; 2009. p. 129-133.

30. Mostert M. Encyclopedia of Measurements and Statistics: Face Validity. [Online].; 2007 [cited 2015 May 7. Available from: http://srmo.sagepub.com/view/encyclopedia-of-measurement-and-statistics/n164.xml.

31. Carpenter C, Sarli C, Fowler S, Kulasegaram K, Vallera T, Lapaine P, et al. Best Evidence in Emergency Medicine (BEEM) Rater Scores Correlate With Publications' Futre Citations. Academic Emergency Medicine. 2013 Oct; 20(10): p. 1004-1012.

32. Berger P, Luckmann T. The Social Construction of Reality: A Treatise in the Sociology of Knowledge Garden City, NY: Anchor Books; 1966.

33. Patterson P, Lave J, Martin-Gill C, Weaver M, Wadas R, Arnold R, et al. Measuring adverse events in helicopter emergency medical services: establishing content validity. Prehosp Emerg Care. 2014 Jan-Mar; 18(1): p. 35-45.

34. Jette N, Reid A, Quan H, Hill M, Wiebe S. How accurate is ICD coding for epilepsy. Epilepsia. 2010 Jan; 51(1): p. 62-9.

35. Viera A, Garrett J. Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine. 2005 May; 37(5): p. 360-363.

36. Lalkhen A, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia, Critical Care & Pain. 2008; 8(6): p. 221-223.

37. O' Malley K, Cook K, Price M, Wildes K, Hurdle J, Ashton C. Measuring Diagnoses: ICD Code Accuracy. HSR: Health Services Research. 2005 Oct; 40(5): p. 1620-1639.

38. Schneeweiss S, Avorn J. A review of uses of health care utilization database for epidemiologic research on therapeutics. Journal of Clinical Epidemiology. 2005; 58: p. 323-337.

39. Elstein A. Clinical reasoning in medicine. In Higgs J. Clinical reasoning in health professions. Oxford, England: Butterworth-Heinemann Ltd.; 1995. p. 49-59.

40. Graber M. The incidence of diagnostic error in medicine. BMJ Qual Saf. 2013; 22: p. ii21-ii27.

41. Van Walraven C, Weinberg A. Quality Assessment of a Discharge Summary System. Canadian Medical Association Journal. 1995 May; 152(9): p. 1437-1442.

42. Ontario Health. A Guide to Better Physician Documentation: Physician Documentation Expert Panel. Ontario Health Ministry; 2006.

43. Quan H, Parsons G, Ghali W. Validity of information on comoridity derived from ICD-9CCM administrative data. Medical Care. 2002; 40(8): p. 675-685.

44. Quan H, Moskal L, Forster A, Brien S, Walker R, Romano P, et al. International variation in the definition of 'main condition' in ICD-coded health data. International Journal for Quality in Health Care. 2014; 26(5).

45. Wennberg J, Gittelsohn A. Small area variations in health care delivery. Science. 1973; 182: p. 1102-8.

46. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, Ghali WA; IMECCHI Investigators. Assessing the validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditons in a unique dually coded database. Health Serv Res. 2008 Aug; 43(4): p. 1424-41.

47. Walker R, Hennessy D, Johansen H, Sambell C, Lix L, Quan H. Implementation of ICD-10 in Canada: how has it impacted coded hospital discharge data? BMC Health Services Research. 2012; 12(149).

48. Van Walraven C, Bennett C, Forster A. Derivation and validation of a MEDLINE search strategy for research studies that use administrative data. Health services research. 2010 Dec; 45(6p1): p. 1836-1845.

49. Daneshvar P, Forster A, Dervin G. Accuracy of administrative coding in identifying hip and knee primary replacements and revisions. J Eval Clin Pract. 2012 June; 18(3): p. 555-9.

50. De Coster C, Li B, Quan H. Comparison and validity of procedures coded with ICD-9-CM and ICD-10-CA/CCI. Med Care. 2008 June; 46(6): p. 627-34.

51. Allen V, Dodds L, Spencer A, Cummings E, MacDonald N, Kephart G. Application of a national administrative case definition of pre-existing diabetes mellitus in pregnancy. Chronic Dis Inj Can. 2012 June; 32(3): p. 113-20.

52. Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. Diabetes Res Clin Pract. 2010 August; 89(2): p. 189-95.

53. Januel J, Luthi J, Quan H, Borst F, Taffe P, Ghali W, et al. Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. BMC Health Serv Res. 2011 August; 18(11): p. 194.

54. Hagen E, Rekand T, Gilhus N, Gronning M. Diagnostic coding accuracy for traumatic spinal cord injuries. Spinal Cord. 2009 May; 47(5): p. 367-71.

55. Reid A, St Germaine-Smith C, Liu M, Sadiq S, Quan H, Wiebe S, et al. Development and validation of a case definition for epilepsy for use with administrative health data. Epilepsy Res. 2012 Dec; 102(3): p. 173-9.

56. Southern D, Norris C, Quan H, Shrive FGP, Humphries K, Gao M, et al. 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. ; 1.

57. Muhajarine N, Mustard C, Roos L, Young T, Gelskey D. Comparison of survey and physician claims data for detecting hypertension. J Clin Epidemiology. 1997 June; 50(6): p. 711-8.

58. Rector T, Wickstrom S, Shah M, Thomas Greeenlee N, Rheault P, Rogowski J, et al. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. Health Serv Res. 2003 Dec; 39(6): p. 1839-57.

59. Lix L, Yogendran M, Burchill C, Metge C, McKeen N, Moore D, et al. Defining and Validating Chronic Diseases: An Administrative Data Approach. Winnipeg, Manitoba:, Manitoba Centre for Health Policy; 2006.

60. Tu K, Campbell N, Chen X, Cauch-Dudek K, McAlister F. Accuracy of administrative databases in identifying patients with hypertension. Open Med. 2007 Apr; 1(1): p. e18-e26.

61. Alberta Government. Occupational Profile: Health Information Management Professional. [Online]. [cited 2015 May 10. Available from: http://occinfo.alis.alberta.ca/occinfopreview/info/browse-occupations/occupation-profile.html?id=71002536.

62. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic Review and Meta-Analysis of Validation Studies on a Diabetes Case Definition from Health Administrative Records. PLOS one. 2013 Oct; 8(10): p. e75256.

63. Marrie RA, Yu BN, Leung S, Elliott L, Caetano P, Warren S, Wolfson C, Patten SB, Svenson LW, Tremlett H, Fisk J, Blanchard JF; CIHR Team in Epidemiology and Impact of Comorbidity on Multiple Sclerosis. Rising prevalence of vascular comorbidities in multiple sclerosis: validation of administrative definitions for diabetes, hypertension, and hyperlipidemia. Mult Scler. 2012 Sep; 18(9): p. 1310-9.

64. Kokotailo R, Hill M. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. Storke. 2005 Aug; 36(8): p. 1776-81.

65. Quach S, Blais C, Quan H. Administrative data have high variation in validity for recording heart failure. Can J Cardiol. 2010 Oct; 26(8): p. 306-12.

66. Thomsen R, Lange P, Hellquist B, Frausing E, Bartels P, Krog B, et al. Validity and underrecording of diagnosis of COPD in the Danish National Patient Registry. Respiratory Medicine. 2011; 105: p. 1063-1068.

67. College of Physicians & Surgeons of Alberta. Administration of Practice: Patient Records. [Online].; 2014 [cited 2015 May 10. Available from: http://www.cpsa.ab.ca/Resources/StandardsPractice/administration-of-practice/patient-records.

68. Juurlink D, Preyra C, Croxford R, Chong A, Austin P, Tu J, et al. Canadian Institute for Health Information discharge Abstract Database: A Validation Study. Toroton: Institute for Clinical Evaluative Sciences; 2006.

69. Levy A, Tamblyn R, Fitchett D. Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction. Can J Cardiol. 1999; 15: p. 1277-1282.

70. Hennessy D, Quan H, Faris P, Beck C. Do coder characteristics influence validity of ICD-10 hospital discharge data? BMC Health Serv Res. 2010 Apr; 21(10): p. 99.

71. Biro S, Barber D, Kotecha J. Trends in the use of electronic medical records. Canadian Family Physician. 2012 Jan; 58(1): p. e21.

72. Sanghavi D. Letter to the Editor: Early Identification of Cerebral Palsy Using a Profile of Abnormal Motor Patterns. Pediatrics. 1997 Nov; 100(5): p. 902.

73. Eastwood K, Else P, Charlett A, Wilcox M. Comparison of Nine Commercially Available Clostridium difficile Toxin Detection Assays, a Real-Time PCR Assay for C. difficile tcdB, and a Glutamate Dehydrogenase Detection Assay to Cytotoxin Testing and Cytotoxigenic Culture Methods. J Clin Microbiol. 2009 Oct; 47(10): p. 3211-3217.

74. Quan H, Smith M, Bartlett-Esquilant G, Johansen H, Tu K, Lix L; Hypertension Outcome and Surveillance Team. Mining administrative health databases to advance medical science: geographical considerations and untapped potential in Canada. Can J Cardiol. 2012 Mar-Apr; 28(2): p. 152-4.

75. Oderkirk J, Ronchi E, Klazinga N. International comparisons of health system performance among OECD countries: Opportunities and data privacy protection challenges. Health Policy. 2013 Sept; 112(1-2): p. 9-18.

76. Lix LM, Walker R, Quan H, Nesdole R, Yang J, Chen G; CHEP-ORTF Hypertension Outcomes and Surveillance Team. Features of physician services databases in Canada. Chronic Dis Inj Can. 2012 Sep; 32(4): p. 186-93.

77. Cunningham C, Cai P, Topps D, Svenson L, Jette N, Quan H. Mining rich health data from Canadian physician claims: features and face validity. BMC Res Notes. 2014 Oct; 1(7): p. 682.