University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2015-02-10

Validation of an ICD-10 coded case definition for the identification of patients diagnosed with sepsis and severe sepsis using administrative data

Jolley, Rachel

Jolley, R. (2015). Validation of an ICD-10 coded case definition for the identification of patients diagnosed with sepsis and severe sepsis using administrative data (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/28626 http://hdl.handle.net/11023/2101

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Validation of an ICD-10 coded case definition for the identification of patients diagnosed with sepsis and severe sepsis using administrative data

by

Rachel J Jolley

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

JANUARY, 2015

© Rachel J Jolley 2015

Abstract

Background: We assessed the validity of existing ICD case definitions used to identify sepsis in administrative data and validated and optimized an existing ICD-10-CA coding algorithm to identify patients diagnosed with sepsis.

Methods: Standard systematic review methodology was applied to assess the validity of ICD case definitions for sepsis. The CIHI ICD-10-CA coding algorithm for sepsis was validated and optimized using a randomly selected cohort of ICU and non-ICU patients. Sensitivity (Sn), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) were calculated.

Results: Twelve studies were identified in the systematic review with a range of diagnostic accuracy reported indicating that sepsis is highly under-coded. We increased the accuracy of the CIHI ICD-10-CA coding algorithm for sepsis (Sn: 71.9%, NPV: 66.6%) and severe sepsis (Sn: 65.1%, NPV: 70.1%) while slightly decreasing Sp and PPV.

Conclusions: Sepsis is highly under-coded in administrative data. The new definition has a much higher sensitivity and negative predictive value.

Contributions of Authors

Dr. Hude Quan- Co -author on all written manuscripts, aided in conceptualization of project, training for data abstraction, reviewed results, edited and reviewed final manuscript

Dr. Nathalie Jette - Co-author on all written manuscripts aided in conceptualization of project, training for data abstraction, reviewed results, edited and reviewed final manuscript

Dr. Christopher J Doig – Co-author on all written manuscripts, primary investigator for all articles alongside thesis, provided concept of project, training for data abstraction, review of medical chart review, reviewed and aided in results analysis, edited and reviewed final manuscript

Keri Jo Sawka – Co -author on all written manuscripts, acted as secondary reviewer for both manuscripts, reviewed all titles, abstracts and full-text articles for systematic review, performed secondary data abstraction and quality assessment as well as discussed format and results, acted as chart reviewer and data abstractor for validation study, reviewed and discussed results, edited and reviewed final manuscripts

Dean W Yergens – Provided literature review software training and support for systematic review, reviewed study protocol, edited and reviewed final manuscript

Dr. Bryan G Yipp – Performed initial consensus review including chart review and data abstraction, provided feedback, edited and reviewed final manuscript

Dr. Derek Robert – Performed initial consensus review including chart review and data abstraction, provided feedback, edited and reviewed final manuscript

Lucy Diep – Acted as chart reviewer and data abstractor for validation study, reviewed final manuscript for validation study

Jade Goliath – Acted as chart reviewer and data abstractor for validation study, reviewed final manuscript for validation study

Acknowledgements

I would like to thank my committee members Dr. Hude Quan and Dr. Nathalie Jetté who were more than generous with their expertise, encouragement, support and precious time.

A massive thanks to my supervisor Dr. Chip Doig, for his motivation, support and encouragement throughout this process allowing me to reach great heights with my personal and academic development.

I would also like to thank my dear friends Kelsey, KJ, Lucy and Dean for their moral support in the form of coffee breaks, pizza dinners, and general motivational and life chats. And to my Geneva family George, Annika, Pete, Line, Kieren, Tori and Clio who were there to pick me up when I needed it most and inspired me to keep going.

To my mom, for all of her love, support, perseverence, and encouragement

And to Glenn Thomas – reach for the stars

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
ACS-NSQIP	American College of Surgeons National Surgical Quality Improvement Program
CAP	Community Acquired Sepsis
CDC	Centre for Disease Control
CM	Clinical Modification
DI	Department of Infectious Disease
ED	Emergency Department
FP	False Positive
FN	False Negative
ICD	International Classification of Disease
ICD-10-CA	International Classification of Diseases, Tenth Revision, Canadian modification
ICU	Intensive Care Unit
MODS	Multiple organ dysfunction syndrome
MOF	Multiple organ failure
NPV	Negative Predictive Value
PPV	Positive Predictive Value
SIRS	Systemic Inflammatory Response Syndrome
Sn	Sensitivity
Sp	Specificity
TP	True Positive
TN	True Negative
USD	United States Dollar
WHO	World Health Organization

CHAPTER ONE: INTRODUCTION

1.1 Overview

Sepsis is one of the top causes of death in the world and is the leading cause of death in non-coronary intensive care units (ICU's) with a reported mortality of 28%-50% [1], [2]. It affects individuals of all ages, sex, race and socioeconomic status with increasing incidence and hospitalization rates [3] and high costs of care estimated to be up to $17 billion USD per year [4]. As well, new studies are demonstrating the detrimental impact that sepsis has on survivors in many aspects of life including functioning and cognitive abilities [5], indicating the burden of sepsis does not end upon leaving the ICU.

Reporting the epidemiology and burden of disease for particular conditions such as sepsis is important to uncover the underlying pathogenesis and mechanism of disease. As well, improving the design and administration of treatment interventions, the overall quality of care and health resource utilization is dependent on effective disease surveillance. In order to report and assess diseases, administrative health data is used as it covers a wide variety of health information and is routinely collected on large segments of the population [6]. In Canada, some administrative health data records diseases using the *International Classification of Diseases, Tenth Revision, Canadian modification* (ICD-10-CA) codes [7].

The accuracy of capturing disease through ICD codes may be dependent on multiple factors including both clinical and health care data coded case definitions. In the case of sepsis, a validated clinical case definition is contentious, sepsis is a difficult to define and complex condition that describes a continuum of a heterogeneous array of symptoms with multiple etiologies (both causal organisms and the location of infection) [8]. As well, many studies have

cited the use of previously defined coding algorithms to capture a study cohort of sepsis patients. However these coding algorithms were not validated utilizing a medical chart review reference standard (considered the gold standard for determining diagnostic accuracy). Some of these coding algorithms were also based on the use of the 9th Revision of the ICD codes (ICD-9) and long since has the ICD-10 version been introduced along with newly updated clinical definitions of the sepsis syndrome, thereby out-dating and complicating the translation of the previously used ICD codes. Both the lack of a reliable clinical case definition, standard ICD coded algorithm and the introduction of the new version of ICD coding in administrative data have lead to difficulties in case ascertainment as seen in the multitude of studies using differing methods to identify patients [9].

Another issue arises in that many septic patients who become a major focus of treatment and intervention strategies end up in the ICU where the diagnosis of sepsis becomes more distinguishable. Those patients in non-ICU settings may not be captured in administrative data if the diagnosis is difficult to recognize and not a major reason for admission, leading to the possibility of gross underestimation of the occurrence of sepsis. Therefore the accurate identification of cases of sepsis using ICD-coding in both ICU and non-ICU populations is necessary to contribute to high quality data used for surveillance and reporting disease burden.

1.2 Study Purpose

The purpose of the study was to estimate the accuracy of an ICD-coded case definition for identifying sepsis patients using Canadian administrative health data. The study also provides a methodological basis for future studies to acquire cohorts of sepsis patients specifically aimed at health services research.

1.3 Study Hypotheses

The study hypotheses are:

1. There will be validated case definitions of sepsis, but none with a high degree of accuracy within a generalizable population.

2. The existing Canadian ICD coded case definition will have low accuracy.

3. It will be possible to optimize an existing ICD-coded case definition for sepsis increasing the degree of sensitivity and thereby increasing the accuracy.

1.4 Study Objectives

Based on the rationale above, there are four objectives we hope to fulfill by conducting this study. These objectives are:

1. To review the literature to assess published studies examining the validity of administrative data ICD-9/ICD-10 coded case definitions for identifying patients with sepsis and severe sepsis.
2. To estimate the accuracy of an existing ICD-10 coded case definition for sepsis in Canadian administrative health data in Calgary, Alberta.
3. To optimize the performance of the existing ICD-10 coded case definition (creating a new algorithm).
4. To compare the validity of the existing and the new ICD-10 coded case definitions in both ICU and non-ICU patient populations in Calgary, Alberta.

CHAPTER TWO: LITERATURE REVIEW

2.1 Evolving Definition of Sepsis

2.1.1 Historical definition of sepsis

Sepsis has been an evolving entity ever-since it was first described. The term sepsis was derived from the Greek word *sipsis* meaning “to decay” closely resembling the concept of putrefaction first described by Hippocrates (ca. 460-370 BC) [10]. Only in 1914 was a more formal definition given to the condition of septicemia described by Schottmueller as "a state of microbial invasion from a portal of entry into the blood stream which causes signs of illness” [11]. Since this time, the term sepsis has evolved to loosely describe a group of infectious diseases where the systemic invasion of bacteria or other organisms in the blood cause a spectrum of bodily responses from mild bodily dysfunction to multiple organ dysfunction syndrome (MODS) and multiple organ failure (MOF). Sepsis is unique in that a multitude of organisms may interact with the host in multiple ways and the host response is a multifactorial process involving innate and adaptive immunity.

2.1.2 ACCP/SCCM Consensus Conference Definitions

The lack of a concise clinical definition for sepsis and the frequent interchangeable use of terms describing sepsis in clinical settings such as “sepsis”, “bacteremia” and “septicemia” [12] resulted in the formation of a distinct definition for the term sepsis by Bone *et al.* [13]. Further to this, the common usage of the terms septic shock and septic syndrome without the agreement of which signs and symptoms should be included in the definitions for each of the terms used, lead to the American College of Chest Physicians/ Society of Critical Care Medicine (ACCP/SCCM) Consensus Conference in 1992 [14]. The goal of this meeting was to agree on a set of clinical

definitions and standardization of terminology for sepsis and the associated spectrum of septic states including systemic inflammatory response syndrome (SIRS), sepsis, severe sepsis, septic shock, infection, sepsis-induced hypotension and multiple organ dysfunction [14]. Sepsis refers to the "systemic inflammatory response in the presence of an infection" [14] while severe sepsis is defined as sepsis plus the presence of organ dysfunction, hypoperfusion abnormality or sepsis-induced hypotension (Table 1). Although not a "gold" standard, the terms that arose out of this conference became the basis for clinically defining the different manifestations of sepsis and differentiating between sepsis and SIRS [16].

Table 1: The International Consensus Conference Distinctions in Definition of Severe Sepsis (1991 ACCP/SCCM Consensus Conference Committee definitions)
Adapted from Bone *et al.* Chest, 1992 [14].

SIRS	Systemic Inflammatory Response Syndrome (a body-wide inflammatory response)
Sepsis	SIRS caused by suspected or proven infection
Severe Sepsis	Sepsis with acute organ dysfunction, hypoperfusion or hypotension (organ dysfunction refers to the presence of altered organ function in an acutely ill patient such that homeostasis cannot be maintained without intervention
Septic Shock	Sepsis induced hypotension despite adequate fluid resuscitation along with presence of hypoperfusion abnormalities or organ dysfunction

In 2005, Trzeciak *et al*. [17] showed that there was a significant increase in clinical trials with inclusion criteria based on the 1992 consensus conference definitions. Alongside clinical trials, sepsis has been reported and studied in widespread population-based and hospital-based epidemiological studies, many of which retrospectively identify cases incorporating the 1992 consensus terms and symptomatology into their case definitions. However the use of the 1992 case definitions has been criticized for its over-sensitivity and lack of case specificity meaning the inclusion of symptoms were very sensitive to sepsis but also could not rule out other disease [18].

In 2001, the terms of sepsis were updated to reflect the growing body of knowledge regarding disease etiology, pathology and treatment and to be useful not only to clinicians but as well to researchers [19]. This conference established definitions and more specific clinical presentation characteristics for the different stages of sepsis. Although the terms were similar, the updated definitions expanded the lists of signs and symptoms for identifying and diagnosing SIRS and sepsis, specifically the SIRS criteria into a list consisting of 7 general, 5 inflammatory, 3 hemodynamic, 7 organ dysfunction, or 2 tissue perfusion criteria needing to be present (Table 2). Although these conference definitions have improved the understanding of the etiology and clinical definition of sepsis, diagnosing the disease clinically and capturing the disease in data is still complex and has lead to an ongoing challenge in accurately reporting the burden of disease.

Table 2: The clinical diagnostic criteria of sepsis taken from Levy M.M *et al.* "2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference"

Infection (Infection defined as a pathologic process induced by a microorganism) - documented or suspected, and some of the following:	
General variables:	Fever (core temperature >38.3°C)
	Hypothermia (core temperature <36°C)
	Heart rate >90 min^{-1} or >2 SD above the normal value for age
	Tachypnea
	Altered mental status
	Significant edema or positive fluid balance (>20 mL/kg over 24 hrs)
	Hyperglycemia (plasma glucose >120 mg/dL or 7.7 mmol/L) in the absence of diabetes
Inflammatory variables	Leukocytosis (WBC count >12,000 μL^{-1})
	Leukopenia (WBC count <4000 μL^{-1})
	Normal WBC count with >10% immature forms
	Plasma C-reactive protein >2 SD above the normal value
	Plasma procalcitonin >2 SD above the normal value
Hemodynamic variables	Arterial hypotension (SBP <90 mm Hg, MAP <70, or an SBP decrease >40 mm Hg in adults or <2 SD below normal for age)
	SvO_2 >70%
	Cardiac index >3.5 $L*min^{-1}*M^{-23}$
Organ dysfunction variables	Arterial hypoxemia (PaO_2/FIO_2 <300)
	Acute oliguria (urine output 0.5 $mL*kg^{-1}*hr^{-1}$ or 45 mmol/L for at least 2 hrs)
	Creatinine increase >0.5 mg/dL
	Coagulation abnormalities (INR >1.5 or aPTT >60 secs)
	Ileus (absent bowel sounds)
	Thrombocytopenia (platelet count <100,000 μL^{-1})
	Hyperbilirubinemia (plasma total bilirubin >4 mg/dL or 70 mmol/L)
Tissue perfusion variables	Hyperlactatemia (>1 mmol/L)
	Decreased capillary refill or mottling

2.2 The Epidemiology of Sepsis

2.2.1 Incidence and mortality

Epidemiological estimates of sepsis have varied dependent on the case definition and type of data used however the incidence of sepsis has been reported to be on the rise worldwide. One of the first large-scale epidemiological studies of sepsis in the United States by Angus *et al*. [4] in 2001 reported the incidence of sepsis as 3 cases per 1000 population or 751,000 cases per year. Another major study by Martin *et al.* [1] described the rise in sepsis incidence over a 22-year period of 8.7%. In June 2011, the Center for Disease Control (CDC) reported hospitalization rates for sepsis had doubled over an eight-year period from 2000 (11.6 per 10 000) to 2008 (24.0 per 10 000) [20] with sepsis-related hospitalizations increasing from 326,000 in 2000 to 727,000 in 2008. A study done by the Canadian Institute for Health Information (CIHI) had reported approximately 30,587 sepsis hospitalizations occurred in 2008-2009 (outside of Quebec) [21].

Infection-related deaths are a leading cause of morbidity and mortality with sepsis having been listed as one of the top fifteen leading causes of death by the CDC in 2010 [22]. Approximately 150,000 people die annually in Europe and 200,000 die annually in the United States from severe sepsis [4] with the crude mortality rate reported to range from 28% to 50% [23], [24] . Canada has reported in-hospital mortality due to sepsis to be at 38.1% in 2009 [25].

2.2.2 Long-term impact for survivors

Emerging evidence has indicated that successful treatment and survival of sepsis does not end upon leaving the ICU. Survivors of sepsis have been shown to have significant and persistent cognitive decline and functional impairment similar to that of dementia [26]. Iwashyna *et al*. [27] showed that survivors of severe sepsis were 3.34 times more likely to have cognitive

impairments and a recent study by Odden *et al*. [28] reported a high degree of functional disability in severe sepsis survivors, even those not admitted to the ICU. Sepsis survivors have a higher risk of cardiovascular events [29] and subsequent infections within the first year post-hospitalization with a high risk of admission [30].

2.2.3 Costs and resource use associated with sepsis

With increasing incidence and mortality of sepsis, the costs attributed to caring for a sepsis patient can be one of the highest priced care-types offered as many of these patients are generally treated in the ICU where care from a multi-disciplinary ICU team and multiple interventions may be given. Annual costs for a sepsis patient have been estimated at $17 billion per year in the US [4]. The mean costs of care in Quebec has been estimated at $11 474 per episode of care with annual costs ranging from $36.4 to $72.9 million per year [31] for each hospitalization. As well, the costs of care for a survivor of severe sepsis have been estimated at CAD $20, 859 during the first year post-hospital discharge, CAD $7,145 during year 2 and CAD $7099 during year three [32].

2.3 Administrative data and disease surveillance for sepsis

2.3.1 Administrative data

Administrative data is collected through the routine administration of programs and is primarily for specific decision-making administrative purposes. For health care, the health system collects data for administrative purposes such as tracking in-patient visits, outpatient or community care, physician claims and billing. For health related studies, the use of administrative health data has been implemented in multiple ways, one of which is to capture disease-based cohorts using health administrative coded case definitions. This type of data

however was not originally intended for research but has been used for this purpose since it covers large populations, is relatively inexpensive and readily available [33].

2.3.2 Defining disease in administrative data

To define diseases in administrative data, the World Health Organization's (WHO) International Classification of Diseases (ICD) can be used. The ICD was originally created and adopted to monitor and compare mortality statistics and causes of death but over time has expanded to include codes for all diseases as well as codes for signs and symptoms and external causes of injury or diseases [34], [35]. The most recent update of the ICD was released in 1994, from the *International Classification of Diseases, Ninth Revision,* (ICD-9) to the *International Classification of Diseases and Related Health Problems, Tenth Revision* (ICD-10). One of the major advantages of ICD-10 is that it contains an alphanumeric classification system with a code size ranging from 3 to 6 characters starting with a letter followed by the minimum 2 digits. It is far more detailed (there are a total of 12,420 codes in ICD-10 compared to 6,969 in ICD-9) permitting richer and more precise capture of clinical information [36].

The transition from ICD-9 to ICD-10 was adopted by many jurisdictions along with country-specific modifications. However multiple countries did not immediately adopt it and in some cases are still in the process of implementing ICD-10. Countries such as the United States (in which many of the epidemiological studies of sepsis have been performed) have been lagging on the uptake of ICD-10 resulting in difficulty for international comparisons. For sepsis, the translation for some of the codes between the ICD-9 and ICD-10 version are presented in Table 3. There are multiple ICD-10 codes presented for certain former ICD-9 codes as well as more specific diagnosis attached to each code.

Table 3: Comparison of ICD-9-CM and ICD-10 codes for sepsis

ICD-9 Code	Nomenclature	ICD-10 Code	Nomenclature
038.0	Streptococcal septicemia	A40.0	Sepsis due to Streptococcus, group A
		A40.1	Sepsis due to Streptococcus, group B
		A40.2	Sepsis due to Streptococcus, group D
		A40.8	Other streptococcal sepsis
		A40.9	Streptococcal sepsis, unspecified
038.2	Pneumococcal septicemia [Streptococcus pneumoniae septicemia]	A40.2	Sepsis due to Streptococcus pneumoniae
	No code	A41	Other sepsis
038.11	Methicillin susceptible Staphylococcus aureus septicemia	A41.0	Sepsis due to Staphylococcus aureus
038.19	Other staphylococcal septicemia	A41.1	Sepsis due to other specified staphylococcus (Sepsis due to coagulase-negative staphylococcus)
038.10	Staphylococcal septicemia, unspecified	A41.2	Sepsis due to unspecified Staphylococcus
038.41	Septicemia due to Hemophilus influenza [H. influenza]	A41.3	Sepsis due to Hemophilus influenza
038.3	Septicemia due to anaerobes	A41.4	Sepsis due to anaerobes
038.40	Septicemia due to gram-negative organism, unspecified	A41.5	Gram-negative sepsis, unspecified
	No code	A4188	Other specified sepsis
038.9	Unspecified septicemia	A41.9	Sepsis, unspecified

2.3.3 Defining sepsis in administrative data

The use of administrative data requires the translation of the clinical definition of sepsis to the ICD codes. When defining a cohort of sepsis patients, there are multiple ICD codes that could be used to capture the spectrum of illness. Canada created its own enhancement resulting in the *International Classification of Diseases and Related Health Problems, Tenth Revision, Canadian Enhancement* (ICD-10-CA) and was formally nationally adopted in 2002. This revision expands the content and specificity to conditions and situations that are not diseases relevant for use outside of a hospital setting [37]. The Canadian Institute for Health Information (CIHI) created an ICD-10-CA (Canadian Revision) coding algorithm to define sepsis in administrative data including 49 different ICD codes for sepsis, 28 codes specific to organ dysfunction and 3 procedure codes which are used to define severe sepsis [21].

2.3.4 Validity of administrative data

Determining the validity of information contained in administrative data is an important issue that needs to be addressed when utilizing this data for research purposes. Virnig and McBean [38] proposed that these issues are addressed through efforts aimed at answering one of two questions: "Do diagnoses and procedures contained in administrative data match data from other sources? Are administrative data sufficient or are additional data needed to obtain a complete clinical picture?" The accuracy of a coding definition is dependent on many factors including the codes chosen and population studied. Using ICD codes to study disease can be problematic as lack of specificity, misclassification of diagnosis resulting from human error in recording the diagnosis, and ranges of codes may affect the quality of the data. For many diseases, studies have compared whether administrative data matches other sources of data or a reference or "gold standard" level of data such as a medical record, specifically when selecting a cohort of patients through a diagnostic code.

For sepsis, measuring the extent and impact of the disease has been difficult due to the discrepancies in the clinical definition resulting in large differences in estimates. Different coding algorithms can markedly affect the reported incidence or hospitalization rates reported for a given disease [6]. A recent US study in 2013 compared four different methods of identifying cases of severe sepsis using administrative data (the National Inpatient Hospital Data) from 2004 to 2009 [9]. These included the studies performed by Angus *et al.* [4], Martin *et al*. [1], Dombrovskiy *et al.* [3], and Wang *et al.* [39] with the results showing that up to a 3.5-fold variance depending on method used. The differences in these estimates are reflective of how well the different algorithms capture sepsis and the need for a more consistent definition for comparability.

2.3.5 Disease Surveillance

Disease surveillance has been described as the ongoing systematic monitoring of disease through the collection and analysis of data [40]. Surveillance systems are used to provide descriptive information regarding when and where health problems are occurring and who is affected [41]. A range of health problems such as infectious disease, acute and chronic health conditions, injuries and disabilities can be monitored through surveillance measures. Surveillance can also track health services utilization and program monitoring functions such as types and timing of treatments and interventions or monitoring of health care practices. The definition of the condition of interest defined in a surveillance system will affect the estimate and its comparability both within a country and between countries. There are studies that have examined incidence, prevalence or mortality rates of sepsis in Canada [25], however there is no passive surveillance system set in place in Canada to monitor episodes of sepsis.

2.4. Measuring Diagnostic Accuracy

2.4.1 Measuring accuracy

Diagnostic accuracy examines how well a test can identify a disease in question or how well it can discriminate between a person with a disease and a person without a disease. To determine accuracy, the population in question must be grouped according to either having the disease or not having the disease.

2.4.2 Sensitivity

Sensitivity (Sn) is defined as the ability of a test to correctly identify those with disease or in other terms it is the probability in which a positive result for disease occurs given the disease is present (Probability (T+/Dx+)). It can be measured by dividing the number of true positives (TP) by the sum of the TP and false negatives (FN).

$$Sn = \frac{TP}{TP + FN}$$

2.4.3 Specificity

Specificity (Sp) is the ability of a test to correctly identify those without disease or is the probability in which a negative result for disease occurs given the disease is absent (Probability (T-/Dx-)). It can be measured by dividing the number of true negatives (TN) by the sum of the TN and false positives (FP).

$$Sp = \frac{TN}{TN + FP}$$

2.4.4 Positive predictive value

The positive predictive value (PPV) is the probability the disease is present (Dx+) given a test is positive (T+) (Probability (Dx+/T+). It can be measured by dividing the number of TP by the sum of the TP and FP.

$$PPV = \frac{TP}{TP + FP}$$

2.4.5 Negative predictive value

The negative predictive value (NPV) is the probability the disease is absent (Dx-) given a test is negative (T-) (Probability (Dx-/T-). It can be measured by dividing the number of TP by the sum of the TP and FP.

$$NPV = \frac{TN}{TN + FN}$$

The calculations for all four diagnostic measures described are presented in Table 4 below.

Table 4: Diagnostic measures of a screening test

	Medical Chart Review Reference standard		
Administrative data	Sepsis (+)	Not Sepsis (-)	
Sepsis	(TP)	(FP)	PPV= TP/(TP+FP)
Not Sepsis	(FN)	(TN)	NPV= FN/(FN+TN)
	Sn = TP/(TP+FN)	Sp = TN/(TN+FP)	

Only PPV and NPV are affected by the prevalence of the disease in the population whereas sensitivity and specificity are the given or fixed properties of the test.

CHAPTER THREE: MANUSCRIPT #1

3.0 Validity of Administrative Data in Recording Sepsis: a systematic review

Rachel J Jolley, KeriJo Sawka, Dean W Yergens, Hude Quan, Nathalie Jetté, Christopher J Doig

Submitted manuscript to *Critical Care Medicine*

Formatted for manuscript-based thesis

3.1 Abstract

Background: Administrative data are used to study sepsis in large population-based studies. Validity of these study findings depends largely on the quality of the administrative data source and the validity of the case definition used. We systematically reviewed the literature to assess the validity of case definitions of sepsis used with administrative data.

Methods: Embase and Medline were searched for published articles with International Classification of Diseases (ICD) coded data used to define sepsis. Abstracts and full text articles were reviewed in duplicate. Data were abstracted from all eligible full text articles, including ICD-9 and/or ICD-10 based case definitions, sensitivity (Sn), specificity (Sp), positive predictive values (PPV) and negative predictive values (NPV).

Results: Of 2317 individual studies identified, 12 full text articles met all eligibility criteria. A total of 38 sepsis case definitions were tested, which included over 130 different ICD codes. The most common ICD-9 codes were 038.x, 790.7, and 995.92 and ICD-10 codes were A40.x and A41.x. The PPV was reported in 10 studies, and ranged from 5.6% to 100% with a median of

50%. Other tests of diagnostic accuracy were only reported in some studies. Sn ranged from 5.9% to 82.3%, Sp ranged from 78.3% to 100% and NPV ranged from 62.1% to 99.7%.

Conclusions: Validity of administrative data in recording sepsis varied across individual studies and ICD coded case definitions. Further studies are required to identify harmonized ICD-coded case definitions of sepsis.

3.2 Background

Sepsis is a life-threatening condition associated with a high mortality rate, significant healthcare costs and long-term consequences [1], [4], [27]. It is characterized by a spectrum of severity from mild acute organ dysfunction to multi-organ failure with complex pathophysiologic processes. Differentiating sepsis as a cause of multiple organ dysfunction syndrome (MODS) compared to other acute systemic inflammatory conditions can be difficult [42].

Many large-scale studies have relied on administrative data to identify patients with sepsis [1], [4]. Examples of administrative data include hospital discharge abstract data, emergency visit data and physician claims data. The most common source are hospital discharge abstract data whereby diseases are coded using the World Health Organization's (WHO) International Classification of Disease (ICD) [34]. These data are advantageous as they are readily available and reasonably inexpensive, and can include a large cohort of patients, control for some confounders such as chronic disease [43], and include individual outcomes [33]. The most recent version of the ICD manual in use is the 10th revision or ICD-10. This manual exists alongside country modifications such as ICD-10-CA (Canadian Edition) and ICD-10-AM

(Australian Modification). As well, a modification of the ICD-9 version, i.e. ICD-9-CM "Clinical Modification", is still being used in a number of countries such as the United States [44] and Italy.

Prior to 1992 there was a lack of consensus regarding ICD-based case definitions for sepsis and related conditions. The Centre for Disease Control (CDC) reported sepsis admissions using administrative data in which the term "septicemia" referring to the presence and spread of micro-organisms via circulating blood [45], was used as a case definition and did not fully incorporate the spectrum of illness that was later defined in more detail by the 1992 ACCP/SCCM Consensus Conference clinical definitions [15].

Angus *et al.* [4] performed a large-scale multi-centre epidemiological study that implemented the identification of severe sepsis patients using an ICD-9-based algorithm that required evidence of both an infection and new-onset organ dysfunction during a single hospitalization, thereafter described as the "Angus" implementation coding scheme. The "Angus" implementation is one of the most well-known and highly cited implementations of an ICD-coded case definition for sepsis. This definition was originally validated by the authors through a comparison of aggregate data showing hospital incidence rates and patient characteristics of the cohorts captured through the ICD-9-CM algorithm versus a previous cohort captured through a prospective study of sepsis patients by Sands *et al.* [46] A recent study [47] validated the "Angus" implementation and another well-known algorithm known as the "Martin" implementation [1] using a reference standard based on physician-based medical chart review. The Angus implementation was reported as having a moderate to low sensitivity (Sn) of 50.3% and a positive predictive value (PPV) of 70.7%, while the Martin implementation had a very low Sn of 16.8% but had a high PPV of 97.6%. As such, they concluded that a population of severe

sepsis patients could be captured through administrative data using the Angus case definition but that cases would be underestimated. Studies that examined the performance of ICD coding algorithms to identify other conditions have also highlighted the great variability that exists when multiple codes are used to define a specific condition [48], [49].

There is currently no consensus regarding which ICD-9 or ICD-10 codes should be used to define sepsis in administrative data. A reasonable step towards the harmonization of an ICD-based definition for sepsis is to examine the literature and report the validity of published ICD-coded case definitions in administrative data.

3.3 Methods

3.3.1 Search Strategy

We applied a modification of the search strategy methodology of St. Germaine-Smith [49]. Using the OVID interface, we conducted searches in MEDLINE and EMBASE from 1992 (based on the 1992 publication date of the establishment of definition criteria for sepsis/severe sepsis by ACCP/SCCM) to September 15th, 2013 applying 'Humans' and 'English language' filters. In order to identify studies assessing the diagnostic accuracy of ICD codes for identifying sepsis, the Boolean operator AND was used to combine three search concepts: sepsis, coding and validity. Articles concerning sepsis were sought using the Boolean operator OR to combine Medical Subject Headings (MeSH) term "sepsis" and EMTREE terms relevant to the condition of sepsis including severe sepsis, and septic shock. Articles concerning the concept of coding were sought using the Boolean operator OR to combine the MeSH terms and keyword searches for: administrative data, hospital discharge data, ICD-9, ICD-10, ICD-9xM or ICD-10xM (country versions), medical record, health information, surveillance, physician claims, claims,

hospital discharge, coding, codes. Articles concerning validity were sought using Boolean operator OR to combine the MeSH and keyword searches for the terms validity, validation, case definition, algorithm, agreement, accuracy, sensitivity (Sn), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) (Table 5).

Table 5: Search Strategy terms

Category	Key Words
Sepsis	exp Sepsis/ OR Sepsis.mp OR Septicemia.mp OR exp Shock, Septic/ OR shock, septic.mp OR Septic shock.mp OR blood poisoning*.mp OR bloodstream infection.mp OR exp Bacteremia/ OR Bacteremia.mp OR exp Fungemia/ OR Fungemia.mp OR exp Parasitemia/ OR Parasitemia.mp OR exp Viremia/ OR Viremia.mp
AND	
Coding	exp Health Services Research/ OR health services research.mp OR administrative data.mp OR exp Hospital Records/ OR hospital records.mp OR exp Medical Records/ OR medical record*.mp OR health information.mp OR surveillance.mp OR physician claims.mp OR claims.mp OR hospital discharge.mp OR coding.mp OR codes.mp OR exp "International Classification of Diseases"/ OR international classification of disease.mp OR ICD.mp OR ICD9.mp OR ICD-9.mp OR ICD-9-CM.mp OR ICD10.mp OR ICD-10.mp OR ICD-10-CM.mp OR ICD-10-CA.mp
AND	
Validation	(validity or validation or case definition or algorithm or agreement or accuracy or sensitivity or specificity or positive predictive value or negative predictive value).mp

3.3.2 Study Inclusion

To be eligible for inclusion, articles had to compare the accuracy of ICD-9 or ICD-10 codes for sepsis, severe sepsis, or septic shock in an administrative database to a reference standard. The following diagnostic accuracy measures were abstracted, if provided, from each study: Sn, Sp, PPV and NPV. All bibliographical references were imported into a custom written java software application [50] for improved reference management and data collection. This software called “Synthesis” is described in more detail elsewhere [51]. The title and abstract of each citation identified was screened in duplicate for eligibility by two reviewers (RJJ and KJS); any article selected as meeting eligibility criteria by either or both reviewers was then retrieved and reviewed by the same two authors) for eligibility criteria (articles excluded based

on title and abstract with reasons for exclusion are shown in Appendix C). To determine inter-rater agreement, the Cohen's kappa score was calculated at both the title and abstract review stage and the full text article review stage. All discordant articles at the abstract review stage went on to full text review. Any discordant full text articles were reviewed a second time and further disagreements about study eligibility at the full text review stage were resolved through discussion until full consensus was obtained.

3.3.3 Data Extraction and Quality Assessment

One author (RJJ) abstracted data from included studies using the standardized abstraction form, including, country location of study, years of data collection, validation database, sample size, and type of sample population. The validated ICD codes and algorithms, diagnostic field position and ICD version used from each study were recorded along with Sn, Sp, NPV and PPV. The authors calculated Sn or Sp in cases where these values were not reported but raw data were available to calculate them.

The included studies were assessed for quality by two reviewers (KJS and RJJ) using a standardized validation study quality checklist, adapted from Benchimol *et al.* (2011) [52] (Table 6). In instances where it was unclear whether a checklist item was fulfilled by the study, it was marked as uncertain. Any discrepancies between the two reviewers were resolved through discussion.

This study was reviewed and approved by the Conjoint Health Research Ethics Board at the University of Calgary.

3.4 Results

3.4.1 Study Characteristics

Of 2317 abstracts reviewed, 96 fulfilled eligibility criteria for full text review. Amongst these articles, the kappa score for inter-rater agreement was 0.87 resulting in a near perfect agreement [53]. Twelve articles met all eligibility criteria and were included in the study [47], [54-64] (Figure 1). Characteristics of the studies are shown in table 8. All 12 studies examined hospital discharge abstract data (also called inpatient administrative health data or inpatient claims administrative dataset). Eight of the 12 studies were performed in the United States [47], [54], [56], [58], [60], [62-64], one in Australia [57], one in Denmark [59], one in Sweden [55] and one in Canada [61]. Publication dates ranged from 1998 to 2014. Seven studies examined ICD-9-CM version codes, one examined only ICD-9, one examined both ICD-9 and ICD-10 version codes, one study examined ICD-10, one study examined ICD-10 Danish version and one study examined ICD-10-AM (Australian Modification) codes. The studies varied considerably in sample size (ranging from n=34 to n=4181) and had heterogeneity in patients studied including highly selective populations (rheumatoid arthritis) or sepsis clinical trial patients to ICU specific, general-medical or surgical patients. The clinical definition of sepsis varied across studies but generally followed the ACCP/SCCM consensus conference definition's clinical criteria closely [19].

After applying the standardized quality assessment checklist to each of the twelve included studies, the tallied scores ranged from 10 to 30, indicating variable quality among studies (Table 6).

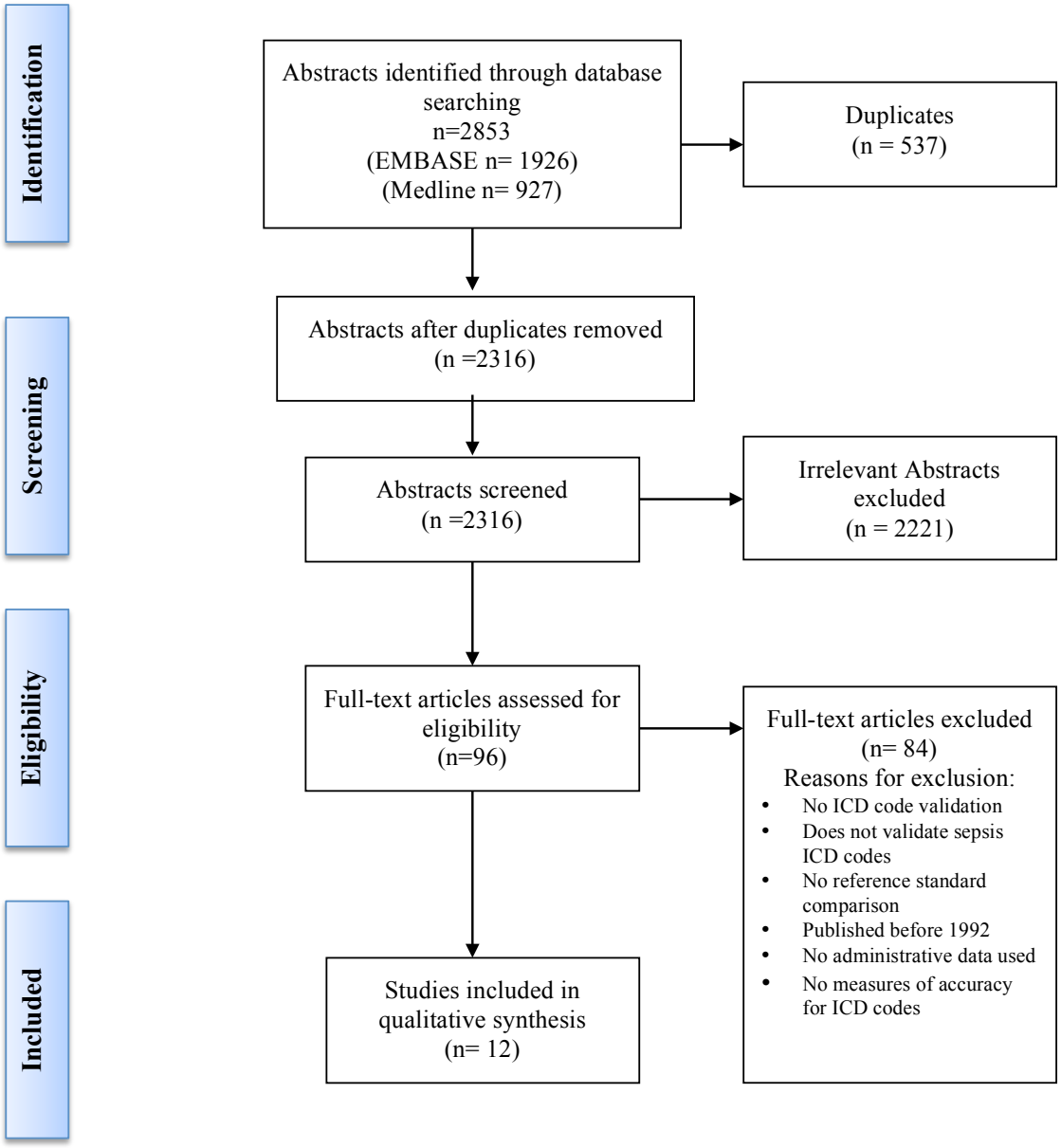

Figure 1. Flow diagram for study screening and article inclusion

Table 6: Quality Assessment checklist of reporting criteria for validation studies of health administrative data (adapted from Benchimol *et al*., 2011 [52])

	Cevasco et al.	Gedeborg et al.	Girjalva et al.	Ibrahim et al.	Lawson et al.	Madsen et al.	Ollendorf et al	Schneeweiss et al.	Quan et al.	Iwashyna et al	Ramanathan et al.	Whittaker et al.
1. Identify article as study of assessing diagnostic accuracy	1	1	1	1	1	1	1	1	1	1	1	1
2. Identify article as study of administrative data	1	1	1	1	1	1	1	1	1	1	1	1
3. State disease identification & validation as goals of study	1	1	1	1	1	1	1	1	1	1	1	1
METHODS: *Participants in validation cohort:*												
4. Age	1	0	1	1	1	0	0	0	1	1	1	1
5. Disease	1	1	1	1	1	1	1	1	1	1	1	1
6. Severity	1	1	1	1	1	1	1	1	1	1	1	1
7. Location/Jurisdiction	1	1	1	1	1	1	1	1	1	1	0	1
8. Describe recruitment procedure of validation cohort	1	0	1	1	1	1	0	1	1	1	1	1
9. Inclusion criteria	1	0	1	1	1	1	0	1	1	1	1	1
10. Exclusion criteria	1	1	1	1	1	1	0	1	1	1	0	1
11. Describe patient sampling (random, consecutive, all, etc.)	1	1	1	1	1	1	0	1	1	1	1	1
12. Describe data collection	1	1	1	1	1	1	0	1	1	1	1	1
13. Who identified patients and did selection adhere to patient recruitment criteria	1	0	1	1	1	1	0	1	1	1	1	1
14. Who collected data	1	0	1	1	1	1	0	1	1	1	1	1
15. A priori data collection form	1	0	1	1	1	1	1	1	1	1	0	1
16. Disease classification	1	1	1	1	1	1	1	1	1	1	1	1
17. Split sample (i.e. re-validation using a separate cohort)	0	0	0	0	0	U	0	0	0	0	0	0
Test Methods:												
18. Describe number, training and expertise of persons reading reference standard	1	1	1	1	1	0	0	1	1	1	1	1
19. If >1 person reading reference standard, quote measure of consistency (e.g. kappa)	1	0	1	n/a	n/a	0	0	n/a	0	0	0	0
20. Blinding of interpreters of reference standard to results of classification by administrative data e.g. Chart abstractor blinded to how that chart was coded	U	1	1	U	0	0	0	0	1	1	0	1
Statistical Methods:												
21. Describe methods of calculating diagnostic accuracy	1	1	1	1	1	1	0	1	1	1	0	1
RESULTS: Participants:												
22. Report when study done, start/end dates of enrollment	1	1	1	1	1	1	0	1	1	1	1	1
23. Describe number of people who satisfied inclusion/ exclusion criteria	1	1	1	1	1	1	1	1	1	1	1	1
24. Study flow diagram	0	0	0	1	0	0	0	0	1	0	0	0
Test results:												
25. Report distribution of disease severity	1	1	1	1	0	1	0	0	1	1	1	1
26. Report cross-tabulation of index tests by results of reference standard	1	1	1	1	0	1	0	0	1	0	0	0
27. Report at least 4 estimates of diagnostic accuracy	0	1	0	1	1	0	0	0	0	1	1	0
Diagnostic Accuracy Measures Reported:												
28. Sensitivity	0	1	0	1	1	1	0	0	0	1	1	1
29. Specificity	0	1	0	1	1	0	0	0	0	1	1	0
30. PPV	1	0	1	1	1	1	1	1	1	1	1	0
31. NPV	0	0	0	1	1	0	0	0	0	1	1	0
32. Likelihood Ratios	0	1	0	0	0	0	0	0	0	0	0	0
33. Kappa	0	0	0	0	0	0	0	0	0	0	0	0
34. Area under the ROC curve / c-statistic	0	0	0	0	0	0	0	0	0	0	0	0
35. Accuracy/agreement	0	0	0	0	0	0	0	0	0	0	0	0
36. Other (specify)	0	0	0	0	0	0	0	0	0	0	0	0
37. Report accuracy for subgroups (e.g. age, geography, etc.)	0	1	0	0	0	0	0	1	1	0	0	1
38. If PPV/NPV reported, does the ratio of cases/controls of validation cohort approximate prevalence of condition in the population	1	1	n/a	n/a	1	n/a	n/a	n/a	n/a	0	0	n/a
39. Report 95% CI for each diagnostic measure	1	1	1	1	1	1	0	0	1	1	1	1
DISCUSSION:												
40. Discuss the applicability of the validation findings	1	1	1	1	1	1	0	1	1	1	1	1
Total Score	27	25	27	30	28	24	10	22	28	29	24	26

3.4.2 Performance Characteristics

Reference standard definitions included medical chart review, ICU registry database (both validated and not validated by ICU physicians), bacteraemia-specific registry database, surgical in-patient database, and a cohort of patients who had been entered into severe sepsis clinical trials based on specified and defined inclusion criteria. A total of 37 ICD sepsis case definitions were tested with over 130 different ICD codes (Table 7). The most commonly used codes were the ICD-9 codes 038.x (septicaemia, not otherwise specified (NOS)), 790.7 (bacteraemia, NOS), and 995.92 (severe sepsis) and the ICD-10 codes A40.x (streptococcal sepsis) and A41.x (other sepsis).

Table 7: ICD version and ICD codes used in included studies

Author	**ICD version**	**ICD codes used**
Cevasco *et al.*	ICD-9-CM	0380, 0381, 03810, 03811, 03812, 03819, 0382, 0383, 78552, 78559, 9980, 99591, 99592, 03840, 03841, 03842, 03843, 03844, 03849, 0388, 0389
Gedeborg *et al.*	ICD-9	*Sepsis-wide criteria:* 020-023, 027A, 032, 037, 040A, 041, 060, 061, 065, 071, 074C, 078G, 078H, 112X, 118, 590, 790H, 790W
	ICD-10	*Sepsis-wide criteria:* A19-A36, A44.0, A49, A54.8, A69.2, A75-A79, B00.7, B00.9, B01.8, B01.9, B02.7,-B02.9, B05.8, B05.9, B34.9, B38-B64, R50, T79.3, T81.3-T81.6, T83.6, T83.8, T84.5-T84.7, T85.7, T88.0, Y95
	ICD-9	*Sepsis-narrow criteria:* 036C-036E, 036X, 038, 084, 112F, 117D, 286G, 999D
	ICD-10	*Sepsis-narrow criteria:* A02.1, A04.0-A04.3, A39-A41, A42.7, A48, A90-A99, B37.7, B38.7, B39.3, B40.7, B41.7, B42.7, B44.7, B45.7, B46.4, B95-B99, D65, T80.2
Grijalva *et al.*	ICD-9-CM	003.1, 036.2, 785.52, 790.7, 038.x
Ibrahim *et al.*	ICD-10-AM	*Sepsis:* A40.0, A40.1, A40.2, A40.3, A40.8, A40.9, A41.0, A41.1, A41.2, A41.3, A41.4, A41.5, A41.52, A41.58, A41.8, A41.9; *Cholecystitis:* K81.0, K83.0; *Peritonitis:* K65.9, *Pneumonia:* J13, J15.9, J18.0, J18.8, J18.9, J85.2; *Perforation:* K22.3, K27.5, K63.1
Lawson *et al.*	ICD-9-CM	038, 78552, 99591, 99592, 9980, 99859, 99931
Madsen *et al.*	ICD-10 Danish version	A42.7, A41.3, A54.8, P36, P36.5, 36.4, P36.8, P36.2, P36.1, A02.1, A40.0, A40.2, A41.9, A40.8, O08.0, O85.9, A41.1, A41.2, A40.9, O75.3, A41.4, A41.5, P36.0, P36.3, P36.9, A41.0, A40.1, A40.3, A28.2, A41.8,
Ollendorf *et al.*	ICD-9-CM	038.3, 022.3, 790.7, 038.42, 038.49, 038.40, 038.41, 054.5, 036.2, 038.2, 038.43, 003.1, 038.8, 038.9, 020.2, 038.44, 038.1, 038.0

Author	ICD version	ICD codes used
Schneeweiss *et al.*	ICD-9-CM	*Bacteremia:* 038.-, 790.7
Quan *et al.*	ICD-10-CA	A40.0, A40.1, A40.2, A40.3, A40.8, A40.9, A41.0, A41.1, A41.2, A41.3, A41.4, A41.5, A41.8, A41.9, R57.8, T81.1
Iwashyna *et al.*	ICD-9-CM	Angus positive: *Severe sepsis:* 995.92; *Septic shock* 785.52; OR Codes used to identify infection: 001, 002, 003, 004, 005, 008, 009, 010, 011, 012, 013, 014, 015, 016, 017, 018, 021, 022, 023, 024, 025, 026, 027, 030, 031, 032, 033, 034, 035, 036, 037, 038, 039, 040, 041, 090, 091, 092, 093, 094, 095, 096, 097, 098, 100, 101, 102, 103, 104, 110, 111, 112, 114, 115, 116, 117, 118, 320, 322, 324, 325, 420, 421, 451, 461, 462, 463, 464, 465, 481, 482, 485, 486, 491.21, 494, 510, 513, 540, 541, 542, 52.01, 562.03, 562.11, 562.13, 566, 567, 569.5, 569.83, 572.0, 572.1, 575.0, 590, 597, 599.0, 601, 614, 615, 616, 681, 682, 683, 686, 711.0, 730, 790.7, 996.6, 998.5, 999.3; AND *Acute organ dysfunction codes:* 785.5, 458, 96.7, 343.3, 293, 348.1, 287.4, 287.5, 286.9, 286.6, 570, 573.4, 584
	ICD-9-CM	*Explicit code positive:* 995.92, 785.52
	ICD-9-CM	*Martin Positive:* 038, 020.0, 112.5, 112.81 AND *Acute organ dysfunction codes:* 785.5, 458, 96.7, 343.3, 293, 348.1, 287.4, 287.5, 286.9, 286.6, 570, 573.4, 584 OR 995.92 OR 785.52
Ramanathan *et al.*	ICD-9-CM	995.91, 995.92, 785.52
Whittaker *et al.*	ICD-9	995.92, 785.52, Angus Coding Method (see Iwashyna *et al.)*

The validity of the ICD sepsis definitions varied greatly among studies. Seven of the 12 studies calculated Sn and 5 studies calculated Sp. Sn ranged from 5.9% to 82.3% (median: 42.4%), and Sp ranged from 78.3% to 100% (median: 98.5%). The PPV was calculated in 10 of the 12 studies and ranged from 5.6% to 100%, (median: 50%); NPV was provided in 4 studies and ranged from 62.1% to 99.7% (median: 97.4%).

One study [55] examined 18 different case definitions using a “sepsis-wide” coded definition and a “sepsis-narrow” coded definition for both ICD-9 and ICD-10 codes. These coding algorithms were then compared. Among these case definitions, Sn varied from 17.2% to 52.5% (median: 37.0%), and Sp ranged from 92.6% to 99.8% (median: 98.5%) (Table 8).

Table 8: Characteristics of Studies Included and Summary of Measures (Sn, Sp, PPV, NPV) reported by validation studies.

Author, Country, Year [ref]	Sample Population	Data Years	Type of Administrative database	Study Size (n)	ICD Version	Diagnostic Coding Field Position	Reference / Gold Standard	Sn	Sp	PPV	NPV
Cevasco *et al.* USA, 2011 [54]	General surgical	2003-2007	Population–based, inpatient veterans hospital (VA)	112	ICD-9-CM	Secondary	Medical chart review	-	-	53%	-
	General Surgical	2005-2007	Population–based, inpatient community hospital	164	ICD-9-CM	Secondary	Medical chart review	-	-	41%	-
Gedeborg *et al.* Sweden, 2007 [55]	ICU-specific	1994-1999	Population–based, inpatient	4181	ICD-9 (i)	Principal, secondary	ICU database	45.7%	97.5%	45.9%	97.5%
	ICU-specific	1994-1999	Population–based, inpatient	3434	ICD-10 (i)	Principal, secondary	ICU database	52.5%	92.6%	28.0%	97.3%
	ICU-specific and DI	1994-1999	Population–based, inpatient	4181	ICD-9 (i)	Principal, secondary	ICU database	17.2%	99.4%	56.1%	96.3%
	ICU-specific and DI	1994-1999	Population–based, inpatient	3434	ICD-10 (i)	Principal, secondary	ICU database	20.1%	98.4%	40.9%	95.7%
	ICU-specific	1994-1999	Population–based, inpatient	45	ICD-9 (i) ICD-10 (i)	Principal Secondary	Sepsis clinical trial patients	42.2%	95.5%	7.4%	99.5%
	ICU-specific	1994-1999	Inpatient Intensivist coded ICU database	45	ICD-9 (i) ICD-10 (i)	Principal Secondary	Sepsis clinical trial patients	51.5%	92.6%	5.6%	99.6%
	ICU-specific	1994-1999	Population–based, inpatient	4181	ICD-9 (ii)	Principal, Secondary	ICU database	43.0%	98.0%	49.7%	97.4%
	ICU-specific	1994-1999	Population–based, inpatient	3434	ICD-10 (ii)	Principal, Secondary	ICU database	43.0%	95.6%	-	-
	ICU-specific	1994-1999	Population–based, inpatient	4181	ICD-9 (i)	Principal	ICU database	31.7%	99.2%	63.4%	97.0%
	ICU-specific	1994-1999	Population–based, inpatient	3434	ICD-10 (i)	Principal	ICU database	21.8%	97.9%	36.4%	95.8%
	ICU-specific CAP	1994-1999	Population–based, inpatient	4181	ICD-9(i)	Principal	ICU database	51.1%	99.4%	66.7%	98.9%
	ICU-specific CAP	1994-1999	Population–based, inpatient	3434	ICD-10 (i)	Principal	ICU database	31.8%	99.0%	41.5%	98.3%
	ICU-specific CAP and DI	1994-1999	Population–based, inpatient	3434	ICD-9 (i)	Principal	ICU database	19.1%	99.8%	64.3%	98.2%
	ICU-specific CAP and DI	1994-1999	Population–based, inpatient	3434	ICD-10 (i)	Principal	ICU database	17.6%	99.4%	42.8%	97.9%
	ICU-specific CAP	1994-1999	Population–based, inpatient	3434	ICD-9 (ii)	Principal	ICU database	47.9%	99.5%	70.3%	98.8%
	ICU-specific CAP	1994-1999	Population–based, inpatient	3434	ICD-10 (ii)	Principal	ICU database	27.1%	99.0%	39.7%	98.2%
	ICU-specific CAP	1994-1999	Population-based, inpatient	45	ICD-9 (ii) ICD-10 (ii)	Principal	Sepsis clinical trial patients	46.9%	97.4%	9.9%	99.7%
	ICU-specific CAP	1994-1999	Inpatient Intensivist coded ICU database	45	ICD-9 (ii) ICD-10 (ii)	Principal	Sepsis clinical trial patients	31.2%	98.5%	10.9%	99.6%
Grijalva *et al.* USA, 2008 [56]	Rheumatoid Arthritis	1995-2004	Inpatient database	45	ICD-9-CM	Principal, Secondary	Medical chart review	-	-	80%	-

Author, Country, Year [ref]	Sample Population	Data Years	Type of Administrative database	Study Size (n)	ICD Version	Diagnostic Coding Field Position	Reference / Gold Standard	Sn	Sp	PPV	NPV
Ibrahim *et al.* Australia, 2012 [57]	General ICU	2000-2006	Inpatient database	1645	ICD-10-AM	Principal	ICU-database	44.1%	98.9%	88.2%	90.6%
	General ICU	2000-2006	Inpatient database	45	ICD-10-AM	Principal	ICU-database	16.5%	99.8%	93.9%	86.8%
Iwashyna *et al.* USA, 2014 [47]	General	2009-2010	Population-based, inpatient	111	ICD-9-CM Angus	All	Medical chart review	50.3%	96.3%	70.7%	91.5%
	General	2009-2010	Population-based, inpatient	111	ICD-9-CM Explicit	All	Medical chart review	9.3%	100%	100%	86.0%
	General	2009-2010	Population-based, inpatient	111	ICD-9-CM Martin	All	Medical chart review	16.8%	99.8%	97.6%	87.0%
Lawson *et al.* USA, 2012 [58]	General surgical	2005-2008	Population-based claims data	13 410	ICD-9-CM	All	ACS-NSQIP in-patient surgical data base	46.3%	94.0%	-	-
Madsen *et al.* Denmark, 1998 [59]	General	1994	Population–based, inpatient	471	ICD-10 – Danish	Unknown	Bacteremia Database	5.9%	-	21.7%	-
Ollendorf *et al.* USA, 2002 [60]	Severe sepsis clinical trial patients	No dates given	Population-based, inpatient claims	122	ICD-9-CM	All	Severe sepsis clinical trial patients	-	-	75.4%	-
Quan *et al.* Canada, 2013 [61]	General Surgical	2007-2008	Population–based, inpatient	117	ICD-10	Secondary	Medical chart review	-	-	9.8%	-
	General Surgical	2007-2008	Population–based, inpatient	34	ICD-10	Secondary	Medical chart review	-	-	12.5%	-
Ramanathan *et al.* USA, 2013 [62]	Surgical mortality patients	2012-2013	Surgical inpatient	243	ICD-9-CM	All	Medical chart review	82.3%	78.3%	91.1%	62.1%
Schneeweiss *et al.* USA, 2007 [63]	General	2001-2004	Population–based, inpatient	158	ICD-9-CM	Principal	Medical chart review	-	-	91%	-
Whittaker *et al.* USA, 2013 [64]	ED admitted inpatients	2005-2009	Population-based, inpatient	1735	ICD-9 (Severe)	All	Medical Chart review	20.5%	-	-	-
	ED admitted inpatients	2005-2009	Population-based, inpatient	1735	ICD-9 (Severe)	All	Medical Chart review	47.2% (Angus)	-	-	-
	ED admitted inpatients	2005-2009	Population-based, inpatient	321	ICD-9 (Shock)	All	Medical Chart review	49.5%	-	-	-
	ED admitted inpatients	2005-2009	Population-based, inpatient	321	ICD-9 (Shock)	All	Medical Chart review	42.4%	-	-	-
	ED admitted inpatients	2005-2009	Population-based, inpatient	321	ICD-9 (Shock)	All	Medical Chart review	75.1% (Angus)	-	-	-

Abbreviations: CAP- Community acquired sepsis (ICU admission within 48 hours), DI - Department of Infectious Disease patients, CM-Clinical Modification, ACS-NSQIP- American College of Surgeons National Surgical Quality Improvement Program, ED – Emergency Department (i) Sepsis-wide criteria codes, (ii) Sepsis-narrow criteria codes

3.5 Discussion

In this review we identified and summarized the published literature evaluating and validating ICD-9 and ICD-10 codes used to identify sepsis in administrative databases. We identified 12 studies that met all eligibility criteria for this systematic review and found large variations in terms of the scope of ICD-codes used and the estimates of validity among studies. All studies validated inpatient data and the majority of the studies found that ICD codes defining a diagnosis of sepsis in administrative data are highly specific but lack sensitivity (in 10 out of the 12 studies less than 53%). A reasonable conclusion is that sepsis is largely under-coded in administrative data using ICD-9 or ICD-10 coded case definitions. However the high Sp and NPV does mean that few false positives would be present in such a dataset.

The heterogeneity seen among the studies in coding accuracy especially with respect to Sn and PPV may be due to multiple factors including the number of codes used, the version of ICD used, the sample population, the reference standard comparison used, and the type of administrative data. For instance Gedeborg *et al.* [55] applied the same coding algorithms to different patient populations including ICU patients with community acquired sepsis and Department of Infectious Disease patients and tested these against two different reference standard definitions (sepsis clinical trial patients and patients from an ICU-specific coded database). They showed the data accuracy to have large variations dependent on the patient population being studied and reference standard used. Not surprisingly, limiting the sample population to one in which an infectious disease service was consulted during the patient stay actually decreased the Sn by 28.5%; while only increasing the Sp by 1.9%. A previous epidemiological study that examined patients outside of the ICU reported that severe sepsis was poorly documented although sepsis was commonly found on non-ICU medical wards [65]. This

suggests that the accuracy of diagnostic codes may be substantially impacted dependent on the population selected or the criteria that is used to define the population and must be taken into account when utilizing ICD codes to define a study cohort. No study examined the expertise of the coders, or the impact of physician documentation on the selected codes.

Validity is also dependent on diagnosis coding field location (primary or secondary or all). Cevasco *et al*. [54] examined a population-based in-patient database but restricted sepsis diagnostic code to a secondary coding field position resulting in lower PPV values (43% and 51%). Grijalva *et al*. [56] restricted the population to a highly specific patient sample (rheumatoid arthritis patients), and examined only 5 ICD-9-CM codes but they allowed the coding field position to be either primary or secondary and reported a PPV of 80%. Gedeborg *et al*. [55] performed multiple comparisons using primary or both primary and secondary code field positions. They reported consistently high Sn estimates when both the primary and secondary coding positions were included. The primary coding field is normally designated to the condition that contributed the most to a patient's length of stay or the main reason for admission (depending on country). Thus sicker patients presenting with severe sepsis or septic shock were likely to be captured using the primary diagnosis alone.

The severity of sepsis and the challenges of determining severity were reflected in the coding of a sepsis diagnosis in administrative data. Peoze *et al*. [66] examined how a physician's awareness and attitude towards the diagnosis of sepsis impacted the recording of sepsis. They reported that 46% of the time in the case of sepsis, the cause of death was incorrectly recorded as due to another disease. Missed reporting of sepsis may also be due to the capacity of practicing physicians to recognize clinical cases of SIRS, sepsis, severe sepsis and septic shock. Assunção

et al. [67] found sepsis was most frequently misdiagnosed, up to 66.5% of the time, as infection without clinical and laboratory signs of inflammatory response.

The variation in diagnosing sepsis alone also translates to variable recording of the diagnosis in the medical record. O'Malley *et al*. [6] describe the patient trajectory from admission to discharge and the process of recording the admitting diagnosis to the assignment of an ICD code post-discharge. A suggested error when a physician records a diagnosis in the medical record is based on the variance across terms and language used to describe the disease, and/or a reporting of an infection without concomitant reporting of systemic inflammation or associated organ dysfunction. Consistent documentation practices among physicians may enhance the coding and enhance the capture rate of sepsis in administrative data.

The results from this systematic review should raise a question of whether reliable research on sepsis can be performed using administrative data. Based on the findings from this systematic review, hospital discharge abstract data is an insufficient source for studies to examine sepsis incidence or for surveillance. Use of administrative data and ICD coding algorithms could still be used for studies that examine risk factors to the development of sepsis, or which examine outcomes, but with a limitation that the studies may include a subset of more easily defined/recognized cases of sepsis. Improving the quality of administrative data for sepsis studies could be accomplished with simple strategies such as 1) improved physician documentation, such as documenting sepsis in the front pages of the chart to get the attention of coders, 2) having a specialized coding procedure for ICU patients perhaps including specific training of health care coders to improve familiarity with the case mix of patients and conditions that are more prevalent in the ICU, and 3) for those countries in which a limited number of diagnosis coding fields exist, there should be at least eight coding fields for diagnosis to capture

conditions such as sepsis [68]. To limit coding error, we recommend that data linkage to other data sources be considered, such as laboratory or microbiology data.

There are limitations to this systematic review. The search strategy was limited to studies only published in English and a grey literature search was not conducted. The target of the study was ICD codes for sepsis specifically, however since sepsis itself is difficult to diagnose and has a range of clinical presentations, there is a possibility that validation studies examining only these other conditions and not sepsis specifically may have been missed. Publication bias in validation studies may also be of concern as authors may only report better-performing case definitions, not publish lower performing case definitions with very low diagnostic accuracy. However our systematic review included studies that published very low values for case definitions and therefore there is little concern that publication bias has occurred.

3.6 Conclusions

Validated case definitions for sepsis have been reported with varying degrees of accuracy from studies using administrative data. Sepsis remains one of the top causes of death, specifically in the ICU, and as more researchers are utilizing administrative data to study sepsis outcomes and health services associated with care an accurate ICD coded case definition is needed. Future studies are warranted to optimize the ascertainment of sepsis in administrative data either by testing new enhanced definitions, optimizing physician documentation and/or considering data linkage.

CHAPTER FOUR: MANUSCRIPT #2

4.0 Development and validation of an ICD-10 case definition for sepsis in ICU patients in Canadian administrative health data.

Rachel J Jolley, Hude Quan, Nathalie Jetté, Keri Jo Sawka, Lucy Diep, Jade Goliath, Derek J Roberts, Bryan G Yipp, Christopher J Doig

Submitted manuscript to the journal *Critical Care Forums*

Formatted for manuscript-based thesis

4.1 Abstract

Objective: Administrative health data are important for health research. We developed a new International Classification of Disease (ICD) coded case definition for sepsis based on an existing algorithm. We validated and compared the existing definitions with the new developed algorithm in intensive care unit (ICU) patients.

Data Sources: The medical records of ICU patients from three tertiary care centers in Calgary, Canada were randomly selected and linked to the discharge abstract database.

Study Selection: All adults (aged ≥18 years) admitted to the ICU were included.

Data Extraction: We validated the Canadian Institute for Health Information (CIHI) ICD-10-CA (Canadian Revision) coding algorithm for sepsis against a reference standard medical chart review. Sensitivity (Sn), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) were calculated.

Data Synthesis: The existing CIHI ICD-10-CA coding algorithm for sepsis had Sn (46.4%), Sp (98.7%), PPV (98.2%), and NPV (54.7%) and for severe sepsis had Sn (47.2%), Sp (97.5%), PPV (95.3%), and NPV (63.2%). Our new ICD coded algorithm for sepsis increased the accuracy by 25.5% (Sn: 71.9%, NPV: 66.6%) with slightly lowered Sp (85.4%) and PPV (88.2%). For severe sepsis both Sn (65.1%) and NPV (70.1%) increased, while Sp (88.2%) and PPV (85.6%) decreased slightly.

Conclusions: This study demonstrates that sepsis is under-coded in administrative data, thus under-ascertaining the true incidence of sepsis. The new definition has a higher validity with higher sensitivity and positive predictive value and should be preferentially considered if used for surveillance purposes.

4.2 Background

Sepsis is a life-threatening condition with a high rate of occurrence in the intensive care unit (ICU) [8], [69]. It is one of the most costly diseases to treat [70], [71] leaving long-term physical and cognitive effects on its survivors [27]. Historically, sepsis has been difficult to define, diagnose and treat [72]. In 1992, the American College of Chest Physicians and Society for Critical Care Medicine (ACCP/SCCM) published the first consensus clinical definitions of sepsis outlining the terminology and clinical characteristics of the spectrum of illness [15]. In 2001, these clinical definitions were updated to provide more clarification on the signs and symptoms of the disease and to identify methodologies to increase the accuracy and reliability of the diagnosis of sepsis [19].

Administrative health data are widely collected and a generally cost-effective way of studying multiple outcomes in large populations [6]. Such administrative data typically record medical conditions using the World Health Organization's (WHO) *International Classification of Diseases* (ICD) [34] codes, an alphanumeric system that is used to code diagnosis. Multiple sepsis-related ICD codes exist. The Canadian Institute for Health Information (CIHI) created an ICD-10-CA (Canadian Revision) coding algorithm to define sepsis in administrative data [21]. This particular definition uses 49 ICD codes to define sepsis and 28 codes specific to organ dysfunction for severe sepsis. Other studies have utilized infection codes [4] or a more limited number of codes for sepsis [1].

Validating these coding definitions is necessary to understand whether or not the patient population for a particular disease condition is identified correctly. Previous studies have assessed the validity of sepsis coding in administrative data [54-64] reporting sensitivities ranging from 5.9% [61] to 82.3% [64]. The number and type of codes applied are just some of the methodological differences seen across studies complicating the comparison of results in both national and international contexts. An accurate coding algorithm is important as it may have direct implications on healthcare resource appropriation and other healthcare system decisions [6].

Sepsis is still not accurately defined in administrative data therefore, we sought to optimize the existing CIHI coding algorithm and create a new algorithm with enhanced validity to increase case capture. We validated each definition for sepsis and severe sepsis patients in an ICU setting.

4.3 Methods

4.3.1 Data sources and study population

This study utilized two databases, the inpatient discharge abstract database (DAD) and an ICU-specific clinical database, TRACER. The DAD has detailed information including demographic, administrative and procedure data on inpatient hospital visits with each inpatient visit record containing up to 50 ICD-10-CA diagnosis coding fields recorded per hospital encounter (25 fields are released for research). Clinical data were abstracted from the ICU database TRACER (details described elsewhere [73]) which contains ICU specific clinical and demographic characteristics including APACHE (acute physiology and chronic health evaluation) II [74] and SOFA [75] (sequential organ dysfunction assessment) scores. Medical charts were also reviewed. All data was linked using the Alberta personal health number (PHN), which is a unique lifetime identifier.

Our study population included adult patients (aged 18 years and older) admitted to an ICU in one of three hospitals in the Calgary region in Alberta, Canada between January 1, 2009 and December 31, 2012. A random sample of 1001 patients was selected. This study was approved by the Conjoint Health Research Ethics Board (CHREB) at the University of Calgary.

4.3.2 Defining sepsis in medical chart and data abstraction

Sepsis was defined using a checklist criteria tool (see appendix D) developed based on the ACCP/SCCM 2001 Consensus Conference updated definitions [19] and consensus of clinical experts (criteria listed in Table 9). The tool was tested through a consensus review completed by two independent physicians, one trained in intensive care medicine and one in surgery (BGY and DJR). Each physician was given the same 10 charts and using the tool, determined if sepsis was present or absent for each case. If sepsis was present, the classification of severity (sepsis, severe

sepsis and septic shock) was indicated. These results were compared and discussed to ensure full consensus. After one round of 10 charts, there was full consensus agreement (kappa statistic=1.00).

Table 9: Diagnostic criteria for a diagnosis of sepsis, severe sepsis or septic shock.
Sepsis is defined as infection plus SIRS criteria, severe sepsis as sepsis plus at least one organ dysfunction variable and septic shock was defined as severe sepsis with one of the shock /hypotension variables.

Infection (Infection defined as a pathologic process induced by a microorganism) - documented or suspected, through microbiological or other equivalent diagnostic confirmation, and some of the following:	
SIRS criteria *Atleast 2 of the following:*	Fever: Temperature >38.3°C
	Hypothermia: Temperature <36°C
	Tachycardia: Heart rate >90/min
	Tachypnea (Respiratory rate of more than 20 breaths per minute)
	Leukocytosis (WBC count >12 x 10^9/L)
	Leukopenia (WBC count 4 < x 10^9/L)
	WBC count with >10% immature granulocytes (bands + myelocytes + metamyelocytes)
Organ dysfunction variables *At least 1 of the following*	$PaO_2/Fi0_2$<300 or <200 in patients with lung injury
	Pa02/Fi02 ratio <250 in patients with Bilateral Pulmonary Infiltrates
	Any FiO_2 ; $Sa0_2$ < 90% w/ $FiO_2 \geq$ 50%
	Decreased $SSVCO_2 \leq$70%
	Need for non-elective invasive or non-invasive mechanical ventilation
	Creatinine >176.8µmol/L >50% increase in SCr from baseline
	Decreased Urine Output <0.5mL/Kg hr for >2 or 45 mL/hr for at least 2 hours, despite adequate fluid resucitation
	Total Bilirubin >34.2µmol/L
	AST (Aspartatransaminase) >80 IU/L
	ALT(Alanine transaminase) >80 IU/L
	Decreased Consciousness or GCS ≤11
	Low platelet count (< 100 x 10^9/L)
	Prolonged capillary refill time (>3 seconds)
	INR> 1.5
	Lactate (arterial) >2.2 mmol/L
Shock /hypotension variables	SBP<90 mmHg
	MAP<65 mmHg
	SBP decrease of >40 mmHg from baseline
	Vasopressers – any continuous infusion, any dose, or otherwise indicated: Epinephrine or Norepinenphrine, Vaspressin (VP) >0.02 u/min, Dobutamine, Dopamine >6mcg/kg/min

WBC, white blood cell; PaO_2, symbol for partial pressure of oxygen in arterial blood; FiO_2, fraction of inspired oxygen; $Sa0_2$, saturated arterial oxygen; $SSVCO_2$, saturated venous gas, SCr, serum creatinine; GCS, Glasgow Coma Scale; INR, international normalized ratio; SBP, systolic blood pressure; MAP, mean arterial blood pressure

Four chart reviewers underwent data abstraction training with two of the principal investigators (CJD and HQ) using the above-described checklist criteria tool. An initial consensus chart review was performed with each reviewer independently reviewing the same 20 charts. The inter-rater agreement among all four reviewers was calculated using the kappa

statistic. This was done until the strength of agreement achieved among all four reviewers was near perfect (kappa statistic between 0.81-1.00) [53]. Two rounds of review were performed including 20 charts per round; the kappa score was calculated after each round until full consensus was reached with any discrepancies discussed and resolved through a third party (CJD). Post-consensus review, data abstraction was completed independently. Cases with uncertainty were discussed to ensure consistency among all reviewers.

4.3.3 Defining sepsis in ICD administrative data

Once patients were classified as having or not having a diagnosis of sepsis, administrative data from the DAD were obtained for each patient corresponding to the specified in-patient visit occurring within the study period. Using the DAD, sepsis was defined as per CIHI's 2009 report [21] by searching through any one of the 25 diagnosis coding fields for any of the codes listed in Table 10. Severe sepsis was indicated by the combination of a code of sepsis and at least one organ dysfunction code. (Detailed code description, Table 11).

Based on the CIHI 2009 algorithm, we revised the ICD-10-CA coded case-definition for sepsis through a review of the existing literature and a thorough analysis of the CIHI coding algorithm. More specifically, we determined the codes that were frequently occurring in the false negative (FN) and false positive (FP) populations. These ICD-10-CA codes were then examined to determine which codes may indicate sepsis and could be included in the new definition based on clinical knowledge of the resulting diagnosis (Appendix D). The additional codes identified were added to the original definition, producing a new ICD-10-CA coded case definition for sepsis (Table 10). We performed an additive analysis in which each code was added individually to the original CIHI definition, as well as the inverse in which all new codes were included in the original definition, with the removal of each individually to determine the changes in accuracy

until the most optimal values of sensitivity (Sn), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) were achieved (Appendix D).

Table 10: ICD-10-CA codes used to define sepsis and severe sepsis in administrative data by algorithm

CIHI ICD-10-CA		New ICD-10-CA	
Sepsis	Severe	Sepsis	Severe
A039, A021, A207, A217, A227, A239, A241, A267, A280, A282, A327, A392, A393, A394, A40, A400, A401, A402, A403, A408, A409, A41, A410, A411, A412, A413, A415, A4150*, A4151*, A4152*, A4158*, A418, A4180*, A4188*, A419, A427, B007, B377, P360, P361, P362, P363, P364, P365, P368, P369, P352, P372, P375	*Sepsis codes with any of the following codes for organ dysfunction:* *Respiratory*: J96.0, J96.9, J80, R09.2 *Cardiovascular*: R57.0, R57.1, R57.2, R57.8, R57.9, I95.1, I95.9 *Renal*: N17.0, N17.1, N17.2, N17.8, N17.9 *Neurologic*: K72.0, K72.9, K76.3, F05.0, F05.9, G93.1, G93.1, G93.4, G93.80 *Haematologic*: D69.5, D69.6, D65 *Procedure codes:* 1GZ31CAND, 1GZ31CRND, 1GZ31GPND	CIHI ICD-10-CA sepsis codes plus following additional codes: A047 B9548 B956 B962 J189 J440 N390	R57.2- septic shock *OR* *Sepsis codes with any one of the codes listed for organ dysfunction (CIHI definition)*

Table 11: ICD-10-CA codes and descriptions

No.	Diagnostic Code	Code Description
1.	A039	Shigellosis, unspecified
2.	A021	Salmonella sepsis
3.	A047	Enterocolitis due to Clostridium difficile
4.	A207	Septicaemic plague
5.	A217	Generalized tularaemia
6.	A227	Anthrax sepsis
7.	A239	Brucellosis, unspecified
8.	A241	Acute and fulminating melioidosis
9.	A267	Erysipelothrix sepsis
10.	A280	Pasteurellosis
11.	A282	Extraintestinal yersiniosis
12.	A327	Listerial sepsis
13.	A392	Acute meningococcaemia
14.	A393	Chronic meningococcaemia
15.	A394	Meningococcaemia, unspecified
16.	A40	Streptococcal sepsis
17.	A400	Sepsis due to Streptococcus, group A
18.	A401	Sepsis due to Streptococcus, group B
19.	A402	Sepsis due to Streptococcus, group D
20.	A403	Sepsis due to Streptococcus pneumoniae
21.	A408	Other Streptococcal sepsis
22.	A409	Streptococcal sepsis, unspecified
23.	A41	Other sepsis
24.	A410	Sepsis due to Staphylococcus aureus
25.	A411	Other sepsis
26.	A412	Sepsis due to unspecified Staphylococcus

No.	Diagnostic Code	Code Description
27.	A413	Sepsis due to Haemophilus influenzae
28.	A414	Sepsis due to anaerobes
29.	A415	Sepsis due to other Gram-negative organisms
30.	A4150*	Sepsis due to Escherichia coli [E.coli]
31.	A4151*	Sepsis due to Pseudomonas
32.	A4152*	Sepsis due to Serratia
33.	A4158*	Sepsis due to other gram-negative organisms, NOS
34.	A418	Other specified sepsis
35.	A4180*	Sepsis due to enterococcus
36.	A4188*	Other specified sepsis
37.	A419	Sepsis, unspecified, *Includes:* Septicaemia
38.	A427	Actinomycotic sepsis
39.	B007	Disseminated herpesviral disease, *Includes:* Herpes viral sepsis
40.	B377	Candidal sepsis
41.	B9548	Other Streptococcus as the cause of diseases classified to other chapters
42.	B956	Staphylococcus aureus as the cause of diseases classified elsewhere
43.	B962	Escherichia coli [E. coli] as the cause of diseases classified elsewhere
44.	J189	Pneumonia, Unspecified Organism
45.	J440	Chronic Obstructive Pulmonary Disease With Acute Lower Respiratory Infection
46.	N390	Urinary tract infection, site not specified
47.	P360	Sepsis of newborn due to Streptococcus, group B
48.	P361	Sepsis of newborn due to other and unspecified streptococci
49.	P362	Sepsis of newborn due to Staphylococcus aureus
50.	P363	Sepsis of newborn due to other and unspecified staphylococci
51.	P364	Sepsis of newborn due to Escherichia coli
52.	P365	Sepsis of newborn due to anaerobes
53.	P368	Other bacterial sepsis of newborn
54.	P369	Bacterial sepsis of newborn, unspecified
55.	P352	Congenital herpes viral [herpes simplex] infection
56.	P372	Neonatal (disseminated) listeriosis
57.	P375	Neonatal candidiasis
58.	R572	Septic Shock

*ICD-10-CA (Canadian Edition) specific codes

4.3.4 Statistical analysis

We calculated Sn, Sp, PPV, NPV and their 95% confidence intervals (95% CIs) for the CIHI and new ICD algorithm. Descriptive statistics were used to describe the study populations acquired by each ICD-10-CA algorithm. The Charlson comorbidity score was calculated using previously described methods [26]. Sn was calculated as the proportion of cases classified as positive by both the administrative data (DAD) and medical record review or "true positives" (TP) compared to all cases positive by the reference standard (medical record review). Sp was calculated as the proportion of cases without sepsis identified by both the DAD and medical

record review or "true negatives" (TN) compared to all cases negative by the reference standard. PPV was calculated as the proportion of TP cases of sepsis compared to all the cases identified as sepsis by the DAD. NPV was calculated as the proportion of cases without sepsis (TN) compared to all the sepsis compared to the all the cases identified as not sepsis by the DAD. All statistical analyses were performed using STATA version 12 (Stata Corp., College Station, TX) [77].

4.4 Results

4.4.1 Patient characteristics for reference standard diagnosis

A total of 1001 patients admitted to the ICU were included and linked to the DAD and TRACER databases. Of these 604 patients were classified as having sepsis [203 (33.6%) with severe sepsis, 315 (52.2%) with septic shock] and 397 were classified as not having sepsis. Of the sepsis patients included in the study, 59.3% were male, their median age was 61 years, 76.5% were admitted through the emergency department (ED), and 44.9% had 2 or more Charlson comorbidities (Table 3). The APACHE II score within the first 24 hours of admission was 20.8 and the admission SOFA score was 6.6. Median hospital length of stay (LOS) was 19 days and median ICU LOS was 5.8 days. ICU mortality was 17.1% and hospital mortality was 24.0%.

The new ICD-based coded case definition (modified algorithm with additional codes) increased the number of cases of sepsis identified by 207 (138 for severe sepsis). The new definition had similar cohort characteristics in both the sepsis and severe sepsis populations to the CIHI ICD-coded algorithm, however the CIHI ICD-coded algorithm patients had higher median APACHE II scores for both sepsis (22.9 versus 20.9) and severe sepsis (23.6 versus 22.4) and higher admission SOFA scores for sepsis (7.5 versus 6.6) and severe sepsis (7.7 versus 7.0) (see table 2). Hospital length of stay was similar among each population, however median

ICU length of stay was higher in the patients identified with the new severe sepsis algorithm at 6.3 days versus 5.9 days in the CIHI definition. ICU mortality was 6.6% higher in sepsis patients and 4.4% higher in severe sepsis patients classified based on the CIHI algorithm. Hospital mortality was 7% higher in sepsis patients and 4.2% higher in severe sepsis patients identified by the CIHI algorithm.

Table 12: Patient clinical characteristics and demographics of the study population by coding algorithm and reference standard definition (n=1001)

Characteristic	Overall sepsis patients (Reference standard) (n=604)	Coded by administrative data definition			
		Sepsis		Severe sepsis	
		CIHI (n=285)	New (n=492)	CIHI (n=257)	New (n=395)
Sex (Male), n (%)	358 (59.3)	162 (56.8)	270 (54.9)	142 (55.3)	206 (52.2)
Age, Median (IQR)	61 (48 – 74)	61 (50 – 74)	63 (50.5 – 74)	62 (50 – 75)	64 (52 – 75)
Admitted through ER, n (%)	462 (76.5)	226 (79.3)	382 (77.6)	202 (78.6)	310 (78.5)
Re-admit, n (%)	43 (8.6)	18 (7.7)	37 (9.1)	15 (7.1)	30 (9.5)
Immunosuppressed n (%)	39 (6.5)	26 (9.1)	36 (7.3)	22 (8.6)	30 (7.6)
Charlson Comorbidity					
0	178 (29.5)	76 (26.7)	127 (25.8)	67 (26.1)	91 (23.0)
1	155 (25.7)	69 (24.2)	124 (25.2)	59 (23.0)	95 (24.0)
2 or more	271 (44.9)	140 (49.1)	241 (49.0)	131 (51.0)	209 (52.9)
Charlson condition, n (%)					
Acute myocardial infarction	56 (9.3)	26 (9.1)	45 (9.2)	26 (10.1)	41 (10.4)
Congestive heart failure	87 (14.4)	41 (14.4)	77 (15.7)	43 (15.4)	69 (17.5)
Peripheral vascular disease	38 (6.3)	22 (7.7)	32 (6.5)	20 (7.8)	30 (7.6)
Cerebrovascular disease	35 (5.8)	13 (4.6)	31 (6.3)	12 (4.7)	23 (5.8)
Dementia	12 (2.0)	5 (1.8)	12 (2.4)	5 (2.0)	12 (3.0)
COPD	125 (20.7)	49 (17.2)	111 (22.6)	46 (17.9)	89 (22.5)
Rheumatoid Disease	11 (1.8)	9 (3.2)	10 (2.0)	7 (2.7)	7 (1.8)
Peptic Ulcer	22 (3.6)	13 (4.6)	20 (4.1)	12 (4.7)	20 (5.1)
Mild Liver Disease	41 (6.8)	23 (8.1)	36 (7.3)	22 (8.6)	34 (8.6)
Diabetes	66 (10.9)	34 (11.9)	53 (10.8)	31 (12.1)	47 (11.9)
Diabetes with complications	113 (18.7)	63 (22.1)	102 (20.7)	60 (23.4)	91 (23.0)
Hemiplegia or paraplegia	19 (3.2)	1 (0.4)	16 (3.3)	1 (0.4)	10 (2.5)
Renal Disease	37 (6.1)	27 (9.5)	38 (7.7)	25 (9.7)	35 (8.9)
Moderate/Severe liver disease	30 (5.0)	22 (7.7)	28 (5.7)	20 (7.8)	26 (6.6)
Cancer	67 (11.1)	37 (13.0)	53 (10.8)	33 (12.8)	45 (11.4)
Metastatic Cancer	21 (3.5)	8 (2.8)	14 (2.9)	6 (2.3)	12 (3.0)
AIDS	2 (0.3)	2 (0.7)	2 (0.4)	2 (0.8)	2 (0.5)
Surgery, n(%)	221(36.9)	95 (33.6)	170 (35.0)	80 (31.4)	124 (31.8)
*Emergent, n(%)	183 (82.8)	82 (86.3)	142 (83.5)	69 (86.3)	105 (84.7)
† APACHE II score, mean $\pm$ SD	20.8 $\pm$ 8.3	22.9 $\pm$ 8.8	20.9 $\pm$ 8.3	23.6 $\pm$ 8.8	22.4 $\pm$ 8.3
ϕ Admission SOFA score, mean $\pm$ SD	6.6 $\pm$ 4.5	7.5 $\pm$ 4.8	6.6 $\pm$ 4.5	7.7 $\pm$ 4.8	7.0 $\pm$ 4.8
Hospital LOS (days) (median [IQR])	19 (9 – 40)	18 (9 – 41)	19.5 (10– 44)	18 (9–42)	19 (10– 44)
ICU LOS (days) (median [IQR])	5.8 (2.8 – 10.7)	5.8 (2.4– 11.2)	5.9 (2.6 – 11.0)	5.9 (2.5 – 11.2)	6.3 (3.1– 11.7)
ICU outcome- Dead, n(%)	108 (17.1)	66 (23.7)	82 (17.1)	64 (25.6)	81 (21.2)
Hospital Outcome- Dead, n(%)	145 (24.0)	90 (31.6)	121 (24.6)	85 (33.1)	114 (28.9)

Abbreviations: IQR, interquartile range; SD, standard deviation; LOS, length of stay; ER, Emergency Room; COPD, Chronic Obstructive Pulmonary Disease *Emergent surgery refers to surgery needed within 24-48 hours since admission with no prior indication of surgery needed before the present admission, †APACHE II score- recorded within the first 24 hours of admission by physician, ϕAdmission SOFA score – the maximum score recorded within the first 24 hours of admission to the ICU

4.4.2 Performance of algorithms for sepsis classification in ICU patients

The CIHI ICD-10-CA algorithm had a moderate Sn of 46.4% and NPV of 54.7%, but was highly specific (98.7%) with a PPV of 98.2%. Patients with severe sepsis had Sn of 47.2%, NPV of 63.2%, Sp of 97.5% and PPV of 95.3%. The Sn for the new coding algorithm for sepsis increased significantly by 25.5% to 71.9% and NPV increased to 66.6%, Sp and PPV decreased to 85.4% and 88.2% respectively. For severe sepsis the same trend was noted, Sn increased by approximately 18% to 65.1%, NPV with an increase to 70.1%. However Sp and PPV decreased to 88.2% and 85.6% respectively.

Table 13: Validity by administrative data definition for ICU patients (n=1001)

Coding Algorithm	TP (n)	FN (n)	FP (n)	TN (n)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
CIHI Sepsis	280	324	5	392	46.4 (42.3-50.4)	98.7 (97.0-99.6)	98.2 (96.0-99.4)	54.7 (51.0-58.4)
CIHI Severe Sepsis	245	274	12	470	47.2 (42.8-51.6)	97.5 (95.7-98.7)	95.3 (92.0 – 97.6)	63.2 (59.6 – 66.6)
New Sepsis	434	170	58	339	71.9 (68.1 – 75.4)	85.4 (81.5-88.7)	88.2 (85.0-90.9)	66.6 (62.3 - 70.7)
New Severe Sepsis	338	181	57	425	65.1 (60.9-69.2)	88.2 (85.0 – 90.9)	85.6 (81.7 – 88.9)	70.1 (66.3 – 73.8)

4.5 Discussion

This study examined the validity of an enhanced and new ICD-10-CA coded case algorithm to identify sepsis and severe sepsis in an inpatient administrative database. We identified ICD codes that optimized the performance of the coding algorithms and our data show the new ICD-10-CA algorithms with added codes achieve a higher validity than the existing CIHI ICD-based algorithm. We increased the sensitivity by over 25% by including codes for pneumonia (J189), enterocolitis due to *Clostridium difficile* (A047), chronic obstructive pulmonary disease with acute lower respiratory infection (J440), other Streptococcus as the cause of diseases classified elsewhere (B9548), *Staphylococcus aureus* as the cause of diseases

classified elsewhere (B956) and *Escherichia coli* as the cause of diseases classified elsewhere (B962). The code for septic shock (R572) was missing from the original CIHI definition, and was also included in the new definition.

When sepsis is identified and coded, it is relatively accurate as determined by the moderate to excellent Sp and high PPV in our results. This new ICD-based definition, although capturing more cases, is still only moderately sensitive suggesting that sepsis is under-coded in administrative data. Our ICD case definition has Sn of 74%, similar to that of other hospital acquired infections internationally [78] and for non-communicable diseases such as hypertension [79] and diabetes [80] in Canadian data. The low NPV achieved by our definition of 66.6% for sepsis and 70.1% for severe sepsis codes may be related to the high prevalence of sepsis in ICU patients [81]. Underreporting may have important implications if used for sepsis surveillance or resource planning and allocation of services. Other conditions have been found to be grossly under-coded, resulting in inaccurate assessments of prevalence and thereby contributing to inadequate allocation of resources for monitoring and appropriate treatment [82]. For sepsis survivors, it is important to have an accurate way of capturing these patients for future planning as they are at a high risk for long-term neurocognitive and physical conditions [83-85]. Further, these coding algorithms could be used for quality assessment surveillance monitoring studies, for example to document the rapidity of antibiotic administration.

The under-coding of sepsis could be due to a variety of other reasons including physician documentation in the medical record. Health care coders may not identify a diagnosis of sepsis based on the physician documentation alone. Physicians may not explicitly state the term "sepsis" within the medical chart, instead terms such as "SIRS" or "shock" are used, or identifying only the infection present. Rothberg *et al.* [86] suggest that patients may be

diagnosed with respiratory failure having the symptoms of pneumonia and or criteria of sepsis without identifying the specific condition or sepsis. As well, selective under-coding of a milder form of sepsis may occur as coders may intentionally disregard coding sepsis if there are other more resource-intensive and very apparent diagnoses present i.e. any highly acute but mild cases of sepsis that clinically resolves quickly but where a patient has an extended hospital stay for another reason complicating the episode of sepsis, sepsis may be missed as contributing to the hospital stay [87].

Other studies that have examined the definition of sepsis in administrative data have also identified variations in reporting. Gaieski *et al*. [7] examined four previously published methods of capturing cases of severe sepsis in administrative data using ICD-9 codes including the well-known "Angus" and "Martin" implementations and compared the incidence and mortality over a 6 year period. They identified up to a 3.5 fold variation among four sepsis case definitions in incidence with number of cases ranging from 894,013 to 3,110,630 and mortality ranging from 14.7% to 29.9% depending on the ICD-9 definition used. Iwashyna *et al*. [47] validated the ICD-9 coding definitions for the Angus and Martin implementations and found these to have low sensitivities when identifying severe sepsis using administrative data. These studies along with our results suggest the need for linkages of administrative to other types of data, such as pharmacy data (e.g. antibiotic or inotropic use), to enhance the ascertainment of sepsis for surveillance purposes.

4.5.1 Limitations

There are several limitations to this study. First, we defined our reference standard using medical record data extracted by reviewers to assess the validity of the ICD-10-CA data. These

criteria are dependent on both the quality of the medical record and the knowledge and experience of the reviewers to discern a diagnosis of sepsis. As such, the possibility of a misdiagnosis of sepsis within the chart review may have occurred. However we used a comprehensive process for training and validation to mitigate this possibility that included a third review by an intensive care physician for any cases that were unclear and under contention by the reviewers. The validation of the tool was done using ICU patients, who have a highly detailed electronic record including specific clinical data exported to a longitudinal database, this may not be applicable to other patient populations. The ICU patient population was selected from tertiary care centres in a large metropolitan area which may then influence the generalizability of case capture to data coming from smaller community hospitals.

4.6 Conclusions

This study validated and optimized ICD-10-CA coded case definitions for the identification of sepsis and severe sepsis in administrative data. We revised these coding algorithms and optimized its performance, improving the Sn, with a small decrease in specificity and PPV. Sepsis regardless of severity level, is under-coded but with the improved Sn and high PPV, these definitions can be used for better defining cohorts of sepsis patients. Further studies are needed to determine if an ICD-coded case definition for sepsis in administrative data in combination with other data can maximize both the sensitivity and specificity to improve diagnostic accuracy.

CHAPTER FIVE: SUPPLEMENTAL METHODS

5.0 Validation of an ICD-10-CA case definition of sepsis in non-ICU patients

5.1 Data sources and study population

This study utilized two databases, the inpatient discharge abstract database (DAD) as described in section 4.3.1 and clinical data were also abstracted from the ICU-specific clinical database, TRACER, as described in section 4.3.1. Medical charts were also reviewed and all data was linked using the Alberta personal health number (PHN) as the unique identifier.

A total of 202 Non-ICU patients were randomly selected from general medical and surgical inpatient hospitalization records not admitted to the ICU during their hospital stay in a single site, the Foothills Medical Centre in the Alberta Health Services Calgary region between the years January 1, 2009 to December 31, 2012. Adults aged 18 and older were eligible for inclusion.

5.2 Sepsis Case Selection

For the reference standard chart review, the validated tool (as described in section 4.3.2) was applied to the non-ICU patients. During the period of data collection, any cases with uncertainty in determining a diagnosis of sepsis were discussed to ensure consistency amongst all reviewers, with a critical care physician acting as a third reviewer in the case of any disagreements.

To determine a case of sepsis in administrative data,the same ICD-10-CA codes were applied (see section 4.3.2, Table 10).

5.3 Statistical Analysis

The diagnostic accuracy of the clinical information coded in the DAD for non-ICU patients was determined as described in section 4.3.4.

CHAPTER SIX: SUPPLEMENTAL RESULTS

6.1 Performance of the algorithms in Non-ICU population

A total of 202 patients not admitted to the ICU were randomly selected and their medical records reviewed. For the non-ICU population, the performance characteristics of each algorithm are shown below in Table 14. The CIHI ICD-10-CA algorithm for a diagnosis of sepsis had extremely low Sn of 6.7%, however it was highly specific at 100%. Severe sepsis cases were slightly higher with Sn of 25% in the DAD and still remained highly specific at 100%. The new coding algorithm improved the Sn for sepsis cases to 60% while the Sn remained the same for severe sepsis at 25%.

Table 14: Validity by Definition for non- ICU patients (n=202)

Definition/ Coding Algorithm	TP (n)	FN (n)	FP (n)	TN (n)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
CIHI Sepsis	1	14	0	187	6.7 (0.1 – 31.9)	100	100	93.0 (88.6-96.1)
CIHI Severe Sepsis	1	3	0	198	25 (0.6-80.6)	100	100	98.5 (95.7-99.7)
New Sepsis	9	6	10	177	60 (32.2 – 83.7)	94.7 (90.4-97.4)	52.6 (28.9-75.6)	96.7 (93.0-98.8)
New Severe Sepsis	1	3	1	197	25 (0.6-80.6)	99.5 (97.2 – 99.9)	50 (1.3 – 98.7)	98.5 (95.7 – 99.7)

CHAPTER SEVEN: OVERALL DISCUSSION AND CONCLUSIONS

7.1 Key Findings

7.1.1 Systematic Review

This part of the study systematically reviewed the international literature for the validity of ICD-9 and ICD-10 coded algorithms for detecting sepsis in administrative data. Major findings include:

- Sepsis is under-coded in administrative data using ICD-9 and ICD-10 based case definitions.
- There is a high degree of heterogeneity across studies for coding sepsis in administrative data that is dependent on the ICD codes and version used, the population studied, the criteria used to define sepsis and the diagnostic coding position.
- To improve the capture of true sepsis cases in administrative data, strategies should be considered including improving physician documentation, implementing specialized coding procedures for ICU patients, and the use of at least 8 coding fields for diagnosis to capture complex conditions such as sepsis.

7.1.2 Validation of an ICD-10-CA case definition for sepsis in ICU population

We validated an existing CIHI ICD-10-CA coding algorithm for identifying sepsis in administrative data using medical records from ICU patients in three tertiary care centers. We identified other ICD codes through an analysis of the false negatives and false positives determined in the study population and optimized the performance of the coding algorithms thus creating a new algorithm. The major findings were:

- The existing CIHI ICD-10-CA definition had low to moderate Sn and NPV and high Sp and PPV indicating that sepsis is highly under-coded but when it is captured in administrative data, it is captured accurately.
- The new ICD-10-CA definitions achieved a higher validity than the existing CIHI definitions (higher Sn and higher NPV in both sepsis and severe sepsis populations) however at the cost of a lower Sp and PPV.
- The new definition increased the overall case capture for both sepsis and severe sepsis.

7.1.3 Validation of an ICD-10-CA case definition for sepsis in non-ICU population

To determine the generalizability of the validation study to populations other than ICU, we determined the validity of ICD coding in a non-ICU population using the same tool validated on an ICU population. The key findings were:

- Sepsis is poorly coded in non-ICU patient records as indicated by the extremely low Sn (6.8%) for the existing definition
- The new coding algorithm increased the Sn for sepsis to 60% however for severe sepsis Sn still remained at 25%
- Sp, PPV and NPV were all high at values of up to 100%, however this is most likely due to the low prevalence of sepsis occurring in this patient population.

7.2 Discussion

In both the systematic review and the validation study, our results try to indicate that ICD-based case definitions for sepsis are highly under-coded in administrative data. The international literature found large variances in accuracy, with only seven studies having calculated Sn and none of these had Sn of more than 82%. Our new algorithm did increase the validity of detecting sepsis in administrative data versus the existing CIHI ICD-10-CA algorithm

for both sepsis and severe sepsis in ICU patients and we achieved a moderately sensitive definition.

According to our analysis of non-ICU patients, sepsis may not be well detected in patients not admitted to the ICU at any point during their hospital stay. This may be due to the low prevalence of severe sepsis in non-ICU settings, potentially causing poor recognition of the disease or missed coding in this population. One study found that 14% of sepsis cases were acquired through an ICD-9-based automated coding screen in non-ICU patients, however they reviewed 111 patient charts and found 66% of the positive screen to be positive for sepsis by chart review and physician assessment [88]. Our results did however demonstrate a high Sp, PPV, and NPV indicating that if sepsis is coded in this non-ICU population, it is highly accurate.

The reasons for possible under-coding are described in section 3.5 and 4.5. Overall this study identified how many different coding algorithms could be used to identify sepsis and the validity of those definitions is dependent on both the definition of the disease and severity level to be captured i.e. bacteremia, septicemia, severe sepsis or septic shock. For a clinically complex and hard-to-define disease that can be transient in nature, the question of whether administrative data is appropriate for research and health services planning arises. To answer this, examining the weaknesses of administrative data in a research context is necessary. The estimated accuracy of the algorithms seen in our study demonstrates the existence of a lack of quality control measures in maintaining a consistent accurate coding of the disease. Although we increased the Sn to above 70%, this is still only a moderately sensitive estimate, indicating that for sepsis there may be multiple issues along the disease coding trajectory causing the issue. Due to this, we can conclude that administrative health data alone is unreliable for program-specific widespread use in making important clinical, research and funding decisions with regards to sepsis. For instance

a program for sepsis surveillance using only administrative health data may be flawed in that not all cases would be captured and therefore would impact health care utilization or resource allocation decisions such as number of available ICU beds.

However our results did show that these algorithms can be used confidently to obtain an accurate cohort of sepsis patients (according to the high Sp and PPV), although the decision of which algorithm or codes to use must take into account the study or decision-making purpose to ensure the most accurate algorithm is applied. If the purpose involves estimating the incidence and burden of disease at the population level, or healthcare planning for bed utilization in the ICU, then prioritizing a higher sensitivity compensated with a lower specificity to detect more positive cases would be more beneficial. Whereas capturing a cohort of sepsis patients to examine outcomes, a high specificity with a reduced sensitivity may suffice in order to minimize the number of false positive cases.

7.3 Future Research Directions

Future research in developing an accurate administrative data case definition for sepsis should involve the linking of ICD coded administrative data with key laboratory, pharmacy or other clinical data to enhance the performance of the algorithms. These could include the linking of variables such as vital signs including temperature and heart rate, blood cell counts, infection confirmation or pharmacy information such as inotropic or antibiotic use.

As well, severity level should be examined separately including larger validation studies of non-ICU populations where milder forms of sepsis are most likely to occur.

7.4 Conclusions

Overall this study showed that the validity of administrative data in capturing sepsis varies and is dependent on the ICD definition used. Firstly, through the systematic review we

determined that there are many different codes used to identify sepsis and there is a high level of heterogeneity among coding algorithms. Secondly, we increased the validity and case capture of sepsis in Canadian administrative data.

REFERENCES

1. Martin GS, Mannino DM, Eaton S, and Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *NEJM* 2003; 348:1546-1554.

2. Brun-Buisson C, Doyon F, Carlet J, Dellamonica P, Gouin F, Lepoutre A, Mercier JC, Offenstadt G, and Regnier B. (1995) Incidence, risk factors, and outcome of severe sepsis and septic shock in adults: A multicenter prospective study in intensive care units. *JAMA;* 274:968–974.

3. Dombrovskiy VY, Martin AA, Sunderram J, and Paz, HL. Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: A trend analysis from 1993 to 2003. *Crit Care Med* 2007; 35: 1244-1250.

4. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinksy MR: Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001; 29:1303–131.

5. Iwashyna TJ, Cooke CR, Wunsch H, Kahn JM. Population burden of long-term survivorship after severe sepsis in older Americans. *JAGS* 2012; 60: 1070-77.

6. O'Malley KJ, Cook KF, Price MD, Raiford Wildes K, Hurdle JF, Ashton CM. Measuring Diagnosis: ICD Code Accuracy. *Health Research Methods 2005;* 40:1620-1639

7. World Health Organization. (1992). *International Statistical Classification of Disease and Health Related Problems, 10th Revision (ICD-10)*. Geneva: World Health Organization.

8. Vincent JL, Sakr Y, Sprung CL, Ranieri VM, Reinhart K, Gerlach H, Moreno R, Carlet J, Le Gall JR, and Payen D; Sepsis Occurrence in Acutely Ill Patients Investigators. Sepsis in European intensive care units: results of the SOAP study. *Crit Care Med* 2006; 34:344-353.

9. Gaieske DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med* 2013; 41(5):1167-74.

10. Majino G. The ancient riddle of sepsis. *J Infect Dis* 1991; 163 (5): 937-945.

11. Schottmueller H. Wesen und Behandlung der Sepsis. *Inn Med* 1914;31:257-280.

12. Riedemann, NC. Guo RF, Ward PA. The enigma of sepsis. *J Clin Invest* 2003; 112(4):460–467.

13. Bone RC, Fisher CJ Jr, Clemmer TP, Slotman GJ, Metz CA, Balk RA. Sepsis syndrome: a valid clinical entity. Methylprednisolone Severe Sepsis Study Group. *Crit. Care Med* 1989;17:389-393.

14. Bone RC. Why new definitions of sepsis and organ failure are needed. *Am J Med* 1993; 95(4):348-50.

15. Bone RC, Balk, RA, Cerra FB, Dellinger RP, Fein, AM, Knaus WA, Schein RM, Sibbald WJ. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; 101:1644-1655.

16. Bone RC. Toward an epidemiology and natural history of SIRS (systemic inflammatory response syndrome). *JAMA* 1992; 268:3452–3455.

17. Trzeciak S, Zanotti-Cavazzoni S, Parrillo JE, Dellinger P. Inclusion criteria for clinical trials in sepsis: did the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference definitions of sepsis have an impact? *Chest* 2005; 127: 242-5.

18. Vincent JL. Dear SIRS, I'm sorry to say that I don't like you. *Crit Care Med* 1997*;* 25: 372-374.

19. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G; International Sepsis Definitions Conference. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference*. Intensive Care Med* 2003; 29(4): 530-8.

20. Hall MJ, Williams SN, DeFrances CJ, Golosinskiy A. Inpatient care for septicemia or sepsis: A challenge for patients and hospitals. NCHS data brief, no 62. Hyattsville, MD: National Center for Health Statistics. 2011.

21. Canadian Institute for Health Information: In Focus: A National Look at Sepsis (Ottawa, Ont.: CIHI, 2009). Available online at: https://secure.cihi.ca/free_products/HSMR_Sepsis2009_e.pdf. Accessed May 2, 2013

22. Murphy SL, Xu JQ, Kochanek KD. Deaths: Final data for 2010. National vital statistics reports; vol 61 no 4. Hyattsville, MD: National Center for Health Statistics. 2013. http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_04.pdf

23. Shorr AF, Micek ST, Jackson WL, Kollef MH. Economic implications of and evidence-based sepsis protocol: can we improve outcomes and lower costs? *Crit Care Med* 2007; 35:1257-1262.

24. Nguyen HB, Corbett SW, Steele R, Banta J, Clark RT, Hayes SR, Edwards J, Cho TW, Wittlake WA. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. *Crit Care Med* 2007; 35:1105-1112.

25. Martin CM, Priestap F, Fisher H, Fowler RA, Heyland DK, Keenan SP, Longo CJ, Morrison T, Bentley D, Antman N; STAR Registry Investigators. A prospective, observational registry of patients with severe sepsis: the Canadian Sepsis Treatment and Response Registry. *Crit Care Med* 2009; 37:81-8.

26. Jackson JC, Hopkins RO, Miller RR, Gordon SM, Wheeler AP, Ely EW. Acute Respiratory Distress Syndrome, sepsis, and cognitive decline: A review and case study. *Southern Medical Journal* 2009; 102: 1150-1157.

27. Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA* 2010; 304: 1787-1794.

28. Odden AJ, Rohde JM, Bonham C, Kuhn L, Malani PN, Chen LM, Flanders SA, Iwashyna TJ. Functional outcomes of general medical patients with severe sepsis. *BMC Infect Dis* 2013 12;13:588.

29. Yende S, Linde-Zwirble W, Mayr F, Weissfeld LA, Reis S, Angus DC. Risk of cardiovascular events in survivors of severe sepsis. *Am J Respir Crit Care Med* 2014;189(9):1065-74.

30. Wang T, Derhovanessian A, De Cruz S, Belperio JA, Deng JC, Hoo GS. Subsequent infections in survivors of sepsis: epidemiology and outcomes. *J Intensive Care Med.* 2014; 29(2):87-95.

31. Letarte J, Longo CJ, Pelletier J, Nabonne B, Fisher HN. Patient characteristics and costs of severe sepsis and septic shock in Quebec. *Journal of Critical Care* 2002; 17(1):39-49.

32. Lee H, Doig CJ, Ghali WA, Donaldson C, Johnson D, Manns B. Detailed costs analysis of care for survivors of severe sepsis. *Crit Care Med* 2004; 32: 981-985.

33. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003; 12(Suppl 2): ii58–ii63.

34. World Health Organization. International Classification of Diseases (ICD). Available online at: http://www.who.int/classifications/icd/en/. Accessed February 3, 2014.

35. History of the Development of the ICD. Available online at: http://www.who.int/classifications/icd/en/HistoryOfICD.pdf. Accessed May 3, 2014.

36. Jetté N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, Paoin W, Sundararajan V, Gao S, Jakob R, Ustün B, Ghali WA; IMECCHI Investigators. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Med Care* 2010; 48(12):1105-10.

37. Canadian Institute for Health Information. Final Report: Canadian Enhancement of ICD-10. June 2001.

38. Virnig BA, McBean M. Administrative data for public health surveillance and planning. Annu Rev Public Health 2001; 22:213-30.

39. Wang HE, Shapiro NI, Angus DC, Yealy DM. National estimates of severe sepsis in United States emergency departments. *Crit Care Med* 2007; 35(8):1928-36.

40. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev* 1988;10:164-90.

41. Rothman K, Greenland S, Lash TL. Modern Epidemiology 3rd Edition.

42. Lynn LA. The diagnosis of sepsis revisited – a challenge for young medical scientists in the 21st century. *Patient Safety in Surgery* 2014; 8:1.

43. Quach S, Hennessy DA, Faris P, Fong A, Quan H, Doig C. A comparison between the APACHE II and Charlson Index Score for predicting hospital mortality in critically ill patients. *BMC Health Serv Res* 2009; 9:129.

44. Jette N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, Paoin W, Sundararajan V, Gao S, Jakob R, Ustun B, Ghali WA; IMECCHI Investigators. The development, evolution, and modifications of ICD-10: challenges to the international comparability of mortality data. *Med Care* 2010; 48(12):1105-1110.

45. Odeh M. Sepsis, septicemia, sepsis syndrome, and septic shock: the correct definition and use. *Postgrad Med J.* 1996, 72(844): 66.

46. Sands KE, Bates DW, Lanken PN, Graman PS, Hibberd PL, Kahn KL, Parsonnet J Panzer R, Orav EJ, Snydman DR, Black E, Schwartz JS, Moore R, Johnson BL Jr, Platt R; Academic Medical Center Consortium Sepsis Working Group. Epidemiology of sepsis syndrome in 8 academic medical centers. *JAMA* 1997; 278:234–240.

47. Iwashyna, TJ, Odden, A, Rohde, J, Bonham, C, Kuhn, L, Malani, P, Chen, L, Flanders, S. Identifying Patients With Severe Sepsis Using Administrative Claims: Patient-Level Validation of the Angus Implementation of the International Consensus Conference Definition of Severe Sepsis. *Medical Care* 2012; 1-5.

48. Quach S, Blais C, and Quan H. Administrative data have high variation in validity for recording heart failure. *Can J Cardiol* 2010; 26(8): 306-312.

49. Germaine-Smith CS, Metcalfe A, Pringsheim T, Roberts JI, Beck CA, Hemmelgarn BR, McChesney J, Quan H, Jetté N. Recommendations for optimal ICD codes to study neurologic conditions a systematic review. *Neurology* 2012; 79(10):1049-1055.

50. Synthesis.info. Synthesis.info. http://synthesis.info/. Updated 2013. Accessed March 20, 2013.

51. Yergens DW, Dutton DJ, Patten SB. An overview of the statistical methods reported by studies using the canadian community health survey. *BMC Med Res Methodol*. 2014; 14(1):15.

52. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, and Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011; 64:821-829.

53. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.

54. Cevasco M, Borzecki AM, Chen, Q, Zrelak PA, Shin M, Romano PS, Itani KMF, Rosen AK. Positive predictive value of the AHRQ patient safety indicator "postoperative sepsis": Implications for practice and policy. *J Am Coll Surg* 2011; 212(6):954-961.

55. Gedeborg R, Furebring M, Michaelsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected the incidence and mortality estimates of ICU patients. *Jour Clin Epid* 2007; 60(2):155-162.

56. Grijalva CG, Chung CP, Stein CM, Gideon PS, Dyer SM, Mitchel EF, and Griffin MR. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. *Pharmacoepidemiol Drug Saf* 2008; 17:890-895.

57. Ibrahim I, Jacobs IG, Webb SAR, Finn J. Accuracy of international classification of diseases, 10th revision codes for identifying severe sepsis in patients admitted from the emergency department. *Crit Care Resusc* 2012; 14(2):112-118.

58. Lawson EH, Louie R, Zingmond DS, Brook RH, Hall BL, Han L, Rapp M, Ko CY. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Annals of Surgery* 2012; 256(6):973-981.

59. Madsen KM, Schonheyder HC, Kristensen B, Nielsen GL, Sorensen HT. Can hospital discharge diagnosis be used for surveillance of bateremia? A data quality study of a Danish hospital discharge registry. *Infect. Control Hosp. Epidemiol.* 1998; 19(3):175-180.

60. Ollendorf DA, Fendrick AM, Massey K, Williams GR, Oster G. Is sepsis accurately coded on hospital bills? *Value in Health* 2002; 5(2):79-81.

61. Quan H, Eastwood C, Cunningham CT, Liu M, Flemons W, De Coster C, Ghali WA; IMECCHI investigators. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open* 2013; 3(10):1-7.

62. Ramanathan R, Leavell P, Stockslager G, Mays C, Harvey D, Duane TM. Validity of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) Screening for Sepsis in Surgical Mortalities. *Surgical Infections* 2014; 14:1-4.

63. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D., Soloman DH. Veteran's affairs hospital discharge database coded serious bacterial infections accurately. *J Clin Epid* 2007; 60(4):397-409.

64. Whittaker SA, Mikkelsen ME, Gaieski DF, Koshy S, Kean C, Fuchs BD. Severe sepsis cohorts derived from claims-based strategies appear to be biased toward a more severely ill patient population. *Crit Care Med* 2013; 41(4): 945-53.

65. Rohde JM, Odden AJ, Bonham C, Kuhn L, Malani PN, Chen LM, Flanders SA, Iwashyna TJ. The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit. *J Hosp Med.* 2013; 8(5):243-7.

66. Poeze M, Ramsay G, Gerlach H, Rubulotta F, Levy M. An international sepsis survey: a study of doctors' knowledge and perception about sepsis. *Crit Care* 2004; 8:R409-R413.

67. Assunção M, Akamine N, Cardoso GS, Mello PV, Teles JM, Nunes AL, Maia MO, Rea-Neto A, Machado FR; SEPSES Study Group. Survey on physician's knowledge of sepsis: do they recognize it promptly? *J Crit Care* 2010; 25(4): 545-52.

68. Drösler SE, Romano PS, Sundararajan V, Burnand B, Colin C, Pincus H, Ghali W; World Health Organization Quality and Safety Topic Advisory Group. How many diagnosis fields are needed to capture safety events in administrative data? Findings and recommendations from the WHO ICD-11 Topic Advisory Group on Quality and Safety. *Int J Qual Health Care* 2014; 26(1):16-25.

69. Alberti C, Brun-Buisson C, Burchardi H, Martin C, Goodman S, Artigas A, Sicignano A, Palazzo M, Moreno R, Boulmé R, Lepage E, Le Gall R. Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. *Intensive Care Med* 2002; 28:108-121.

70. Weycker D, Akhras KS, Edelsberg J, et al: Long-term mortality and medical care charges in patients with severe sepsis. *Crit Care Med* 2003; 31:2316-23

71. Torio CM, Andrews RM. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011. HCUP Statistical Brief #160. August 2013. Agency for Healthcare Research and Quality, Rockville, MD. Available online at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf. Accessed October 1, 2014.

72. Bone RC. Let's agree on terminology: Definitions of sepsis. *Crit Care Med* 1991; 19:973-976.

73. Doig CJ, Zygun DA, Fick GH, Laupland KB, Boiteau PJ, Shahpori R, Rosenal T, Sandham JD. Study of clinical course of organ dysfunction in intensive care. *Crit Care Med* 2004; 32:384-390.

74. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13:818-829.

75. Janssens U, Graf C, Graf J, Radke PW, Königs B, Koch KC, Lepper W, vom Dahl J, Hanrath P. Evaluation of the SOFA score: a single-center experience of a medical intensive care unit in 303 consecutive patients with predominantly cardiovascular disorders. Sequential Organ Failure Assessment. *Intensive Care Med* 2000; 26:1037-1045.

76. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005; 43:1130-1139.

77. StataCorp. 2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

78. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of administrative code data for the surveillance of healthcare-associated infections: A systematic review and meta-analysis. *Clin Infect Dis* 2014; 58:688-696.

79. 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; 54:1423-1428.

80. Southern DA, Roberts B, Edwards A, Dean S, Norton P, Svenson LW, Larsen E, Sargious P, Lau DC, Ghali WA. Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. *Can J Public Health* 2010; 101:61-64.

81. Padkin A, Goldfrad C, Brady AR, Young D, Black N, Rowan K. Epidemiology of severe sepsis occurring in the first 24 hrs in intensive care units in England, Wales, and Northern Ireland. *Crit Care Med* 2003; 31:2332-2338.

82. Fillit H, Geldmacher DS, Welter RT, Maslow K, Fraser M. Optimizing Coding and Reimbursement to Improve Management of Alzheimer's Disease and Related Dementias. *J Am Geriatr Soc* 2002; 50:1871–1878.

83. Griffiths JA, Gager M, Waldmann C. Follow-up after intensive care. *Contin Educ Anaesth Crit Care Pain* 2004; 4:202-205.

84. Pandharipande PP, Girard TD, Jackson JC, Morandi A, Thompson JL, Pun BT, Brummel NE, Hughes CG, Vasilevskis EE, Shintani AK, Moons KG, Geevarghese SK, Canonico A, Hopkins RO, Bernard GR, Dittus RS, Ely EW; BRAIN-ICU Study Investigators. Long-Term Cognitive Impairment after Critical Illness. *N Engl J Med* 2013; 369:1306-1316.

85. Zielske J, Bohne S, Brunkhorst FM, Axer H, Guntinas-Lichius O. Acute and long-term dysphagia in critically ill patients with severe sepsis: results of a prospective controlled observational study. *Eur Arch Otorhinolaryngol* 2014; 271:3085-3093.

86. Rothberg MB, Pekow PS, Priya A, Lindenauer PK. Variation in Diagnostic Coding of Patients With Pneumonia and Its Association With Hospital Risk-Standardized Mortality Rates: A Cross-sectional Analysis. *Ann Intern Med* 2014; 160:380-388.

87. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care* 1994; 32:81-90.

88. Rohde JM, Odden AJ, Bonham C, Kuhn L, Malani PN, Chen LM, Flanders SA, Iwashyna TJ. The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit. *J Hosp Med* 2013; 8: 243–247.

Appendices A: Inclusion Criteria for Eligible Articles title and abstract review
(Form was electronically produced using Knowledge Share Version 2.0 software)

Study Reference (Last name, Year, title)

Inclusion Criteria:

1. Was the study published between 1992 and 2012?	Yes/ No
2. Does this study address sepsis or severe sepsis?	Yes/ No
3. Does this study examine an adult population?	Yes/ No
4. Does this study examine a human population?	Yes/ No
5. Does the study present a measurement statistic of sensitivity, specificity, PPV, or NPV?	Yes/ No
6. Is the report/publication in English?	Yes/ No
7. Does this study meet all of the inclusion criteria for the study?	Yes / No
8. Should this study be included in this study?	Yes/ No

Why or Why not?

Appendices B: Data abstraction form for systematic review

Table 1: Characteristics of Studies Included and Summary of Measures (Sn, Sp, PPV, NPV)

Year and Place of Publication	Author	Data Years	Sample Population	Administrative data Source:	Study Size	ICD Version	ICD codes	Reference Standard	Sn	Sp	PPV	NPV

Appendices C: List of excluded articles and reason for exclusion

No.	Author	Year	Reason for Exclusion
1.	Abrusci T	2011	Does not validate sepsis ICD codes
2.	Ahishakiye D	2009	No ICD code validation
3.	Aitken LM	2011	No ICD code validation
4.	Al-Hadeedi S	1991	No ICD code validation, published before 1992
5.	Al-Juaid A	2012	No ICD code validation
6.	Alavi A	2011	No ICD code validation
7.	Anaya DA	2003	No ICD code validation
8.	Angyo IA	2001	No ICD code validation
9.	Arnold FW	2003	No ICD code validation
10.	Aube H	1992	No ICD code validation
11.	Bacli C	2003	No ICD code validation
12.	Bahrami S	2010	No ICD code validation
13.	Balaban U	2012	No ICD code validation
14.	Banatvala N	1997	Does not validate sepsis ICD codes
15.	Bang AT	2005	No ICD code validation, only a diagnostic system
16.	Barber C	2013	Not original research-review
17.	Bavdekar SB	2005	No ICD code validation
18.	Beekmann SE	2005	No ICD code validation
19.	Begier EM	2005	No ICD code validation, Does not validate sepsis ICD codes
20.	Bell LM	1992	No ICD code validation
21.	Bellini C	2007	No ICD code validation
22.	Bennett NJ	2007	No ICD code validation- only MRSA infection
23.	Bhatt SP	2007	Not original research
24.	Bhattarai J	2011	No ICD code validation
25.	Bouam S	2003	No ICD code validation
26.	Bouletreau A	1999	No ICD code validation
27.	Bouza E	2011	No ICD code validation
28.	Brown C	2009	No ICD code validation
29.	Brown SM	2009	No ICD code validation
30.	Buchanan-Chell M	1992	No ICD code validation
31.	Capp R	2011	No ICD code validation
32.	Carnahan RM	2012	Not original research
33.	Cevasco M	2011	Does not validate sepsis ICD codes – CVC-related BSIs
34.	Chiu D	2011	No reference standard comparison and no administrative data
35.	Cho YK	2011	No ICD code validation
36.	Choe Y	2011	No reference standard comparison
37.	Chowdhury HR	2010	No administrative data used
38.	Cremer OL	2014	Not original research-editorial

No.	Author	Year	Reason for Exclusion
39.	Damjanovic V	1995	Not original research
40.	Davis BH	2005	No ICD code validation
41.	De Prost N	2010	No ICD code validation
42.	De Wals P	1984	Published before 1992, used ICD codes, version unknown
43.	Elramah M	2012	No reference standard comparison
44.	Emori TG	1998	No ICD code validation, Does not validate sepsis ICD codes
45.	Esper AM	2006	Does not validate sepsis codes
46.	Ferreira J	2014	No ICD code validation
47.	Fischer JE	2005	Not original research, No ICD code validation
48.	Fontela PS	2013	Does not validate sepsis ICD codes
49.	Fry DE	2007	No reference standard comparison
50.	Gerald J	2011	Abstract - used ICD-9 as reference standard
51.	Golden WE	1995	No ICD codes validation
52.	Graham PL	2004	No ICD code validation
53.	Guasticchi G	2009	No ICD code validation
54.	Guevara RE	1999	Does not validate sepsis ICD codes -pneumococcal pneumonia
55.	Hayward J	1986	Published before 1992
56.	Horng S	2012	Abstract available with little information
57.	Hsu LY	2008	Published before 1992
58.	Jaimes F	2003	No ICD code validation, Does not validate sepsis ICD codes – only SIRS
59.	Juskewitch JE	2012	No ICD code validation, No administrative data
60.	Lagu T	2011	No ICD code validation -Looking at sepsis severity not ability of codes to capture sepsis
61.	Lagu T	2012	No ICD code validation
62.	Leal J	2008	Not original research
63.	Lesher L	2009	Does not validate sepsis ICD codes- only TSS definition
64.	Leth RA	2006	No ICD code validation
65.	Linde-Zwirble WT	2011	No reference standard comparison
66.	Liu FX	2010	No measures of accuracy for ICD codes
67.	McIntosh EDG	2003	No measures of accuracy for ICD codes
68.	Misset B	2009	No reference standard comparison
69.	Modi N	2013	No ICD code validation
70.	Moehring RW	2009	Does not validate sepsis ICD codes, looking at HAIs
71.	Moore LJ	2013	No ICD codes
72.	Patrick SW	2010	Does not validate sepsis ICD codes-Only CLABSI
73.	Ramanathan R	2013	Does not validate sepsis ICD codes -Abstract
74.	Romano PS	2003	Does not validate sepsis ICD codes
75.	Ruhnke GW	2009	No ICD code validation
76.	Stevenson KB	2008	Does not validate sepsis codes - looking at HAP, HAI

No.	Author	Year	Reason for Exclusion
77.	Tabak YP	2007	No ICD code validation
78.	Thompson DS	2003	No administrative data, tool evaluation
79.	Verelst S	2010	No ICD code validation
80.	Wallgren U	2011	ICD codes used as reference standard, No validation
81.	Watson RS	2012	Not original research
82.	Weiss SL	2012	No reference standard comparison
83.	Yokoe DS	1998	No ICD code validation

Appendices D: Data abstraction diagnostic tool for chart review

Section 1. DEMOGRAPHICS
Chart Number: _____
PHN - Personal Healthcare Number: _____
DOB: _____
Gender: _____

Section 2. ADMISSION INFORMATION:

Hospital Admission Date: _____

Was this patient admitted to the ICU / CICU? Yes No
(Circle which unit)
Was this patient admitted post-code? Yes No
Code 66 Code Blue

ICU Admission Date: _____
ICU Discharge Date: _____
Unit/ location admitted from:_____
Most responsible reason for hospital admission: 1)_____
2)_____
Was this patient re-admitted? Yes No
Code 66 Code Blue

Date admitted:_____ Date Discharged: _____

Are there any pre-existing conditions? (excluding conditions that occurred or were diagnosed during hospitalization)
□ MI- Myocardial Infarction - old and recent
□ CHF- congestive heart failure
□ HTN- hypertension
□ CAD- coronary artery disease
□ Peripheral vascular disease (everything except heart)
□ Cerebrovascular disease- strokes, hemmorhage (ICH-Intracranial Hemmorhage)
□ Hemiplegia or paraplegia - spinal cord injury, paralysis
□ Dementia
□ COPD- Chronic pulmonary disease, respiratory failure, asthma
□ Rheumatologic disease
□ Peptic ulcer disease
□ Diabetes
□ Renal Disease (CRF-chronic renal failure, CRD-chronic renal disease)
□ Any Malignancy/ Cancer, including leukemia and lymphoma – any time in life
□ Mild liver disease
□ Moderate or severe liver disease
□ Esophageal/gastric varices
□ AIDS/HIV
□ Hepatitis C (HepC)
Did this patient have surgery? Yes No

Type of Surgery: Emergent Elective

Was this patient on immunosuppressive medications prior to hospitalization?

□ Glucocorticoids/corticosteroids: Prednisone (Pred), Dexamethasone (Dxms, decadron), Solumedrol, Hydrocortisone
□ Chemotherapy (any)
□ Mycophenylate (cellcept)
□ Azathiopine
□ Cyclosporin A (CyA)
□ Cyclophosphamide (Cyx or Cytoxan)
□ Rituximab (Rtx)
□ Methotrexate (Mtx)

Section 3. SEPSIS WORK UP

1) Infection symptoms/signs (patient complaints/physical findings) (Check all that apply)

□ chills

□ fever

□ rigors (shakes)

□ rash

□ dysuria (painfull urination)

□ dyspnea (shortness of breath/SOB)
□ confusion (delerium, encephalopathy)

□ stiff neck (meningioma)

□ new murmurs

□ bronchial breath sounds

□ pleuritic chest pain

□ peritoneal findings (acute abdomen, rigid abd, rebound)

□ abdominal pain

□ pain out of proportion

□ purulent wound (pus)

□ cellulitis

□ skin changes of necrotizing fasciitis (flesh eating disease, Fournier's gangrene

2) SIRS Criteria (Check all that apply)

□ Temperature >38.3°C

□ Temperature <36°C

□ Tachycardia (Heart Rate >90/min)

□ Tachypnea (Respiratory rate of more than 20 breaths per minute)

□ Leukocytosis (WBC count >12 x 10^9/L)

□ Leukopenia (WBC count 4 < x 10^9/L)

□ WBC count with >10% immature granulocytes (bands + myelocytes + metamyelocytes

3) Microbiologic or Equivalent Confirmation of Infection (Check all that apply)

□ Blood Culture

□ Sputum Respiratory sample culture

□ Urine Culture

□ CSF culture

□ Wound Culture

□ Culture Other _____

□ Laparotomy or surgical findings of infection

□ Surgical Debridement (removing tissue)

□ Chest xray consistent with pneumonia

□ Chest x-ray consistent with ARDS/ALI

□ Any X-Ray/CT consistent with ischemia

□ Any X-Ray/CT consistent with infection

□ Any X-Ray/CT consistent with abscess

□ Abdominal X-Ray/CT consistent with free air

4) Based on the available information does the patient have a suspected or documented/confirmed infection (determined by signs/symptoms and microbiology/DI) confirmation?

□ Yes, type of organism_____

□ No

If Q.4 is "Yes", check the most likely source of new infection: □ Pneumonia, Empyema (lung)

□ Urinary tract infection (UTI, pyelonephritis, perinephric abscess)

□ Acute Abdominal Infection (colitis, perforated bowel, abscess, ischemic bowel)

□ Meningitis (viral encephalitis)

□ Endocarditis

□ Bone/joint infection

□ Endovascular infection (central line, catheter, device-related)

□ Skin/Soft Tissue infection (cellulitis, fasciitis, gangrene)

□ Primary Bloodstream infection

□ Other_____

5) Are any of the following organ dysfunction criteria present at a site remote from the site of infection (except in the case of bilateral infiltrates) that are not considered to be chronic conditions? (check all that apply)

□ PaO2/Fi02 <300 or <200 in patients with lung injury

□ Pa02/Fi02 ratio <250 in patients with Bilateral Pulmonary Infiltrates

□ Any (Fraction of inspired oxygen) FiO_2 ; (saturated arterial oxygen) $Sa0_2$ $< 90\%$ w/ $FiO_2 \geq 50\%$

□ Decreased $SSVCO_2$ (saturated venous gas) $\leq 70\%$

□ Need for non-elective invasive or non-invasive mechanical ventilation

□ Creatinine >176.8µmol/L >50% increase in SCr from baseline

□ Deacreased Urine Output <0.5mL/Kg hr for >2 or 45 mL/hr for at least 2 hours, despite adequate fluid resucitation

□ Total Bilirubin >34.2µmol/L

□ AST (Aspartatransaminase) >80 U/L

□ ALT(Alanine transaminase) >80 U/L

□ Decreased Consciousness or GCS ≤ 11

□ Low platelet count (< 100 10^9/L)

□ Prolonged capillary refill time (>3 seconds)

□ INR> 1.5

□ Lactate >2.2 mmol/L

6) Shock/hypotension criteria (check all that apply)

□ Systolic Blood pressure (SBP) <90 mmHg

□ Mean Arterial Pressure (MAP) <65 mmHg

□ SBP decrease of >40 mmHg from baseline

Vasopressers – any continuous infusion, any dose:

□ Epinephrine or Norepinenphrine

□ Vaspressin (VP) >0.02 u/min

□ Dobutamine

□ Dopamine >6mcg/kg/min

Is there an explicit diagnosis of "sepsis/severe sepsis/septic shock" stated in the physician progress notes or discharge summaries?

Yes No

Physician stated diagnosis (check all that apply):

□ SIRS

□ Sepsis

□ Severe Sepsis

□ Septic Shock

Section 4. DIAGNOSIS

**Questions 8, 9 and 10 are to determine where during the patient's care they developed sepsis ie: upon arrival to ICU, during ICU stay, or after ICU stay.

After due consideration of the clinical details of the patient's medical chart:	**Yes**	**No**
7. Is there an indication of sepsis, severe sepsis, or septic shock? Please circle the most appropriate: sepsis severe sepsis septic shock		
8. If answer is "Yes" to Q.7, was the diagnosis of sepsis/severe sepsis/septic shock present upon admission (within first 24 hours) to the ICU?		
9. If "No" to Q.8, did the patient subsequently develop sepsis/severe sepsis/septic shock at any time during their ICU stay?		
10. If "No" to Q.8, did the patient subsequently develop sepsis/severe sepsis/septic shock at any time during their hospital stay (after leaving ICU)? **If not an ICU patient did patient develop sepsis during their hospital stay?		

Qualifiers for diagnoses of sepsis, severe sepsis, septic shock:

Sepsis = 2 SIRS criteria + suspected or documented/confirmed infection

Severe Sepsis= sepsis criteria + ≥ 1 organ dysfunction criteria

Septic Shock = severe sepsis criteria + ≥ 1 shock/hypotension criteria

Appendices E: Frequently identified ICD-10-CA codes in false negative and false positive populations

Table 1: List of frequently occurring codes in false negative population

ICD-10-CA Code	Description
A047	**Enterocolitis due to Clostridium difficile**
B9548	**Other streptococcus as the cause of diseases classified to other chapters**
B956	**Staphylococcus aureus as the cause of diseases classified to other chapters**
B962	**Escherichia coli [E. coli] as the cause of diseases classified to other chapters**
F059	Delirium, unspecified
I100	Benign hypertension
I500	Congestive Heart Failure
I480	Atrial Fibrillation
J189	**Pneumonia, unspecified**
J440	**Chronic obstructive pulmonary disease with acute lower respiratory infection**
J690	Pneumonitis due to food and vomit
J960	Acute respiratory failure
N179	Acute renal failure, unspecified
N390	**Urinary tract infection, site not specified**
R572	**Septic Shock**

* Bolded codes were used in new definition