Validation of case-finding algorithms derived from health administrative data for identifying neonatal bacterial sepsis

Chunhe Yao

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
MSc degree in Epidemiology specialization in Biostatistics

School of Epidemiology and Public Health
Faculty of Medicine
University of Ottawa

Table of Contents

ABSTRACT

Objectives: The objectives of this thesis were to: 1) develop and validate a coding algorithm to identify true cases of neonatal bacterial sepsis, and 2) apply the algorithm to calculate incidence rates and estimate temporal trends of neonatal bacterial sepsis.

Methods: For Objective 1, the reference cohorts were assembled among neonates born in 2012-2017 using patient-level health care encounter data. Any neonates who met both the Diagnostic Criterion I (microbiological confirmation) and Criterion II (sepsis-related antibiotic administration) were included in the true-positive cohort. Potential coding algorithms were developed based on different combinations of ICD-10-CA codes on the hospitalization discharge abstract. For Objective 2, the coding algorithm with the most optimal characteristics was applied to provincial data to calculate incidence rates in Ontario during 2003-2017. Recent temporal trends were estimated by Poisson regression analysis.

Results: In Objective 1, since all true-positive cases identified were born at preterm gestation, the study population in Objective 2 was limited to preterm infants. The final coding algorithm selected had sensitivity of 75.3% (95% CI, 66.8%-83.7%), specificity of 98.2% (95% CI, 97.8%-98.6%) and PPV of 50.0% (95% CI, 42.1%-58.0%). Using this algorithm, the annual incidence declined over time from 50.2 (95% CI, 45.4-55.4) per 1000 preterm infants in 2003 to 27.5 (95% CI, 20.4-36.9) per 1000 preterm infants in 2017. The trend over time was statistically significant with P-value <0.0001. Significant variation in bacterial sepsis incidence rates was noted across infant sex and gestational age.

Conclusion: The coding algorithm developed in this study could not accurately identify neonates with bacterial sepsis from within health administrative database using the data available to us now.

For the purpose of demonstrating the application of the algorithm, we carried out Objective 2; however, it is important to cautiously interpret the provincial rates given the the poor performance of the case-finding algorithm. Future research is required to improve the case-finding algorithm for use in health administrative databases for neonatal bacterial sepsis research and surveillance.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my Thesis Committee, friends and family, who have given me the invaluable support needed to complete this thesis. To me personally, this work and my study here at University of Ottawa has been a truly transformative experience.

Foremost, I would to like to thank my supervisors Dr. Steven Hawken and Dr. Deshayne Fell for your continuous mentorship and support. Not only have I learned validation study and neonatal epidemiology under you, but I will always be in debt to you for teaching me how to do rewarding research. Thank you for patiently guiding me through this process and giving me great advices and encouragement whenever I met difficulties. I could not have imagined having a better advisor and mentor for my study. My future endeavor will forever benefit from your advices.

I am grateful to Dr. Dayre McNally for providing insightful comments and suggestions on my thesis advisory committee. Despite being incredibly busy, you always took time to share your expertise and your timely response was always there to help me push forward whenever I encountered a question.

I also want to express my gratitude to Dr. Meltem Gulsum Tuna, Robert Talarico, Robin Ducharme, Deanna Rothwell and Jo Ann Colas, who have given me tremendous support on all technical fronts. Thank you to my friends for championing me every step of the way and providing advice and support when I needed it. Last but not least, I offer my deepest gratitude to my family for being the foundation of my life. Special thanks to my beloved husband, Yapeng Wu, having you by my side makes me feel like anything in life is possible. You have given me the motivation I needed to see this research through to completion.

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

WHO World Health Organization

GBS Group B streptococci

E coli Escherichia coli

US United States

ACCP The American College of Chest Physicians

SCCM Society of Critical Care Medicine

SIRS Systemic inflammatory response syndrome

OHIP Ontario Health Insurance Plan

IKN ICES Key Number

ICD International Classification of Diseases

CIHI Canadian Institute for Health Information

TOH The Ottawa Hospital

OHDW Ottawa Hospital Data Warehouse

REB Research Ethics Board

OHSN Ottawa Health Sciences Network

AHFS American Hospital Formulary Service

IV Intravenous

DIN Identification number

TP True Positive

FP False Positive

FN False Negative

TN True Negative

PPV Positive predictive value

NPV Negative predictive value

CI Confidence intervals

DAD Discharge Abstract Database

NACRS National Ambulatory Care Reporting System

SD Standard deviation

Value/DF Ratio of the deviance to the degrees of freedom

CHAPTER 1: INTRODUCTION

1.1 Overview

As a major contributor globally to morbidity and mortality, sepsis has been listed as a key health-care priority for the coming decade by the World Health Organization (WHO). (1) Peak sepsis incidence occurs in neonates and young children. A recent meta-analysis of population-based study estimated a pooled incidence of neonatal sepsis of 22.0 (95% CI 11.0-43.6) per 1000 live births based on data from eight studies originating from six countries (two high-income countries and four middle-income countries) between 1979 and 2016. (2) Neonatal sepsis (i.e., sepsis occurring within the first 28 days of life) can be due to infection with bacterial, viral or fungal pathogens, with bacteria being the predominant cause. (3)

Routinely-collected health administrative data are potentially an appealing source of information for population-based surveillance and research of neonatal sepsis, as the data are readily available, population based, reasonably inexpensive to use and include diagnostic codes assigned in different health care settings (i.e., hospitalizations, emergency departments and physician's offices). (4-7) However, the secondary use of diagnosis codes from health administrative data requires careful validation of case-finding algorithms against gold-standard method. (8) Studies have been performed utilizing administrative data to assess the validity of case definitions of sepsis. (9-12) Based on our searching results for publications between 1992 and 2018, most validation of sepsis studies focused on adult patients (aged ≥ 18 years) and no publications were conducted specifically for validation of algorithms for identifying neonatal sepsis. There were some common issues in most existing validation of sepsis studies. For example, no consensus currently exists regarding which ICD-9 or ICD-10 codes should be applied to define sepsis in health administrative databases. (9, 13) Reported sensitivity and specificity differed considerably across validation studies of sepsis

because of the scope of ICD codes, varied definitions of reference standards and methods for developing the coding algorithms. (14) Additionally, sepsis was largely under-coded in administrative data using ICD-9 or ICD-10 coded case definitions, thus under-ascertaining the true incidence of sepsis. (13) Hence, there is a need for validating a case-finding algorithm to support research on neonatal bacterial sepsis identified using health administrative data.

1.2 Study Objectives

This was a population-based retrospective validation study of neonates with bacterial sepsis in Ontario, Canada. The specific objectives of this study were to:

(1) Develop and validate case-finding algorithms to accurately identify cases of bacterial sepsis in neonates in Ottawa using data from the Ottawa Hospital Data Warehouse (OHDW);

(2) Apply the validated case-finding algorithm to the infant population of Ontario to calculate incidence rates, and estimate temporal trends of neonatal bacterial sepsis by Poisson regression models.

CHAPTER 2: LITERATURE REVIEW

2.1 Clinical Features of Neonatal Sepsis

Sepsis is a syndrome involving physiologic, pathologic, and biochemical abnormalities induced by infection. (15) Neonatal sepsis (i.e., sepsis occurring within the first 28 days of life) can be the result of infections with bacterial, viral or fungal (mostly yeast) microorganisms, with bacteria being the predominant cause. (16) In one large case series of septic preterm infants, rates of infections by Gram-negative organisms (55%) were higher than those due to gram-positive organisms (38%), fungal pathogens (5%) and other unclassified organisms (2%). (17)

Depending on the timing of onset of the infection, neonatal sepsis is classified as either early-onset sepsis or late-onset sepsis. Clinical manifestations of early-onset sepsis appear within the first 72 hours of life, and usually represent vertical mother-to-infant transmission just prior to, or during vaginal delivery. (18) Typical risk factors of early-onset sepsis include low gestational age (< 32 weeks of pregnancy), low birth weight (≤1500 grams), chorioamnionitis, premature rupture of the membranes more than 18 hours and Group B streptococci (GBS) colonization. (19) Cases of late-onset sepsis presents clinical manifestation of infection after 72 hours of life, and are attributed to organisms acquired from interaction with the hospital environment or the community. (20) Similar to early-onset sepsis, the rate of late-onset sepsis is inversely proportional to gestational age and birth weight. (19)

In cases of early-onset sepsis, normally the flora of the maternal birth canal is the causative agent. *Escherichia coli (E coli)* is the most commonly identified etiologic agent causing neonatal sepsis, followed by GBS and more rarely, *Staphylococcus aureus* or *Listeria*. (19, 21, 22) Late-onset sepsis is also associated with GBS, E coli, other Gram-negative aerobes, or Listeria monocytogenes infection. (16) Further, there is a known burden of late-onset sepsis caused by

Gram-negative organisms in resource-poor areas of the worlds. (23) Viral infections such as herpes simplex virus and enteroviruses are more frequently associated with late-onset sepsis compared to early-onset sepsis. (24) Fungi, notably yeast, have also been implicated in neonatal sepsis, and are usually acquired during prolonged hospital stay of preterm neonates. (16)

2.2 Epidemiology of Neonatal Sepsis

Neonatal sepsis has become one of the most frequent causes of death. (1) In 2015, infectious diseases accounted for more than 50% of all deaths in neonates and among these, neonatal sepsis was one of the most frequent causes of death, responsible for an estimate 401,000 neonatal deaths per year. (1, 25) In 2010, 7.6 million children (< 5 years old) died worldwide and neonatal deaths accounted for 40% of the total lives lost, predominantly because of infections including sepsis. (26) In developed countries, four out of every ten infants with sepsis die or experience major disability, including significant permanent neurodevelopmental impairment, in spite of major advances in neonatal care and increasing research. (27) In a retrospective hospital discharge database study in Canada, the incidence of sepsis among children (aged 0-17 years) ranged from 0.13 to 0.15 per 1000 person-years between 2004 and 2008. Among those patients with any diagnosis of sepsis, 56.3% were neonates. (28)

Among all age groups, preterm infants experience the highest incidence and mortality of sepsis. (29, 30) It has been estimated that sepsis in prematurely born neonates is up to 1000-fold more common compared with term infants, and is associated with higher rates of mortality and life-long neurodevelopmental impacts. (31-33)

According to a systematic review published in 2018, the estimated global burden of neonatal sepsis is 3.0 million cases annually. (2) Wide variation exists in estimates of neonatal sepsis burden between countries of different income levels. (16) Across all age groups, the financial burden due to sepsis in the United States (US) has been estimated at more than $24 billion, representing 6.2% of US total hospital costs in 2013. In Europe and Canada, the daily costs of hospital care of a septic patient were estimated to be between $645 and $939 (in US dollars) in 2000. (1) In low- and middle-income countries, the epidemiological burden of sepsis is likely to be much higher, though there is a lack of available data.

2.3 Case Definition of Neonatal Sepsis

Although a case definition is paramount for assessing the epidemiology and impact of any disease (34), prior to 1991, there was a lack of consensus in terms of the clinical criteria used to define sepsis and related conditions. In 1992, the American College of Chest Physicians/ Society of Critical Care Medicine (ACCP/SCCM) Consensus Conference used the term systemic inflammatory response syndrome (SIRS) to describe neonatal sepsis. (35) Recognizing limitations with the existing definitions at that time, the 2001 International Sepsis Definitions Conference tried to improve the definitions by expanding the list of signs and symptoms of sepsis to reflect clinical bedside experience; however, no new definitions for sepsis were introduced. (36) In 2015, an international consensus recommended that sepsis should be defined as “life-threatening organ dysfunction caused by a dysregulated host response to infection”. (15)

Compared with other age groups, neonates (newborn infants up to and including 28 days after birth) have limited natural immunity due to immature immune systems and, therefore, have distinctive responses to infections. (19) In 2015, the Global Alignment of Immunization Safety Assessment in Pregnancy group, part of the Brighton Collaboration, proposed a standardized definition of neonatal infections, which included neonatal bacteremia and sepsis (of early or late onset), meningitis, pneumonia and other respiratory infections. (37) Three levels of diagnostic certainty have been proposed, as shown in **Table 2.3-1**. The microbiological confirmation of infection is included in level 1 of Brighton Collaboration Case Definition, which historically constitutes the "gold standard" for the presence of neonatal sepsis. (37) Laboratory-confirmed neonatal sepsis is diagnosed by isolating the causative agent from a normally sterile body site (*i.e.*, blood, CSF, urine, pleural, joint and peritoneal fluids). (38) Recent studies found common laboratory tests may have limited diagnostic accuracy for neonatal sepsis and some sepsis evaluations prompted by concerning clinical signs are associated with negative blood culture results. (39-42) Therefore, microbiological confirmation is usually combined with clinical criteria to make sure the accurate diagnosis of sepsis. (15, 34, 37) For example, the number of clinical criteria was chosen included in level 2 of Brighton Collaboration Case Definition. (37) It is common for infants with evidence of infection (*e.g.* temperature, WBC count, tachycardia) and organ dysfunction to receive course of antibiotics regardless of culture results. (43)

Table 2.3-1. Brighton Collaboration Case Definition for the diagnosis of neonatal invasive blood stream infections: bacterial/fungal/viral (reproduced from Vergnano et al., 2016) (37)

Level 1	Level 2	Level 3
Recognized pathogen [a] identified using a validated method and from a normally sterile site [b]. If an organism normally considered non-pathogenic is isolated from blood cultures [a]: Level 1 requires its identification from at least 2 blood cultures taken from two different sites, or at 2 different times, PLUS 1 of the criteria as per level 2 of evidence.	Not meeting Level 1 of evidence AND 3 or more criteria: • Temperature ≥37.5 °C or <35.5 °C • Tachycardia or new or more frequent episodes of bradycardia • New or more frequent episodes of apnea or increased oxygen requirement or increased requirement for ventilatory support • Lethargy or moving only when stimulated or hypotonia or irritability • Difficulty in feeding or abdominal distention • Pallor or poor perfusion or hypotension • Abnormal White Cell Count or I/T ratio > 0.2 • Abnormal platelet count • Increased inflammatory markers (CRP, procalcitonin) • Metabolic acidosis as defined by a base excess (BE)	Not meeting Level 1 or 2 of evidence AND 2 or more of the following criteria: • Temperature ≥37.5 °C or <35.5 °C • Tachypnea or severe chest indrawing or grunting or cyanosis • Change in level of activity • History of feeding difficulty • History of convulsions

a. See list of pathogens and non-pathogens in Appendix 1.
b. Sterile site: blood, sterile urine (catheter urine or supra-pubic aspirate), pleural fluid, ascitic fluid, broncho-alveolar lavage, bone biopsy, synovial fluid.

2.4 Administrative Data and Disease Surveillance for Neonatal Sepsis

2.4.1 Health Administrative Data

Health administrative databases are computer data files created when a health-care transaction occurs, that are generally for the purposes of health care management, payment and monitoring

functions. (44) These databases have become valuable sources of information for disease surveillance, since they are often collected routinely, cover wide geographic areas, and have a relatively complete capture of all patient encounters with the health care system; however, they do not have as detailed clinical information as some other data resources (e.g., electronic clinical charts, laboratory records). (45)

In Ontario, a repository of Ontario's health administrative databases is maintained at ICES, which includes administrative data for over 13 million people covered by the Ontario Health Insurance Plan (OHIP). Most health care records extend back to 1991. Health records are linkable across databases via a unique identification number (ICES Key Number [IKN]), which is an encrypted identifier based on the Ontario health card number. Each database is comprised of individual data files for each beneficiary including hospital separations (inpatient records), physician billings (in- and out-patient physician services), and urgent health care visits to an emergency room. These databases typically contain detailed clinical and demographic data, including the use of standardized diagnostic codes to characterize the reasons for the health care encounter.

2.4.2 Defining Neonatal Sepsis in Health Administrative Data

The main diagnostic coding system used for medical record abstraction in health administrative databases in Canada is the International Classification of Diseases (ICD). While the ICD was originally developed to codify and compare mortality statistics and causes of death, but over time, it has expanded to include codes for virtually all diseases as well as codes for specific signs, symptoms, pathogens, and external causes of injury or diseases. (46) When a patient is admitted

to hospital and treated for one or more medical problems, the medical diagnoses and treatment procedures pertinent to the admission are documented by hospital staff in the patient's medical chart. After the patient is discharged, the chart is transferred to a health records department, where trained coders go through the clinical notes and assign ICD-10 diagnostic codes and procedural codes based on the information documented in the chart.

Canada adopted the International Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) in 2001, using an enhanced version (ICD-10-CA), developed by the Canadian Institute for Health Information (CIHI). There is a total of 12,420 codes in ICD-10 compared to 6,969 in the previous version, ICD-9, intended to allow for more precise and comprehensive capture of clinical information, and facilitate improved international comparability. (47)

2.4.3 Validity of Health Administrative Data

There is increasing interest among epidemiologists and policymakers in the use of health administrative data to identify patients with sepsis. (49, 50) However, it is important to assess the validity of case ascertainment using health administrative data given that these data are collected for the purpose of managing and evaluating the healthcare system, not specifically for disease surveillance or research purposes. (8)

Studies have been performed utilizing administrative data to assess the validity of case definitions of sepsis utilizing administrative data. The main bibliographic databases (MEDLINE, Embase, Pubmed, CINAHL, Psycinfo, Scopus and web of science) were searched for sepsis validation publications between 1992 and 2018. The full search strategy use to identify studies for this

literature review is provided in Appendix 5. Studies were identified that compared the accuracy of ICD-9 or ICD-10 codes for sepsis to a reference standard using administrative data and reported any diagnostic accuracy measures (i.e., sensitivity, specificity, positive predictive value or negative predictive value). We identified 12 articles, most of which focused on adult patients (aged ≥ 18 years). (9, 13, 51, 52) A few of the identified validation studies just specified the study population as all patients without specifying the age distribution (53-55). None of above 12 publications through the above searching strategy were conducted specifically for validation of algorithms for identifying neonatal sepsis.

There were some common issues in most existing validation of sepsis studies. First, no consensus currently exists regarding which ICD-9 or ICD-10 codes should be applied to define sepsis in health administrative databases. (9, 13) According to a systematic review which examined the validity of published ICD-coded case definitions in administrative data from 1992-2014, a total of 38 sepsis case definitions were tested, which included over 130 different ICD codes. (13) For instance, Gaieski *et al* examined four different methods by using different ICD-9 codes to capture cases of severe sepsis and identified up to a 3.5-fold variation in incidence across four ICD-coded case definitions applied to the health administrative databases. (56) Second, reported sensitivity and specificity differed considerably across validation studies of sepsis because of the scope of ICD codes, varied definitions of reference standards and methods for developing the coding algorithms. (14) For example, reported sensitivities in validation studies for sepsis coding within administrative databases have ranged from 5.9% to 82.3% and positive predictive value (PPV) ranged from 5.6% to 100%, with a median of 50%. (51, 52, 54, 57-59) Third, it was reported that sepsis was largely under-ascertained in administrative data using ICD-9 or ICD-10 coded case definitions, thus under-ascertaining the true incidence of sepsis. (13) For example, a population

based study in Denmark from 2010 to 2011 suggested that using ICD-10 discharge diagnosis to identify cases may result in an up to 7-fold underestimation of the incidence of traditional severe sepsis compared to medical record review. (60) Last, most previous studies did not test algorithms against clinical data from medical chart review, and thus they could be more accurately referred to as database agreement studies, rather than true clinical chart review validation studies. (2, 13)

In view of the limitations on existing validation studies, the need for developing and validating an enhanced algorithm using health administrative data is urgent, to increase the accuracy of case capture for a diagnosis of neonatal sepsis to allow for population-based surveillance and research on neonatal sepsis.

CHAPTER 3: METHODS

3.1 Overview of Study Design

The initial sections of this Chapter (Sections 3.1 and 3.2) provide information that is relevant to both Objectives 1 and 2. Objective-specific methods are provided in subsequent sections. Objective 1 was to develop and validate coding algorithms based on The Ottawa Hospital Data Warehouse (OHDW) databases and Objective 2 was to apply the prefered coding algorithm identified to population-based health administrative databases in ICES to calculate incidence rates of neonatal bacterial sepsis and estimate temporal trends of neonatal bacterial sepsis in Ontario.

Table 3.1-1. Overview of Study Design for Objectives 1 and 2.

Objective 1	**Chart Review**	Define the study population in Objective 1 using the OHDW databases.
		Identify the true-positive reference cohort by meeting diagnostic criteria I (microbiological confirmation) and II (prolonged sepsis-related antibiotic administration). Identify the true-negative reference cohort by representing all infants in the study population who do not meet the diagnostic criteria for the true-positive cohort.
	Algorithm Derivation and Evaluation	Examine the ability of ICD-10-CA codes to identify true-positive cases, as well as coding practices in true-negative reference cohorts to develop and refine coding algorithms for identification of neonatal bacterial sepsis.
		Select the best coding algorithm by comparing sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio and c-statistic across tested algorithms.
Objective 2	**Algorithm Application**	Define the study population in Objective 2 using the ICES databases. Apply the selected coding algorithm to calculate incidence rates within the study population in Ontario; Apply Poisson regression models to estimate temporal trends of neonatal bacterial sepsis in Ontario.

3.2 Ethics and Privacy Statements

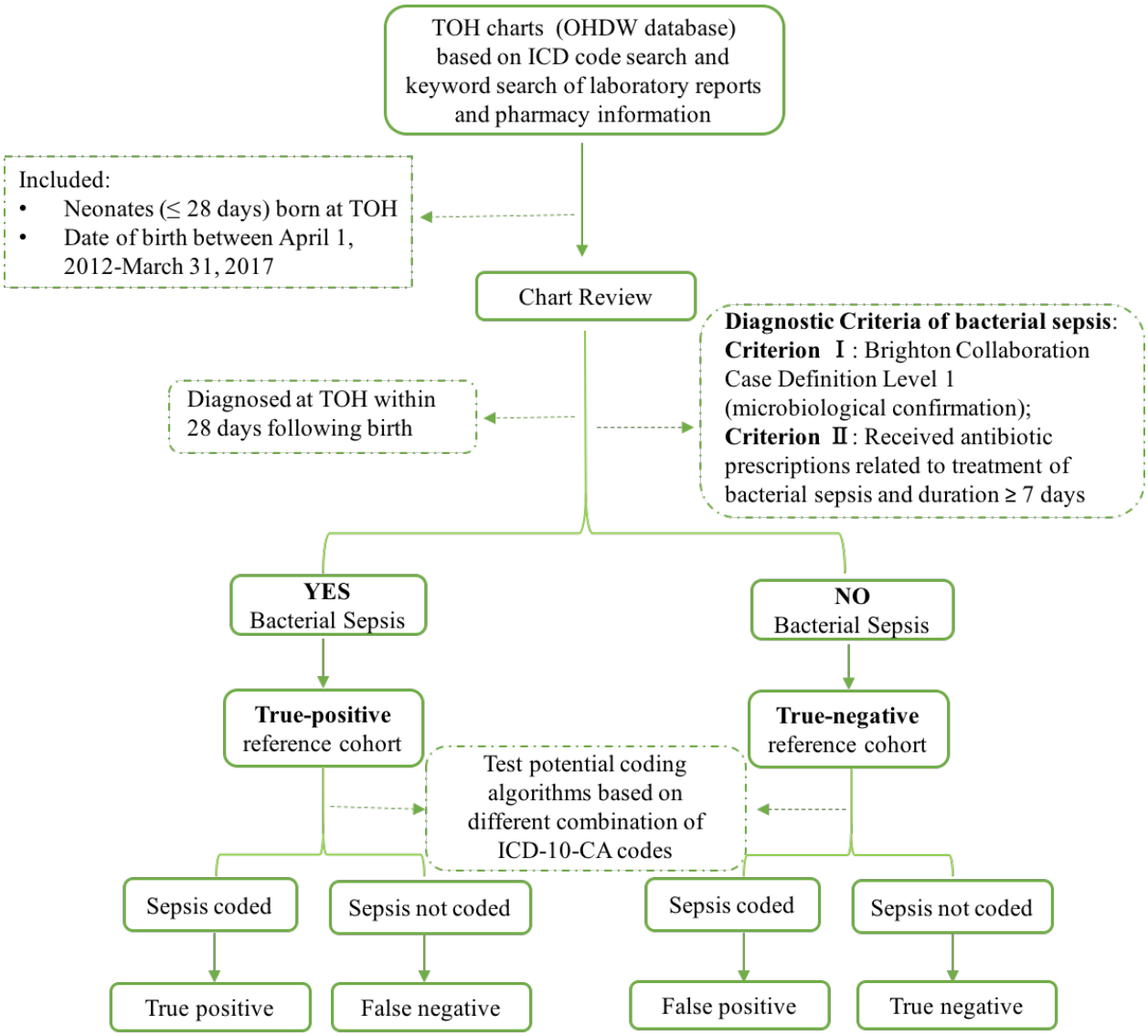
This study was approved by the Research Ethics Board (REB) of the Ottawa Health Sciences Network (OHSN) and the Children's Hospital of Eastern Ontario (CHEO), as well as by the ICES Privacy Officer. The datasets from this study are held securely at the Ottawa Hospital (TOH) and ICES. All administrative data were de-identified prior to analysis and linked using encrypted unique identifiers. To protect privacy, cell sizes fewer than six individuals were suppressed and reported as n<6.

3.3 Objective 1: Develop and Validate Coding Algorithms

3.3.1 Overview

Figure 3.1-1 demonstrates the steps taken in order to accomplish the Objective 1 to identify cases of neonatal bacterial sepsis and develop coding algorithms.

Figure 3.3-1. Flow diagram of algorithm development for Objective 1.



3.3.2 Data Sources

To conduct this validation study and identify neonatal bacterial sepsis, the OHDW databases housed in the secure network environment at TOH were used. TOH includes three campuses, two of which provide obstetrical and neonatal care (the General and Civic sites). The OHDW is a

relational database containing information from multiple TOH information systems, including data relating to hospital admissions, outpatient visits, emergency department visits, day care visits and procedures for all patients seen at TOH. In addition, the OHDW contains a clinical data repository of laboratory results, radiology and pathology reports, and pharmacy data for all patients seen after April 1, 2002. The following OHDW data tables were used in this study:

- Encounter Table: This contains patient-level clinical encounter information, including patient demographics (at the time of the encounter), inpatient and outpatient details, and discharge disposition.
- Health Record Abstract Table: This table describes general abstract information, which captures administrative, clinical and demographic information on hospital discharges (including deaths, sign-outs and transfers).
- Diagnoses Table: This contains patient-level diagnostic information including diagnosis codes. The diagnosis coding practice was changed from ICD-9-CA to ICD-10-CA in 2002, prior to our derivation time frame.
- Pharmacy Service Table: This contains pharmacy information for drugs received during health care encounters. It includes the start and end date/time of administration, route and frequency of administration, service status (i.e. discontinued, cancelled), brand name/ingredient for intravenous (IV) and drug dispensation location (i.e. pharmacy, ward stock).
- Pharmacy Service Ingredient Table: This describes drug ingredient information, including drug identification number (DIN), order description, dose unit and amount and therapeutic category (e.g., local anesthetics, anti-viral, antibiotic).

- Laboratory Service Table: This describes laboratory services received during health care encounters, including type of service (e.g., blood/urine culture, PCR), date/time of service, lab description, lab result, abnormal code and location of service (e.g. Civic campus, General campus), exam reports, and microbiology test results.

3.3.3 Study Population and Creation of True-Positive and True-Negative Reference Cohorts

The study population included all neonates (≤ 28 days) born at TOH with date of birth between April 1, 2012 to March 31, 2017. Since the OHDW in TOH was a fully electronic health record database, a data extraction algorithm was programmed in SAS to extract the electronic data into a dataset where each record could be reviewed and a determination made with respect to case status. Any record for health care received at TOH for neonates was examined. We identified true-positive and true-negative reference cohorts within the study population to support the development and validation of coding algorithms. Cases of neonatal sepsis diagnosed at TOH within 28 days following birth were used as the true-positive reference cohort in our study. We used several database search strategies to ensure that we captured all of cases of potential neonatal bacterial sepsis in TOH during the study period.

In this study, we had microbiology test results including pathogenic and non-pathogenic bacteria from culture samples, which enabled the identification of neonates meeting Level 1 of the Brighton Collaboration Case Definition. However, a small fraction of cases with microbiological confirmation is known to be false positives due to contamination of sample collection and culture media. It is common for infants with evidence of infection and organ dysfunction to receive a

course of antibiotics regardless of culture results. Therefore, we combined microbiological confirmation with a practical approach (sepsis-related antibiotic administration) as the diagnosis criteria based on expert clinical guidance (DM Pediatric Intensivist) to improve accuracy of diagnosis. Therefore, any neonates who met both the Diagnostic Criterion Ⅰ (microbiological confirmation) and Criterion Ⅱ (prolonged sepsis-related antibiotic administration) were included in the true positive reference cohort. Since free text clinical notes were not available in the electronic chart review and prolonged sepsis-related antibiotic administration alone didn't contain sufficient information, we were not able to evaluate Level 2 of Brighton Collaboration Case Definition. A more extensive clinical chart review following up from our study is underway by using sepsis-related antibiotic administration as a screening tool to evaluate Level 2 of Brighton Collaboration Case Definition.

To evaluate Diagnostic Criterion Ⅰ, we checked microbiology test results from the patient data by comparing them with the list of recognized pathogens and non-pathogens in the Level 1 of Brighton Collaboration Case Definition (**Table 2.3-1**), including identification of pathogenic and non-pathogenic bacteria from culture samples. If either of the following criteria was satisfied, this patient met the Diagnostic Criterion Ⅰ:

- Recognized pathogenic bacteria (Appendix 1) were identified from a normally sterile site (blood, sterile urine, pleural fluid, ascitic fluid, broncho-alveolar lavage, bone biopsy, synovial fluid).
- Non-pathogenic bacteria (Appendix 2) were identified from two blood cultures taken at different time periods.

To evaluate Diagnostic Criterion Ⅱ, we identified patients with sepsis-related antibiotic administration and computed their treatment duration.

- All therapeutic categories used by the study population (Appendix 3) were listed based on the American Hospital Formulary Service (AHFS) Pharmacologic-Therapeutic Classification system. We only included therapeutic categories of antibiotic prescriptions related to treatment of bacterial sepsis. Any neonates who received antibiotics that were unrelated to treatment of bacterial sepsis were not included in the true-positive reference cohorts.
- The duration of antibiotic treatment was calculated for all neonates who received antibiotic prescriptions related to treatment of bacterial sepsis. If a particular treatment period was equal to or greater than seven days, we classified the neonate as having met the Diagnostic Criterion Ⅱ for the true positive reference cohort.

The true-negative reference cohort was defined as all infants in the study population who did not meet both Criterion I and II of the diagnostic requirements for the true-positive cohort.

In the sensitivity analyses, we assessed a true-negative cohort that excluded neonates who only met Criterion Ⅰ (i.e., with microbiological confirmation, but no prolonged sepsis-related antibiotic administration) as well as those who only met Criterion Ⅱ (i.e., with prolonged sepsis-related antibiotic administration, but no microbiological confirmation) to determine these cases' impact on potential coding algorithms.

For this study, we also reported characteristics of the true-positive and true-negative reference cohorts, including: infant sex, gestational age (weeks), birth weight (grams) and hospital length of stay (days). Preterm birth was subdivided on the basis of gestational age according to the World

Health Organization (WHO) definition (61): extremely preterm (<28 weeks); very preterm (28-31 weeks); moderate or late preterm (32-36 completed weeks of gestation).

3.3.4 Algorithm Development

Once the study population was classified as true-positive and true-negative cohorts on the basis of diagnostic criteria Ⅰ and Ⅱ in the OHDW database, the development of coding algorithms proceeded in two steps.

First, we explored the related sepsis ICD-10-CA codes on the hospital abstracts that were frequently documented in correctly classified, false-negative and false-positive cases. The specific neonatal (P360, P361, P362, P363, P364, P365, P368, P369) and adult (A021, A227, A267, A327, A400, A401, A402, A403, A408, A409, A410, A411, A412, A413, A414, A4150, A4151, A4152, A4158, A4180, A4188, A419, A427, A5486, B377, O85002, O85004, O85009) sepsis codes were shown in Appendix 4. Besides neonatal and adult sepsis codes, we also explored other sepsis-related codes, which could also have potentially been commonly used codes on false negative records. These codes were then examined to determine which codes might indicate bacterial sepsis and could be included in the new definition based on clinical knowledge of the resulting diagnosis. All these codes identified were added sequentially to develop Coding algorithms 1-4. The best coding algorithm was chosen for the next step based on the performance.

Second, the impact of individual codes was explored based on inspection of the number of false positive and true positive contributed by each. The ratios of the false positive to true positives were used to prioritize codes and determined which code was removed from the coding algorithm which was determined from the last step to produce new coding algorithms.

3.3.5 Statistical Analysis

All potential coding algorithms were evaluated using the true-positive and true-negative reference standard groups in the study cohort for their ability to correctly classify neonates with and without bacterial sepsis. Measures of diagnostic accuracy for those algorithms were calculated, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio, and c-statistic. Details of the how these measures were computed are illustrated in **Table 3.3-1**. Exact binomial 95% confidence intervals (CIs) were calculated for estimated proportions for sensitivity, specificity, PPV, NPV and positive likelihood ratio. The c-statistic, which is analogous to the area under the receiver operating characteristic curve, was computed for each coding algorithm to provide a measure of an algorithm's ability to correctly classify bacterial sepsis cases vs. non-cases. The c-statistic values were calculated by using "proc logistic" in SAS, with true case status as the outcome and algorithm case status as the predictor. The c-statistic can range from 0.5 to 1.0, with higher values indicating better discriminating algorithms. When an algorithm has no ability to discriminate, the expected value of c-statistic is 0.5. A c-statistic much smaller than 0.5 is evidence of an algorithm model that can discriminate but is improperly parameterized (higher predictions are associated with fewer cases). (62) Based on the performance characteristics of each algorithm, we set a threshold of minimum acceptable sensitivity of 70%, and then selected the algorithm with the maximum PPV given minimally acceptable sensitivity was achieved. The algorithm with the best performance using the above criteria was selected for use completing Objective 2.

Table 3.3-1. Measures of diagnostic accuracy for all potential coding algorithms.

Coding Algorithm [b]	**Gold Reference Standard** [a]		
	Condition Present	**Condition Absent**	
Sepsis coded	True Positive (TP)	False Positive (FP)	PPV=TP/(TP+FP)
Sepsis not coded	False Negative (FN)	True Negative (TN)	NPV=TN/(FN+TN)
	Sensitivity=TP/(TP+FN)	Specificity=TN/(FP+TN)	Positive likelihood ratio = sensitivity / (1-specificity)

[a] Determined on the basis of clinical information on the medical charts.
[b] Determined on the basis of ICD-10-CA codes.

3.4 Objective 2: Application of the Algorithm in the Ontario-wide population using ICES data to Estimate Incidence Rates and Trends in Neonatal Bacterial Sepsis

3.4.1 Introduction

The aim of this objective was to apply the optimal validated case-finding algorithm developed in Objective 1 in order to estimate population-based incidence rates of neonatal bacterial sepsis in Ontario. For the purpose of demonstrating the application of the algorithm developed in Objective 1, we have carried out this objective, with the caveat that it is important to cautiously interpret the results generated in this objective given the poor performance of the case-finding algorithm.

3.4.2 Study Population and Data Sources

The study population in Objective 2 included all preterm neonates with date of birth between April 1, 2003 to February 28, 2017 in Ontario. Since Canada adopted the ICD-10 coding system in 2001 and the transition from ICD-9 to ICD-10 was completed in 2003, it was reasonable to apply the algorithm in the data beginning in 2003. We used databases housed in the secure network environment at ICES.

- **MOMBABY Database**. As an ICES-derived database, the MOMBABY Database has the inpatient hospital admission records of delivering mothers and their newborns. It includes all inpatient admission records from the Discharge Abstract Database (DAD) for mothers and their newborns delivered from fiscal year 2002/03 onward. Each record in the MOMBABY Database contains the unique encrypted health card number of the mother and baby, as well as a reduced set of clinical variables. Since midwifery home births comprise <2% of all Ontario births and most obstetrical deliveries in the province of

Ontario take place in hospital (63), we were able to use this database to capture the vast majority of population of infants born in the province during our study period.

- **Discharge Abstract Database (DAD)**. The DAD database captures administrative, clinical, and demographic information for hospital discharges (inpatient acute, chronic, rehabilitation) and day surgeries. Specifically, the key variables include diagnosis codes, admission/discharge dates, type of admission, discharge disposition, demographic data (e.g. gender, postal code etc.). Up to 25 clinical diagnoses identified with ICD-10-CA codes are available on each record. (64) The DAD was linked to MOMBABY database to provide the diagnosis codes of hospitalization for newborns.
- **National Ambulatory Care Reporting System** (NACRS). The NACRS database captures information on patient visits to all hospital- and community-based ambulatory care facilities. Data about visits are collected at the time of service in participating facilities. The major variables include patient identifiers, emergency department visits, day surgery, and outpatient clinic visits, demographic information, diagnostic codes and procedure codes for the services that were provided during the patient's visit. The NACRS database was linked to the MOMBABY database to provide the diagnosis codes of emergency department visits for newborns.

3.4.3 Incidence Rates of Neonatal Bacterial Sepsis

For the purpose of demonstrating the application of the algorithm, we computed incidence rates for neonatal bacterial sepsis as the number of new sepsis cases identified by our coding algorithm 5, divided by the number of preterm neonates at risk in Ontario in a defined period of time (April

1, 2003 to February 28, 2017). The preferred ICD-10-CA coding algorithm determined in Objective 1 was applied to identify the cases of neonatal bacterial sepsis within the health administrative databases (MOMBABY, DAD and NACRS databases). The incidence rates were calculated per 1000 live births. Incidence rates of bacterial sepsis were calculated overall in the whole study population, and also in subgroups of infant sex and degree of prematurity: extremely preterm (<28 weeks), very preterm (28-31 weeks) and moderate to late preterm (32-36 weeks) in each year. Incidence rates were also calculated by gestational age (weeks). Meanwhile, assuming the sensitivity and specificity of the algorithm remains consistent in the applied settings and time period, the estimated incidence rate could be adjusted based on the following equation, according to the information from the literature. (65) But, it is not generally advisable to do this in practice.

$$p = \frac{\tau + \beta - 1}{\alpha + \beta - 1}$$

$\boldsymbol{p}$: adjusted incidence rate; $\boldsymbol{\tau}$: estimated incidence rate; $\boldsymbol{\alpha}$: sensitivity; $\boldsymbol{\beta}$: specificity.

3.4.4 Poisson Regression Analysis

The Poisson regression model has been widely used for modeling counts and rates of disease. (66) In this study, we used Poisson regression to estimate temporal trends of neonatal bacterial sepsis in Ontario by year of birth, infant sex, gestational age (weeks), as well as categories of preterm birth.

3.4.4.1 Assumption Assessment and Correction for Overdispersion

Before analysis, we assessed whether the assumptions of the Poisson model were met in our data. In Poisson regression, the dependent (Y) variable is an observed count that follows the Poisson distribution, the probability of y events with the formula

$$\Pr(Y = y|\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \quad (y = 0,1,2 \ldots) \quad \ldots\ldots [1]$$

The Poisson distribution has the property that the sum of independent Poisson random variables is also Poisson. Specifically, if Y_1 and Y_2 are independent with $Y_i \sim P(\mu_i)$ for i = 1, 2 then

$$Y_1 + Y_2 \sim P(\mu_1 + \mu_2) \quad \ldots\ldots [2]$$

The key implication of equation [2] is that individual and grouped data can both be analyzed with equivalent results. (67) According to the properties of the Poisson distribution, Poisson regression models should satisfy the assumptions that the responses y_i must follow the Poisson distribution, including the requirement that the expected value of the mean should be equal to the variance. (68) In other words, the variance of the response variable approximates the mean of the response variable; i.e., $V\big(E(Y)\big) = \emptyset E(Y)$, where $\emptyset$ is referred to as the dispersion parameter. The dispersion parameter is considered as the variation unaccounted for by the fitted model. If $\emptyset$>1, then the model is considered to be overdispersed, which means the variance is larger than the mean.

In this study, the response variable y_i was the counts of cases of neonatal bacterial sepsis in each year during the study period. We assessed the model assumption by exploring the mean-variance relationship to test whether the distribution of sepsis cases (y_i) followed a Poisson random process in our data. Since the consequence of such overdispersion is that standard errors are incorrectly estimated (69), we planned to take measures if there was evidence of any excess variation. There

are several options for addressing overdispersion, one of which is to correct the standard errors of the estimates. (70) More specifically, dispersion parameter ∅ is estimated as the ratio of the deviance to the degrees of freedom (Value/DF) by using scaled Pearson Chi-Square criteria to accommodate the excess variation. The ratio of the Value/DF should be approximate one, whereas large ratio values indicate an over-dispersed response variable. The goodness of fit was assessed for the model before and after applying this correction.

3.4.4.2 Model Building

Poisson regression models are generalized linear models with the logarithm as the (canonical) link function, that is $g(\mu_i) = log(\mu_i) = \eta_i$. (71) Hence, the Poisson regression model is expressed as

$log(\mu_i) = \log(t_i) + \beta' X_i + \beta_0$..... [3]

Where β is a vector of regression coefficients, β_0 is the model intercept, X_i is a vector of covariates for subject I, and $\log(t_i)$ is the natural logarithm of the denominator, referred to as an offset.

In this study, Poisson regression was applied to estimate temporal trends of neonatal bacterial sepsis in Ontario by year of birth, infant sex, gestational age (weeks) and categories of preterm birth. The vector of covariates in equation [3] could be expressed as year of birth, infant sex, gestational age (weeks) and category of preterm birth specifically in our model. To be more exact, the Poisson regression models in this study would be:

$\log[E(Y)] = \beta_0 + \sum_{j=1}^{n} \beta_i Year_j + \log(births)$..... [4]

$\log[E(Y)] = \beta_0 + \sum_{j=1}^{n} \beta_i Year_j + \beta_r Sex_i + \log(births)$..... [5]

$$\log[E(Y)] = \beta_0 + \sum_{j=1}^{n} \beta_i Year_j + \beta_h (gestational\ age)_i +$$

$$\log(births) \quad [6]$$

$$\log[E(Y)] = \beta_0 + \sum_{j=1}^{n} \beta_i Year_j + \beta_k (\text{category of preterm birth})_i +$$

$$\log(births) \quad [6]$$

The reference groups were set "year of birth" as 2016, "infant sex" as female, "gestational age" as 36 weeks, "category of preterm birth" as moderate to late preterm (32-36 weeks) group. Where Y denote the number of newly initiated bacterial sepsis cases in each year, log(births) was an offset for the number of preterm births at risk in each year in Ontario, β_0 is the regression coefficient estimated from the models when covariates were set as 0, β_i is the regression coefficient for the effect of year of birth, β_r is the regression coefficient for the effect of infant sex, β_h is the regression coefficient for the effect of gestational age, β_k is the regression coefficient for the effect of category of preterm birth. Model parameters were estimated by using the method of maximum likelihood estimation. Temporal trends in incidence rate were estimated by the multivariable Poisson regression model across the study period. 95% confidence intervals (CI) and p-values for relative incidences were reported, with intervals excluding 1 and P-values of <0.05 considered to be evidence supporting statistically significant effects.

CHAPTER 4: RESULTS

4.1 Objective 1: Develop and Validate Coding Algorithms

4.1.1 Defining Reference Cohorts

First, to meet Diagnostic Criterion Ⅰ (as described in Section 3.3.2), we reviewed microbiology test results for all neonates and flagged both pathogenic and non-pathogenic bacteria identified in the Brighton Case Definition (**Table 4.1-1**).

Table 4.1-1. Recognized pathogens and non-pathogens identified in the study cohort.

Recognized pathogens (Bacteria)	
Acinetobacter baumannii	Klebsiellapneumonia
Bacillus cereus	Morganellamorganii
Citrobacter freundii	Pseudomonas aeruginosa
Enterococcus faecalis	Serratia marcescens
Enterobacter aerogenes	Staphylococcus aureus (Methicillin-sensitive and -resistant MSSA or MRSA)
Enterobacter cloacae	
Escherichia coli	
Haemophilusinfluenzae	Streptococcus agalactiae or group B streptococcus
Klebsiellaoxytoca	
Non-pathogenic organism (Bacteria)	
Coagulase-negative staphylococci (Coagulase-negative Staphylococcus)	Streptococcus oralis
Streptococcus mitis	Streptococcus salivarius

Next, we screened all antibiotic administration based on AHFS Pharmacologic-Therapeutic Classification System (Appendix 3) and identified prolonged (≥ 7 days) bacterial sepsis-related antibiotic administration, which is part of the Diagnostic Criterion Ⅱ (as described in Section 3.3.2). **Table 4.1-2** lists therapeutic categories for bacterial sepsis-related antibiotic administration used in this study cohort.

Table 4.1-2. Therapeutic categories for bacterial sepsis-related antibiotic administration.

Code	Therapeutic Category*
081202	Aminoglycosides
081206	Cephalosporins
081207	Misc.Beta-Lactam antibiotics
081212	Erythromycins
081216	Penicillins
081220	Sulfonamides (systemic)
081228	Misc. antibiotics
083600	Urinary anti-infectives

* Based on AHFS Pharmacologic-Therapeutic Classification system.

On the basis of the Diagnostic Criteria Ⅰ and Ⅱ, we identified the true-positive and true-negative reference cohorts. There were 129 cases who met the Diagnostic Criterion Ⅰ and 314 cases who met the Diagnostic Criterion Ⅱ. Among them, 101 cases met both diagnostic criteria. **Table 4.1-3** shows the characteristics of the true-positive and true-negative reference cohorts by sex, gestational age (weeks), birth weight (grams) and hospital length of stay (days). Overall, there were a total of 31609 neonates born between April 1, 2012-March 31, 2017 in the full study cohorts. Of these, 101 (0.3%) neonates were classified as having sepsis (true-positive reference cohort) and 31508 (99.7%) were classified as not having sepsis (true-negative reference cohort). Of the neonates diagnosed as having sepsis included in the study, 52.5 % were male. There was different mean birth weight of 2342.6 (95% CI, 2211.2 - 2474.0) between the true-positive and true-negative cohorts (true positive: 941.4 grams; true negative: 3284.0 grams). Infants in the true-positive reference cohort had a median length of hospital stay of 46 days (interquartile range, 28-84), while those in the true-negative reference cohort had a median length of hospital stay of 2 days (interquartile range, 1-3). All neonates with bacterial sepsis in this study cohort were born at preterm gestation. In the true-negative reference cohort, 13.2% of neonates were preterm. The

average gestational age was 26.1 weeks for the true-positive cohort and 38.4 weeks for the true-negative cohort. The majority of bacterial sepsis cases (75.3%) were in the extremely preterm category (<28 weeks).

Table 4.1-3. Characteristics of the true-positive and true-negative reference cohort in the full study population.

Characteristics	**True-positive Reference Cohort (Neonatal sepsis)**		**True-negative Reference Cohort**	
	n	**%**	**n**	**%**
All Neonates	101	0.30	31508	99.7
Sex				
Female	48	47.5	15288	48.5
Male	53	52.5	16220	51.5
Gestational Age (weeks)				
Mean ± SD	26.1 (2.62)		38.4 (2.70)	
<28 weeks	76	75.3	280	0.90
28-36 weeks	25	24.7	3927	12.4
≥ 37 weeks	0	0	27295	86.6
Missing	0	0	6	0.02
Birth Weight (grams)				
Mean ± SD	941.4 (398.8)		3284.0 (673.5)	
Hospital Length of Stay (days)				
Median (interquartile range)	46 (28,84)		2 (1,3)	

4.1.2 Algorithm Development

The development of coding algorithms was based on two steps. Considering that 100% of cases in the true-positive cohort (**Table 4.1-3**) were preterm births, we decided to add preterm gestation to

our true positive and true negative reference cohort definitions, and all coding algorithms we developed and tested using only preterm births.

First, through the process of exploring the ICD-10-CA codes on the hospital abstracts, the results show in addition to the neonatal and adult sepsis codes, the codes B962, B9681, A499 and P393 were identified, which might assist with the identification of neonatal bacterial sepsis (**Table 4.1-4**). The B962 and B9681 organism codes were frequently recorded and relate to E. coli and Enterococcus, both of which are pathogens associated with bacterial sepsis. The code A499, indicating bacterial infection, was frequently coded in association with sepsis as well. The code P393 indicates neonatal urinary tract infection. Urinary tract infection can be considered bacteruria irrespective of the site of infection in the urinary tract, which is very common in neonatal sepsis. (72) Therefore, these additional codes identified were added to the pre-specified ICD-10 codes to develop first four coding algorithms. The performance of these four coding algorithms was tested against the true positive and negative reference cohorts (Table 4.1-7). Coding algorithm 4, which had the best performance, relative to algorithms 1 through 3, was selected for the next step.

Table 4.1-4. Additional ICD-10-CA codes frequently recorded and related to sepsis.

	ICD-10-CA	Nomenclature
1	B962	Escherichia coli [E. coli] as the cause of diseases classified to other chapters
2	B9681	Enterococcus as the cause of diseases classified to other chapters
3	A499	Bacterial infection, unspecified
4	P393	Neonatal urinary tract infection

Next, the impact of individual ICD-10-CA codes was explored based on the ratios of false positive to true positive cases by looking at incremental effects of codes added individually (**Table 4.1-5**).

Table 4.1-5. Incremental effects of ICD-10-CA codes on ratios of false positive cases to true positive cases.

<table>
<tr><th>IC-10-CA Codes</th><th>Ratio of FP/TP</th><th>IC-10-CA Codes</th><th>Ratio of FP/TP</th></tr>
<tr><td colspan="4">Neonatal sepsis codes</td></tr>
<tr><td>P360</td><td>3.5</td><td>P364</td><td>1.3</td></tr>
<tr><td>P361</td><td>2.0</td><td>P365</td><td>0.0</td></tr>
<tr><td>P362</td><td>1.3</td><td>P368</td><td>1.4</td></tr>
<tr><td>P363</td><td>1.1</td><td>P369</td><td>5.5</td></tr>
<tr><td colspan="4">Add adult sepsis codes</td></tr>
<tr><td>A410</td><td>0.0</td><td rowspan="7">A021, A227, A267, A327, A400, A401, A402, A403, A408, A409, A412, A413, A414, A4152, A4188, A427, A5486, B377, O85002, O85004, O85009</td><td rowspan="7">N/A</td></tr>
<tr><td>A411</td><td>0.5</td></tr>
<tr><td>A4150</td><td>N/A</td></tr>
<tr><td>A4151</td><td>0.0</td></tr>
<tr><td>A4158</td><td>0.0</td></tr>
<tr><td>A4180</td><td>0.0</td></tr>
<tr><td>A419</td><td>2.0</td></tr>
<tr><td colspan="2">Add B962, B9681, A499</td><td colspan="2">Add P393</td></tr>
<tr><td>A499</td><td>0.5</td><td>P393</td><td>1.5</td></tr>
<tr><td>B962</td><td>0.4</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>B9681</td><td>0.1</td></tr>
</table>

Based on the ratios of the false positive to true positives, six new coding algorithms (Coding algorithm 5 to 10) were produced. **Table 4.1-6** shows the ICD-10-CA codes of ten potential coding algorithms.

Table 4.1-6. Coding algorithms used to identify neonatal bacterial sepsis based on ICD-10-CA codes.

Coding Algorithm (ICD-10-CA)	
1	Neonatal Sepsis Codes*
2	Neonatal Sepsis Codes or Adult Sepsis Codes[&]
3	Neonatal Sepsis Codes or Adult Sepsis Codes or A499, B962, B9681
4	Neonatal Sepsis Codes or Adult Sepsis Codes or A499, B962, B9681 or P393
5	All codes from Coding Algorithm 4, excluding P369
6	All codes from Coding Algorithm 4, excluding P369, p360
7	All codes from Coding Algorithm 4, excluding P369, p361
8	All codes from Coding Algorithm 4, excluding P369, p360, p361
9	All codes from Coding Algorithm 4, excluding P369, p360, p361, A419
10	All codes from Coding Algorithm 4, excluding P369, p360, p361, A419, P393

Neonatal sepsis codes*: P360, P361, P362, P363, P364, P365, P368, P369;

Adult sepsis codes[&]: A021, A227, A267, A327, A400, A401, A402, A403, A408, A409, A410, A411, A412, A413, A414, A4150, A4151, A4152, A4158, A4180, A4188, A419, A427, A5486, B377, O85002, O85004, O85009

4.1.3 Algorithm Evaluation

We assessed the performance characteristics of all potential coding algorithms against TOH reference cohorts among preterm births. The characteristics (sensitivity, specificity, PPV, NPV, positive likelihood ratio and c-statistic) of all coding algorithms for identifying neonatal bacterial sepsis are presented in **Table 4.1-6**.

From Coding Algorithm 1 to 4, ICD-10-CA codes were progressively added from neonatal sepsis codes to more sepsis-related ICD-10-CA codes and we inspected of the number of false positive and true positive contributed by each. The sensitivity was increased from 44.6% (95% CI, 34.9% - 54.3%) in Algorithm 1 to 78.2% (95% CI, 70.2% - 86.3%) in Algorithm 4. The PPV was increased from 36.6% (95% CI, 28.1% - 45.1%) in Algorithm 1 to 45.1% (95% CI, 37.4% - 52.7%) in Algorithm 3, and then decreased to 44.6% (95% CI, 37.3% - 52.0%) in Algorithm 4. The specificity for these four algorithms ranged from 98.2% to 97.7%. Across Coding Algorithm 5 through 10, we progressively removed specific ICD-10-CA codes to evaluate the impact on performance. Among these six algorithms, the specificity, PPV and NPV values changed very little. Although Algorithm 10 had a slightly higher PPV of 53.3%, it came at the expense of an 11 percentage point drop in the sensitivity. Weighing the performance of these six algorithms against one another, Algorithm 5 was selected as being the one that resulted in the highest sensitivity of 75.3% (95% CI, 66.8% - 83.7%), while maintaining satisfactory specificity (98.2%; 95% CI, 97.8% - 98.6%) and a high NPV (99.4%; 95% CI, 99.2 - 99.6). The PPV was low (50.0%; 95% CI, 42.1% - 58.0%), indicating that only half of preterm infants who had a sepsis-related ICD-10 on their hospital abstract actually had sepsis according to their medical chart information. Although the PPV was low at 50%, none of the algorithms performed well on this metric, so Coding Algorithm 5 was the preferred algorithm for the purpose of demonstrating the application of the algorithm

developed in Objective 1, selected from a group of poorly performing algorithms. Given the poor performance of the algorithm, we have cautiously interpreted the results.

Table 4.1-7. Characteristics of all potential coding algorithms.

Coding algorithm	TP (n)	FN (n)	FP (n)	TN (n)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Positive likelihood Ratio	c-statistic
1	45	56	78	4129	44.6 (34.9-54.3)	98.2 (97.7-98.6)	36.6 (28.1-45.1)	98.7 (98.3-99.0)	24.0 (16.1-31.9)	0.71
2	56	45	84	4123	55.5 (45.8-65.1)	98.0 (97.6-98.4)	40.0 (31.9-48.1)	98.9 (98.6-99.2)	27.8 (19.7-35.9)	0.77
3	73	28	89	4118	72.3 (63.6-81.0)	97.9 (97.5-98.3)	45.1 (37.4-52.7)	99.3 (99.1-99.6)	34.2 (25.5-42.8)	0.85
4	79	22	98	4109	78.2 (70.2-86.3)	97.7 (97.2-98.1)	44.6 (37.3-52.0)	99.5 (99.3-99.7)	33.6 (25.7-41.5)	0.88
5	76	25	76	4131	75.3 (66.8-83.7)	98.2 (97.8-98.6)	50.0 (42.1-58.0)	99.4 (99.2-99.6)	41.7 (30.6-52.7)	0.87
6	74	27	70	4137	73.3 (64.6-81.9)	98.3 (98.0-98.7)	51.4 (43.2-59.6)	99.4 (99.1-99.6)	44.0 (31.9-56.2)	0.86
7	75	26	74	4133	74.3 (65.7-82.8)	98.2 (97.8-98.6)	50.3 (42.3-58.4)	99.4 (99.1-99.6)	42.2 (30.9-53.6)	0.86
8	73	28	68	4139	72.3 (63.6-81.0)	98.4 (98.0-98.8)	51.8 (43.5-60.0)	99.3 (99.1-99.6)	44.7 (32.1-57.3)	0.85
9	71	30	66	4141	70.3 (61.4-79.2)	98.4 (98.1-98.8)	51.8 (43.5-60.2)	99.3 (99.0-99.5)	44.8 (31.9-57.7)	0.84
10	65	36	57	4150	64.4 (55.0-73.7)	98.7 (98.3-99.0)	53.3 (44.4-62.1)	99.1 (98.9-99.4)	47.5 (32.6-62.4)	0.82

4.1.4 Sensitivity Analyses of Potential Coding Algorithms

In our definition (as described in Section 3.3.3), the true-negative reference cohort represented all neonates in the study population who did not meet the requirement for the true-positive cohort. Given our approach, the true-negative cohort would have contained some neonates meeting alternative definitions of bacteria sepsis: (1) neonates with microbiological confirmation who did not receive a prolonged (≥ 7 days) course of sepsis-related antibiotic; and (2) neonates who

received a prolonged (≥ 7 days) course of sepsis-related antibiotic but without microbiological confirmation. The potential contribution of these cases to the performance of the different coding algorithms was considered. **Table 4.1-8** summarizes the numbers of two categories in the false positives and true negatives.

Table 4.1-8. The components of the false positives and true negatives related to diagnostic criteria.

Coding Algorithm	False Positive (FP, n)			True Negative (TN, n)		
	Meet Criterion Ⅰ only or Ⅱ only [a]	Meet neither Criterion Ⅰ nor Criterion Ⅱ [b]	Sum [c]	Meet Criterion Ⅰ only or Ⅱ only [a]	Meet neither Criterion Ⅰ nor Criterion Ⅱ [b]	Sum [d]
1	54	24	78	155	3974	4129
2	60	24	84	149	3974	4123
3	63	26	89	146	3972	4118
4	70	28	98	139	3970	4109
5	60	16	76	149	3982	4131
6	58	12	70	151	3986	4137
7	60	14	74	149	3984	4133
8	58	10	68	151	3988	4139
9	56	10	66	153	3988	4141
10	49	8	57	160	3990	4150

a: Refers to neonates who met Criterion I only or Criterion II only (but not both).
Criterion Ⅰ refers to microbiological confirmation.
Criterion Ⅱ refers to a prolonged (≥ 7 days) course of sepsis-related antibiotic.
b: Refers to neonates who met neither Criterion I nor Criterion II.
d: This is the total number of false positives shown in the **Table 4.1-7**.
e: This is the total number of true negatives shown in the **Table 4.1-7**.

To evaluate the extent to which performance of the coding algorithms were affected by the inclusion of neonates who might have met clinical criteria of sepsis (level 2 of Brighton Collaboration Case Definition, **Table 2.3-1**), we removed these cases meeting Criterion Ⅰ only or Ⅱ only from the true-negative cohort and performed the sensitivity analyses. More specifically,

we only kept category of neonates who met neither Criterion I nor Criterion II and excluded neonates who met either Criterion I or Criterion Ⅱ (but not both) in the true negative cohort to reevaluate the performance of all previous coding algorithms.

As shown in **Table 4.1-9**, PPV values ranged from 65.2% (95% CI, 54.0% - 76.5%; Coding Algorithm 1) to 89.0% (95% CI, 81.9% - 96.2%; Coding Algorithm 10), while sensitivity remained unchanged compared with our original analyses (**Table 4.1-7**). For Coding Algorithm 5, the PPV was improved to be 82.6% (95% CI, 74.9% - 90.4%) while maintaining a sensitivity of 75.3% (95% CI, 66.8% - 83.7%).

Table 4.1-9. Sensitivity analyses of coding algorithms after changing the definition of the true-negative cohort (excluding infants who met Criterion Ⅰ only and met Criterion Ⅱ only).

Coding algorithm	TP (n)	FN (n)	FP (n)	TN (n)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Positive likelihood Ratio	c-statistic
1	45	56	24	3974	44.6 (34.9-54.3)	99.4 (99.2-99.6)	65.2 (54.0-76.5)	98.6 (98.3-99.0)	74.2 (38.4-110.0)	0.72
2	56	45	24	3974	55.5 (45.8-65.1)	99.4 (99.2-99.6)	70.0 (60.0-80.0)	98.9 (98.6-99.2)	92.4 (49.7-135.0)	0.77
3	73	28	26	3972	72.3 (63.6-81.0)	99.4 (99.1-99.6)	73.7 (65.1-82.4)	99.3 (99.1-99.6)	111.1 (63.8-158.5)	0.86
4	79	22	28	3970	78.2 (70.2-86.3)	99.3 (99.0-99.6)	73.8 (65.5-82.2)	99.5 (99.2-99.7)	111.7 (66.3-157.1)	0.89
5	76	25	16	3982	75.3 (66.8-83.7)	99.6 (99.4-99.8)	82.6 (74.9-90.4)	99.4 (99.1-99.6)	188.0 (87.9-288.1)	0.87
6	74	27	12	3986	73.3 (64.6-81.9)	99.7 (99.5-99.9)	86.1 (78.7-93.4)	99.3 (99.1-99.6)	244.1 (94.6-393.6)	0.87
7	75	26	14	3984	74.3 (65.7-82.8)	99.7 (99.5-99.8)	84.3 (76.7-91.8)	99.4 (99.1-99.6)	212.1 (91.6-332.5)	0.87
8	73	28	10	3988	72.3 (63.6-81.0)	99.8 (99.6-99.9)	88.0 (81.0-95.0)	99.3 (99.1-99.6)	289.0 (95.6-482.3)	0.86
9	71	30	10	3988	70.3 (61.4-79.2)	99.8 (99.6-99.9)	87.7 (80.5-94.8)	99.3 (99.0-99.5)	281.1 (92.6-469.5)	0.85
10	65	36	8	3990	64.4 (55.0-73.7)	99.8 (99.7-99.9)	89.0 (81.9-96.2)	99.1 (98.8-99.4)	321.6 (80.2-563.0)	0.82

4.2 Objective 2: Application of the Algorithm in the Ontario-wide population using ICES data to Estimate Incidence Rates and Trends in Neonatal Bacterial Sepsis

4.2.1 Study Population and Incidence Rates of Neonatal Bacterial Sepsis

The study population in Objective 2 included 1,858,169 infants born in Ontario with date of birth between April 1, 2003 to February 28, 2017. Of those, 150,312 (8.09%) were born at preterm gestation (<37 weeks). Since in Objective 1, all cases in the true-positive reference cohort were born at preterm gestation and we developed the algorithms among preterm births only, the study population in Objective 2 focused on neonates born at preterm gestation. The preferred coding algorithm identified in Objective 1 was applied in Objective 2 (i.e., Coding Algorithm 5).

Table 4.2-1 shows the characteristics of the study population in Objective 2 by sex, gestational age (weeks) and birth weight (grams). According to the preferred coding algorithm selected for this application, there were 6126 (4.10%) bacterial sepsis cases and 144186 (95.9%) non-sepsis cases among a total of 150,312 preterm infants, which produced an overall estimated incidence rate of 40.8 (95% CI, 39.8-41.8) per 1000 preterm live births in neonates born at preterm gestation in Ontario during the study period. To account for the poor performance of the case-finding algorithm, we computed an adjusted incidence rate based on the data inputs from Objective 1 (sensitivity: 75.3%; specificity: 98.2%) and Objective 2 (estimated incidence rate: 40.8 per 1000 preterm live births), which was found to be 31.0 per 1000 preterm live births.

Table 4.2-1. Characteristics of the study population in Objective 2.

Characteristics	Sepsis		Non-Sepsis	
	n	%	n	%
All Neonates	6126	4.10	144186	95.9
Sex				
Female	2541	41.5	66320	46.0
Male	3585	58.5	77866	54.0
Gestational Age (weeks)				
Mean (SD)	29.7 (3.88)		34.0 (3.02)	
<28 weeks	2164	35.3	7532	5.2
28-31 weeks	1793	29.3	11524	8.0
32-36 weeks	2169	35.4	125130	86.8
Birth Weight (grams)				
Mean (SD)	1480.0 (738.0)		2346.7 (699.5)	

Of the neonates identified by the algorithm as having bacterial sepsis, 3585 (58.5%) were male infants. The average gestational age was 29.7 weeks for bacterial sepsis cases and 34.0 weeks for non-sepsis cases. More specifically, 35.3% of sepsis cases were in the extremely preterm category (<28 weeks), and 29.3%, 35.4% were in the very preterm (28-31 weeks) and moderate or late preterm (32-36 weeks) categories respectively. The majority of non-sepsis cases (86.8%) were in the moderate or late preterm (32-36 weeks) category. The average birth weight of bacterial sepsis cases was 1480.0 grams while that of non-sepsis cases was 2346.7 grams.

4.2.2 Incidence Rates and Trends of Neonatal Bacterial Sepsis by Year of Birth

Incidence rates of neonatal bacterial sepsis were calculated by year of birth among the study population in Ontario during the study period. The estimated annual incidence rate decreased from 50.2 (95% CI, 45.4-55.4) per 1000 preterm live births in 2003 to 27.5 (95% CI, 20.4-36.9) per

1000 preterm live births in 2017. The difference of incidence rates between 2013 and 2017 was 22.7 (95% CI, 20.2-25.2) per 1000 preterm live births.

Table 4.2-2. Annual incidence rates of neonatal bacterial sepsis.

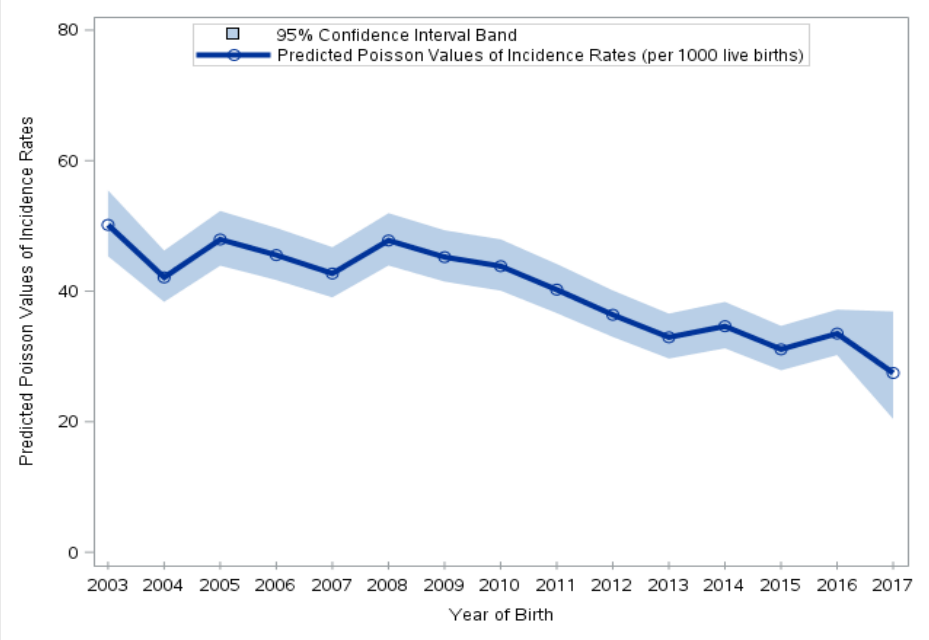
Year of Birth	Incidence Rate (per 1000 preterm live births, 95% CI)
2003	50.2 (45.4-55.4)
2004	42.1 (38.4-46.2)
2005	47.9 (43.9-52.3)
2006	45.5 (41.7-49.7)
2007	42.7 (39.1-46.7)
2008	47.8 (43.9-52.0)
2009	45.2 (41.5-49.3)
2010	43.8 (40.1-47.9)
2011	40.2 (36.6-44.2)
2012	36.4 (33.0-40.1)
2013	32.9 (29.7-36.6)
2014	34.6 (31.2-38.4)
2015	31.1 (27.9-34.7)
2016	33.5 (30.2-37.2)
2017	27.5 (20.4-36.9)

The Poisson regression model was applied to estimate temporal trends of neonatal bacterial sepsis. We assessed the model assumptions by exploring the mean-variance relationship. The Value/DF of the scaled Pearson Chi-Square was 1.6862, which, providing evidence that the variance (model estimate=79.84) was not equal to the mean (model estimate=30.69). Therefore, there was evidence to suggest that the mean-variance assumption of the Poisson model was not met in our study. To address this, one of the recommended remedies for overdispersion was used, in which the variance

of the model was adjusted by dividing it by 1.6862, also known as the scale parameter, to correct the error term for all tests and confidence intervals.

Figure 4.2-1 presents the trends in estimated annual incidence rates (per 1000 preterm live births) of neonatal bacterial sepsis by year of birth by showing predicted Poisson values of incidence rates with 95% confidence interval band. According to the line graph, there was a downward trend in incidence rates of neonatal bacterial sepsis in neonates born at preterm gestational age between 2003 and 2017 in Ontario. The trend by year of birth was statistically significant with P-value <0.0001.

Figure 4.2-1. Trends in incidence rates of neonatal bacterial sepsis in Ontario based on the Poisson model.



4.2.3 Incidence Rates and Trends of Neonatal Bacterial Sepsis Based on Infant Sex

Incidence rates of neonatal bacterial sepsis were calculated by infant sex among the study population in Ontario in the 15 years spanning from 2003 through 2017 (**Table 4.2-3**). During the whole study period, the incidence rate for female infants was 36.9 (95% CI, 35.5-38.3) per 1000 preterm live births, while that of male infants was 44.0 (95% CI, 35.5-38.3) per 1000 preterm live births.

Table 4.2-3. Sepsis cases and incidence rates by infant sex during 2003 and 2017.

Infant Sex	Number of incident cases	Total	Incidence rate per 1000 preterm live births (95% CI)
Female	2541	68861	36.9 (35.5-38.3)
Male	3585	81451	44.0 (42.6-45.4)

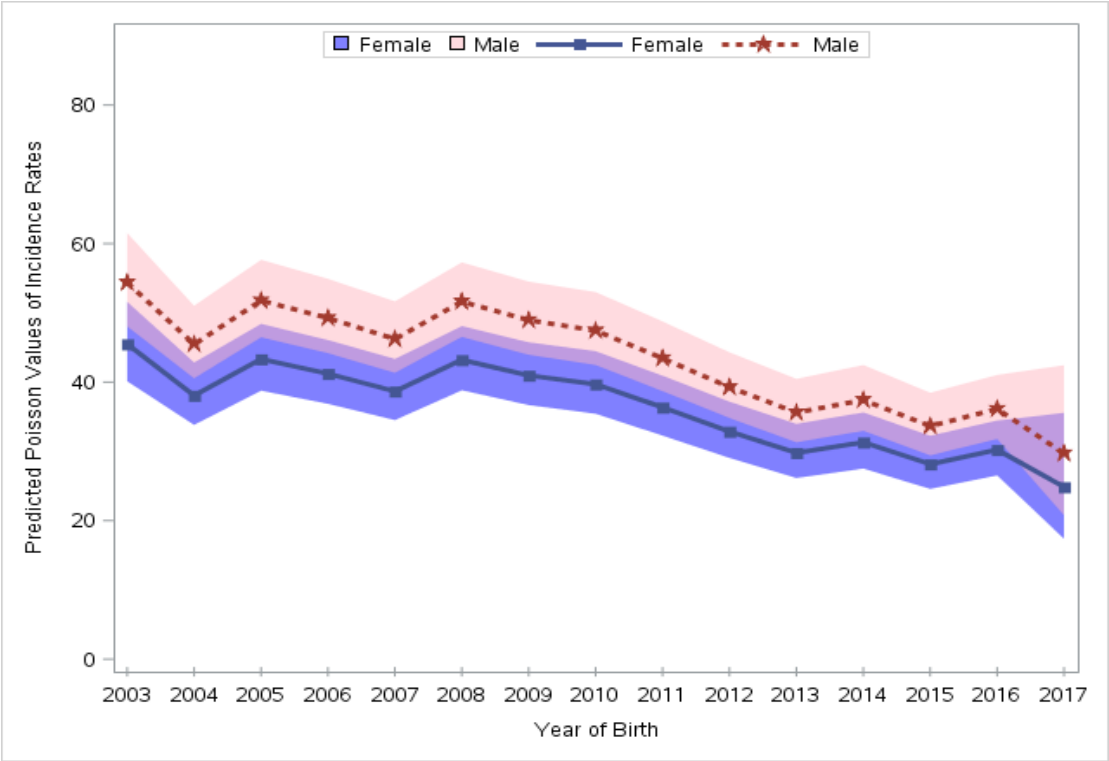
Incidence rates of neonatal bacterial sepsis were calculated by sex in each year among the study population. As shown in **Table 4.2-4**, the incidence rate of 45.5 (95% CI, 40.1-51.6) per 1000 preterm live births in 2003 for female infants declined to 24.9 (95% CI, 17.4-35.6) per 1000 preterm live births in 2017. The difference of incidence rates between 2003 and 2017 for female infants was 20.6 (95% CI, 17.0-24.2) per 1000 preterm live births. Among male infants, the incidence rate of 54.4 (95% CI, 48.0-61.5) per 1000 preterm live births in 2003 decreased to 29.7 (95% CI, 20.8-42.4) per 1000 preterm live births in 2017. The difference of incidence rates between 2003 and 2017 for male infants was 24.7 (95% CI, 21.2-28.2) per 1000 preterm live births.

Table 4.2-4. Annual incidence rates of neonatal bacterial sepsis by infant sex.

Year of Birth	Incidence Rate (per 1000 preterm live births, 95% CI)	
	Female	**Male**
2003	45.5 (40.1-51.6)	54.4 (48.0-61.5)
2004	38.1 (33.8-42.8)	45.5 (40.5-51.0)
2005	43.3 (38.8-48.4)	51.8 (46.5-57.7)
2006	41.2 (36.9-46.1)	49.2 (44.2-54.9)
2007	38.7 (34.5-43.3)	46.2 (41.4-51.6)
2008	43.2 (38.8-48.1)	51.6 (46.5-57.3)
2009	41.0 (36.7-45.7)	48.9 (43.9-54.5)
2010	39.7 (35.4-44.5)	47.4 (42.5-53.0)
2011	36.4 (32.3-40.9)	43.4 (38.7-48.8)
2012	32.9 (29.1-37.2)	39.3 (34.8-44.3)
2013	29.8 (26.1-34.0)	35.6 (31.3-40.5)
2014	31.3 (27.5-35.6)	37.4 (33.0-42.5)
2015	28.2 (24.6-32.3)	33.6 (29.4-38.5)
2016	30.2 (26.5-34.4)	36.1 (31.8-41.0)
2017	24.9 (17.4-35.6)	29.7 (20.8-42.4)

The Poisson regression model was used to estimate temporal trends of neonatal bacterial sepsis by infant sex in each year (**Figure 4.2-2**). Between 2003 and 2017 there were steady decreases in incidence rates both for both male and female infants. The trends for each were statistically significant with P-value < 0.0001. The incidence rates in the male infants were higher than in the female infants during the whole period. Based on the estimate of the fitted Poisson regression model, the relative incidence was calculated to be 1.22 (95% CI, 1.14-1.31), indicating that, on average, male infants had a 22% higher incidence of neonatal bacterial sepsis than female infants across the study period.

Figure 4.2-2. Trends in incidence rates of neonatal bacterial sepsis by infant sex based on the Poisson model.



4.2.4 Incidence Rates and Trends of Neonatal Bacterial Sepsis Based on Gestational Age

We evaluated incidence rates of neonatal bacterial sepsis according to gestational age at birth among the study population in Ontario in the 15 years spanning from 2003 through 2017. **Table 4.2-5** shows the number of sepsis cases and incidence rates within three categories of preterm birth. The extremely preterm group (<28 weeks) had the highest incidence of neonatal bacterial sepsis, 233.2 (95% CI, 214.9-231.6) per 1000 preterm live births, which was approximately 13.7 times that in the moderate to late preterm group (32-36 weeks) with 17.0 (95% CI, 16.3-17.7) per 1000 preterm live births. The incidence rate in the very preterm group (28-31 weeks) was 134.6 (95% CI, 128.8-140.4) per 1000 preterm live births.

Table 4.2-5. Sepsis cases and incidence rates by categories of preterm birth between 2003 and 2017.

Preterm Group	Number of incident cases	Total	Incidence of neonatal sepsis per 1000 preterm live births (95% CI)
Extremely preterm (<28 weeks)	2164	9696	233.2 (214.9-231.6)
Very Preterm (28-31 weeks)	1793	13317	134.6 (128.8-140.4)
Moderate to late preterm (32-36 weeks)	2169	127299	17.0 (16.3-17.7)

Annual incidence rates of neonatal bacterial sepsis were calculated by categories of preterm birth among the study population. As shown in **Table 4.2-6**, the incidence rate of 258.1 (95% CI, 223.6-297.8) per 1000 preterm live births in 2003 for the extremely preterm group (<28 weeks) was declined to 140.5 (95% CI, 93.9-210.1) per 1000 preterm live births in 2017. The difference of incidence rates between 2003 and 2017 was 117.6 (95% CI, 42.5-192.7) per 1000 preterm live births. For the very preterm category (28-31 weeks), the incidence rate of 155.0 (95% CI, 134.1-179.2) per preterm 1000 live births in 2003 was dropped to 84.4 (95% CI, 56.3-126.4) per 1000 preterm live births in 2017. The difference of incidence rates in 15 years was 70.6 (95% CI, 12.9-128.3) per 1000 preterm live births. In the moderate-to-late preterm group (32-36 weeks), the incidence rate of 19.7 (95% CI, 17.1-22.8) per 1000 preterm live births in 2003 was dropped to 10.7 (95% CI, 7.2-16.1) per 1000 preterm live births in 2017. The difference of incidence rates was 9.0 (95% CI, 7.03-10.9) per 1000 preterm live births.

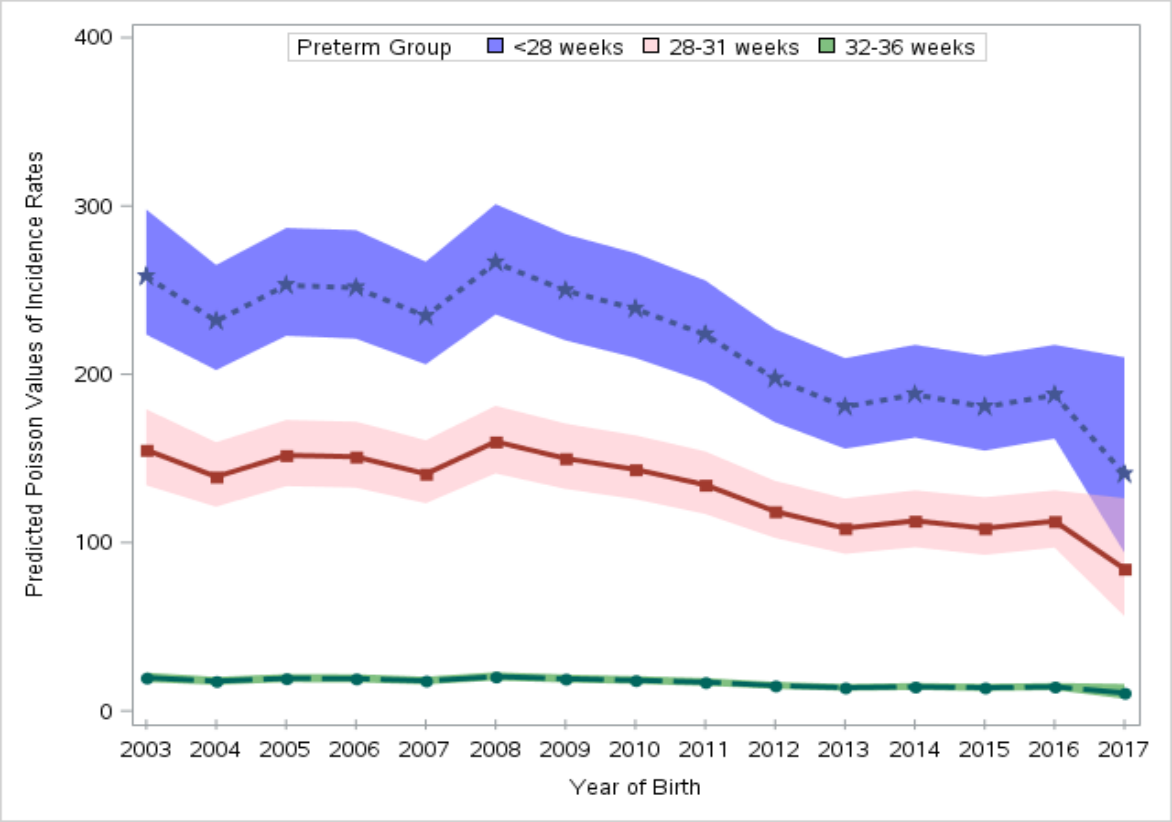
Table 4.2-6. Annual incidence rates of neonatal bacterial sepsis by categories of preterm birth.

Year of Birth	Incidence Rate (per 1000 preterm live births, 95% CI)		
	Extremely preterm (<28 weeks)	Very Preterm (28-31 weeks)	Moderate to late preterm (32-36 weeks)
2003	258.1 (223.6-297.8)	155.0 (134.1-179.2)	19.7 (17.1-22.8)
2004	231.6 (202.5-264.9)	139.1 (121.3-159.6)	17.7 (15.5-20.2)
2005	252.9 (222.8-287.0)	151.9 (133.5-172.8)	19.3 (17.0-21.9)
2006	251.4 (221.2-285.6)	151.0 (132.7-171.9)	19.2 (16.9-21.8)
2007	234.4 (205.9-266.9)	140.8 (123.3-160.8)	17.9 (15.7-20.4)
2008	266.3 (235.5-301.0)	160.0 (141.2-181.2)	20.3 (18.0-23.0)
2009	249.7 (220.1-283.3)	150.0 (131.8-170.6)	19.1 (16.8-21.6)
2010	238.8 (209.8-271.9)	143.5 (125.7-163.7)	18.2 (16.0-20.8)
2011	223.5 (195.4-255.8)	134.3 (117.0-154.1)	17.1 (14.9-19.5)
2012	197.2 (171.5-226.9)	118.5 (102.7-136.6)	15.1 (13.1-17.3)
2013	180.7 (155.7-209.6)	108.5 (93.3-126.2)	13.8 (11.9-16.0)
2014	187.9 (162.4-217.5)	112.9 (97.3-131.0)	14.4 (12.4-16.6)
2015	180.6 (154.6-210.9)	108.5 (92.7-126.9)	13.8 (11.8-16.1)
2016	187.6 (161.8-217.5)	112.7 (97.0-131.0)	14.3 (12.4-16.6)
2017	140.5 (93.9-210.1)	84.4 (56.3-126.4)	10.7 (7.2-16.1)

Based on the Poisson regression model, temporal trends of neonatal bacterial sepsis were estimated by categories of gestational age in each year (**Figure 4.2-3**). We can observe three trend lines with 95% confidence interval bands for three preterm categories. There was a decrease in incidence rates in the 15 years spanning from 2003 through 2017 for the three groups. The trends over time were statistically significant with P-value < 0.0001 for each of the preterm groups. The relative incidence of the extremely preterm (<28 weeks) versus that of moderate to late preterm (32-36 weeks) group was calculated to be 13.1 (95% CI, 12.1-14.2), indicating that, on average, infants

born in gestation age of <28 weeks had a 12.1 times higher incidence of neonatal bacterial sepsis than infants born in gestation age of 32-36 weeks across the study period. Additionally, the relative incidence of the very preterm (28-31 weeks) group versus that of moderate to late preterm (32-36 weeks) group was calculated to be 7.87 (95% CI, 7.53-8.56), indicating that, on average, infants born in gestation age of 28-31 weeks had a 6.87 times higher incidence of neonatal bacterial sepsis than infants born in gestation age of 32-36 weeks. Therefore, the annual decrease was more pronounced for the extremely preterm (<28 weeks) and very preterm (28-31 weeks) categories, which was consistent with previous calculation of difference in incidence rates in 15 years.

Figure 4.2-3. Trends in incidence rates of neonatal bacterial sepsis by categories of preterm birth based on the Poisson model.



Incidence rates of neonatal bacterial sepsis were also calculated based on gestational age (weeks). **Table 4.2-7** shows there is an increase of incidence rates from gestational age of 24 weeks (342.1, 95% CI, 311.6-375.7 per 1000 preterm live births) to 25 weeks (356.8, 95% CI, 327.5-388.8 per 1000 preterm live births). From 25 weeks onwards, incidence rates were steadily declined, reaching 8.44 (95% CI, 7.74-9.21) per 1000 preterm live births at gestational age of 36 weeks.

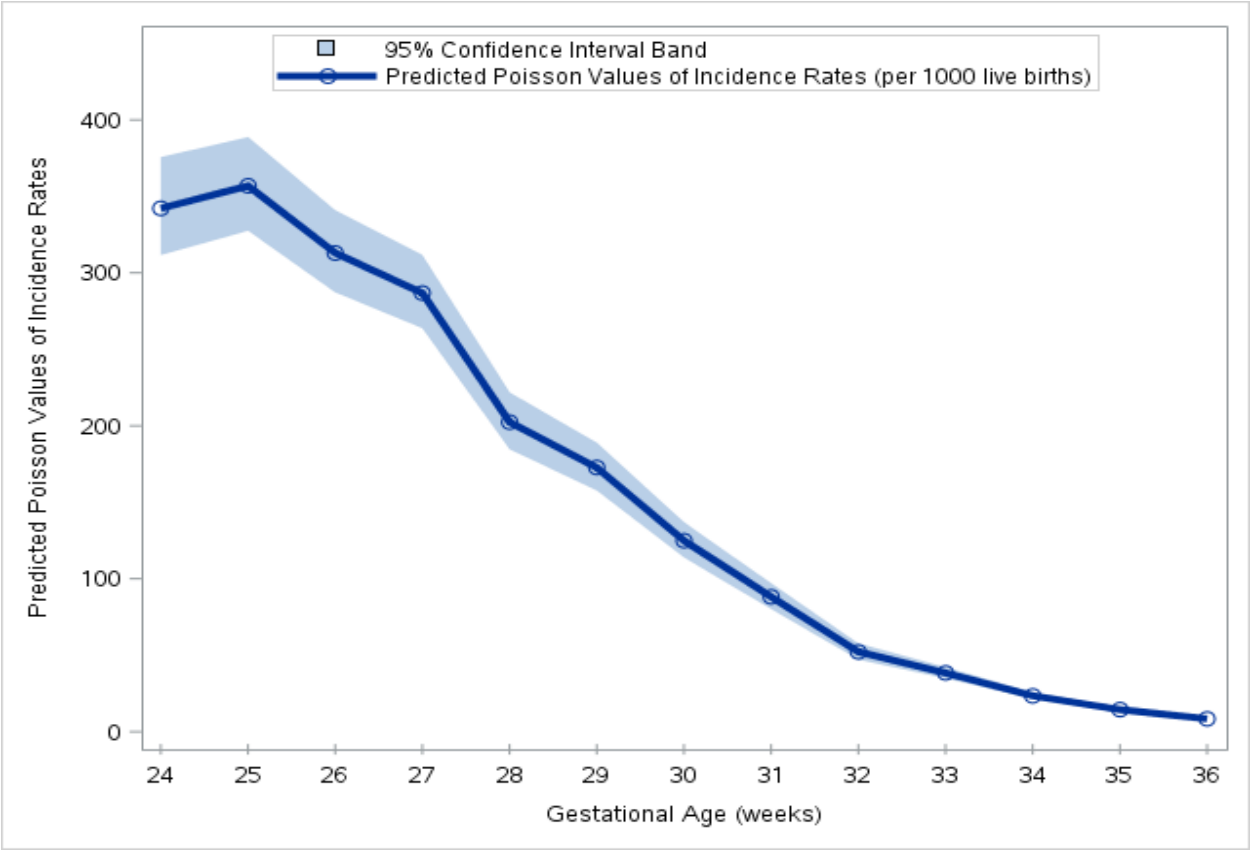
Table 4.2-7. Incidence rates of neonatal bacterial sepsis by gestational age (weeks).

Gestational Age (weeks)	Incidence Rate (per 1000 preterm live births, 95% CI)
24	342.1 (311.6-375.7)
25	356.8 (327.5-388.8)
26	313.0 (287.3-341.0)
27	286.8 (263.8-311.8)
28	202.4 (184.6-221.9)
29	172.7 (157.7-189.0)
30	124.9 (113.9-137.0)
31	88.2 (80.2-97.1)
32	52.2 (47.2-57.8)
33	38.4 (34.9-42.3)
34	23.5 (21.4-25.8)
35	14.4 (13.1-15.8)
36	8.44 (7.74-9.21)

- The infants below 24 weeks' gestation had few numbers of cases and were not included in trend analysis.

Figure 4.2-4, based on estimates from a Poisson regression model, shows temporal trends of neonatal bacterial sepsis from gestational age of 24 weeks to 36 weeks. It can be seen that there is a decreasing trend of incidence rates from a gestational age of 25 weeks. From the Poisson model, the trend of incidence rates by gestational age was statistically significant with P-value < 0.0001.

Figure 4.2-4. Trends in incidence rates of neonatal bacterial sepsis by gestational age based on the Poisson model.



CHAPTER 5: DISCUSSION

5.1 Objective 1: Develop and Validate Coding Algorithms

This study evaluated the feasibility of an ICD-10-CA based case-coding algorithm to identify neonatal bacterial sepsis in health administrative databases. The true-positive and true-negative reference cohorts were identified in the study population from electronic medical chart data, including patient-level information on microbiology test results and pharmacological treatment. We identified the ICD-10-CA codes that optimized the performance of the coding algorithms among infants born at preterm gestation.

In our TOH reference cohort, we identified 101 true positive sepsis cases from 2012 to 2017. All babies in the true-positive group in the reference cohort were preterm. It was likely related to the fact that we were only able to capture sepsis diagnoses at TOH and would have missed any neonates born at TOH who were initially discharged home with no diagnosis of sepsis recorded, but developed sepsis later during the neonatal period. These infants would have presented for care at CHEO, and therefore, would have been considered as part of the true negative reference cohort in our study since the CHEO chart data were not available to our study. Infants discharged home in the first 28 days of life are more likely to be term infants. In addition, currently in the city of Ottawa, there are 3 hospitals that provide Level 3 (TOH - General Campus and CHEO) or Level 2c (TOH - Civic Campus) neonatal care for preterm and/or unwell infants. (73) The high-risk pregnancies delivered in Ottawa are mainly at TOH - General Campus and Civic Campus, which is an urban tertiary academic teaching hospital (comprised of three campuses) that provides obstetrical care, including maternal Level 3 care. Due to the fact that TOH provides care to high-risk pregnant women, the hospital has a high preterm birth incidence (13.6%), relative to the general population of Ontario (7.9%). (74) Furthermore, in the reference cohort, the extremely

preterm (<28 weeks) and very preterm (28-31 weeks) categories accounted for 2.7% of the TOH birth population, while the moderate to late preterm (32-36 weeks) category was 10.9%. In the general population of Ontario, the corresponding rates are approximately 1.1% and 6.8%, respectively. (74) Because increasing prematurity is strongly associated with increased incidence of bacterial sepsis, and preterm infants were over-represented in our study population compared to the general population, the estimated incidence of neonatal sepsis in Objective 1 was likely increased.

The performance characteristics of all potential coding algorithms, based on various ICD code combinations, were evaluated by calculating measures of diagnostic accuracy. All of the algorithms derived in the study suffered from imperfect sensitivity and low PPVs. Because PPV is the probability of a subject having the disease given a positive test result, low PPVs pose a risk of overestimation because of high false positive rate. For example, coding algorithm 5 had a PPV of 50.0%, suggesting that about 50% of the cases captured by the coding algorithm would be false positive cases. Our findings indicate that health administrative data is not an ideal source for estimating incidence of neonatal bacterial sepsis, at least not with currently available OHDW data without free text clinical notes.

In the sensitivity analyses, the range of PPV values was greatly improved from 36.6% - 53.3% to 65.2% - 89.0%. The reason for this was that the neonates who received a prolonged (≥ 7 days) course of sepsis-related antibiotic but without microbiological confirmation accounted for a large number of false positives for those coding algorithms. Although those neonates didn't meet level 1 of Brighton Collaboration Case Definition (microbiological confirmation), they were at high risk to meet clinical criteria (level 2 of Brighton Collaboration Case Definition) of sepsis since they received a prolonged (≥ 7 days) course of sepsis-related antibiotic. Once those cases were excluded

in the sensitivity analyses, the PPV values were greatly improved. Following up from our study, a more extensive chart review is underway in a subset of the false positive and false negative records to review free text clinical notes and other information not available in the electronic chart review (*e.g.*, fever, clinical judgement, other observations from physical examination not easily captured using current ICD-10 coding practices). It might be beneficial to optimize the coding algorithms to capture more true sepsis cases and reduce the false positives.

Generally speaking, an algorithm will be less likely to misclassify a non-case as positive case if the case identifying algorithm is more stringent. However, as a tradeoff, the sensitivity of the coding algorithm would decrease because it becomes more difficult to meet the case definition. In this study, the preferred algorithm of those we tested was only moderately sensitive and had imperfect specificity. Lower sensitivity would lead to underestimation if specificity were perfect, and lower specificity would lead to overestimation if sensitivity were perfect. In reality, there is always a mixture of both types of misclassification, which may offset each other, and it is difficult to know what the true impact is. Regardless, the sensitivity of the preferred algorithm in our study (75.3%, 95% CI, 66.8%-83.7%), was similar to the optimized ICD-10-CA coding algorithm for adult sepsis (71.9%, 95% CI 68.1%-75.4%, aged 18 years and older) in a previous study conducted in the Calgary region of Alberta, Canada. (9) Our results were also consistent with those of other hospital-acquired infections reported in a systematic review, which reported moderate sensitivity (76.0%, 95% CI, 56.2%-88.7%) and high specificity (99.9%, 95% CI, 99.6%-100%) using either ICD-9 or ICD-10 in administrative data from literature during the period of 2000-2012. (75)

For the purpose of surveillance, under-ascertaining or over-ascertaining the true incidence of neonatal bacterial sepsis results in inaccurate assessments of prevalence which would contribute to inadequate allocation of resources for monitoring and treatment. There are some probable

reasons for the inaccurate assessments of neonatal bacterial sepsis. First of all, it is strongly associated with the accuracy of the code. The accuracy of the codes assigned to the electronic record that is eventually available for secondary use in research may be influenced by several factors, including the completeness and accuracy of the information that is documented in the chart, as well as the interpretation of the coder abstracting the information. When information in the chart is incomplete (e.g., if the lab test results are not available at the time the chart information is abstracted by the coders) or illegible, the assignment of codes may be also be incomplete or inaccurate. (48) In addition, it may be related to the natural limitations of administrative data, which are developed for managing the healthcare system and generally focus on physician documentation rather than clinical information on the patient's status. Healthcare coders may not identify a diagnosis of sepsis based on physician documentation alone, since physicians may not explicitly state the term "sepsis" in the chart, instead identifying only the infection present. Further, sepsis may go uncoded or undocumented if there are other more pressing clinical diagnoses present during the health care encounter. For example, if 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. (76, 77) Sepsis may be underreported having been coded using organ-specific infection instead of being classified as sepsis. (28)

In summary, the coding algorithm developed in this study could not accurately identify neonates with bacterial sepsis from within health administrative database using the data available to us now.

5.2 Objective 2: Application of the Algorithm in the Ontario-wide population using ICES data to Estimate Incidence Rates and Trends in Neonatal Bacterial Sepsis

As an exercise to illustrate the application of using coding algorithms for incidence rate estimation, we used coding algorithm 5 from Objective 1 to identify preterm infants with bacterial sepsis from within the population-based health administrative databases in ICES. The coding algorithm detected a decreased trend in the estimated annual incidence during the study time period (2003 to 2017) and significant variation of sepsis incidence rates was noted across sex and gestational age. However, we are unable to draw any firm conclusions regarding true temporal trends in bacterial sepsis incidence during the study period, because of the suboptimal performance of the coding algorithm available for this application.

Overall, the estimated incidence rate of bacterial sepsis for preterm infants in Ontario during the whole study period was 40.8 (95% CI, 39.8-41.8) per 1000 preterm live births. To account for the suboptimal performance of the coding algorithm developed in Objective 1, we adjusted the overall incidence rate based on the observed sensitivity of 75.3% and specificity of 99.6% of coding algorithm 5, and found the adjusted incidence rate to be 31.0 per 1000 preterm live births. This result implies that we likely overestimated the incidence by approximately 10 per 1000 preterm live births, assuming that the sensitivity and specificity of coding algorithm 5 were consistent in the applied setting during the study period.

Our estimation of annual sepsis incidence rates declined during the study period and the difference of incidence rates between 2013 and 2017 was 22.7 (95% CI, 20.2-25.2) per 1000 preterm live births. Although we cannot rely on the accuracy of the incidence estimates due to the poor performance of the algorithm, the declining trends are generally consistent with prior research. For

example, a Canadian national (excluding Quebec) epidemiologic report of sepsis showed a decline in hospitalization rates for neonates with diagnostic codes for sepsis from 2005 to 2009. (28) A population-based study in Taiwan reported sepsis incidence decreased from 598.0 to 336.4 cases per 100000 people from 2002 to 2012 in the infant population. (78) Further, Chanu et al reported hospital mortality rates at two academic hospitals in Boston, U.S. declined for sepsis patients with positive blood cultures, from 15.7% to 11.6% between 2003 and 2012. (79) Regarding the estimated global incidence of neonatal sepsis, although an overall decline in sepsis-related mortality has occurred over the past decade, there has been a dramatic difference in terms of incidence rates related to sepsis in the infant population, with incidence of neonatal sepsis around 40-times higher in middle-income countries than in high-income countries. (2, 80) To our knowledge, there were few representative epidemiological studies on incidence rates of sepsis in preterm infants.

Some measures were believed to have improved the outcome of neonatal sepsis. First, development of clear guidelines, concentrated use of specific educational initiatives, improved supportive care and early evidence-based bundle care have contributed to substantial reductions in morbidity and mortality of sepsis. (81, 82) An International guideline "Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock" was originally published in 2004 and updated in 2008 and 2012. (83-85) Second, studies have found that screening and initiation of antibiotic prophylaxis of pregnant women in the intra-partum period for those who screen positive for Group B Streptococcus (GBS) colonization has contributed to the reduction in early-onset neonatal sepsis infections. (86) Early-onset neonatal sepsis can be caused by vertical transmission of infectious organisms from the mother, before or during birth. Maternal GBS colonization in the current pregnancy and GBS bacteriuria are the risk factors most commonly associated with

neonatal bacterial sepsis. (87) Efforts have been made to screen for GBS colonization during late pregnancy (35-37 weeks) and provide intrapartum antimicrobial prophylaxis to colonized mothers in order to prevent vertical transmission to the newborn. For example, recommendations for intrapartum prophylaxis to prevent perinatal GBS disease were issued in 1997 by the American Academy of Pediatrics. (88) Revised guidelines for the prevention of early-onset GBS disease issued in 2002 recommended universal culture-based screening of all pregnant women at 35-37 weeks' gestation to optimize the identification of women who should receive intrapartum antibiotic prophylaxis. (89) In the United States, widespread acceptance of intrapartum antibiotic prophylaxis to reduce vertical transmission of GBS infections in high-risk women has resulted in a significant decline in rates of neonatal sepsis by GBS infection. (90) In 2013, the Society of Obstetricians and Gynaecologists of Canada published recommendations for regimens for antimicrobial prophylaxis among pregnant women in labour for the prevention of early-onset neonatal GBS disease. (91) All these efforts were beneficial to the treatment of maternal infections or colonization, which might prevent a significant proportion of early-onset neonatal sepsis. It was likely to be consistent with our observation in substantially decline of trend in incidence of neonatal bacterial sepsis in Ontario in the following 6 years from 2008. Also, it is still possible that other factors might have varied over time which could have affected the temporal trends that we estimated in Objective 2. For example, expertise of the coders, quality of coding, or the impact of physician documentation could also had changed over this time period.

We observed a decreasing trend in incidence rates for both male and female infants; however, incidence rates were higher for male infants during ever year of our study period. Overall, the relative incidence from Poisson regression model shows male infants had a 22% increased incidence of bacterial sepsis compared with female infants. Similar results were presented by

Watson and Carcillo, who observed a higher incidence of severe sepsis in infant boys compared to girls. (92) Moreover, sex-related difference in children with sepsis were reported in some other studies. (28, 93) Although a mechanism of hormonal and immunologic change related to sex differences has been postulated, the exact mechanism of the sex difference in sepsis is still to be determined. (94)

Variation in incidence rates of sepsis by categories of preterm birth and gestational age (weeks) in incidence rates were observed in each year over the entire study period. Overall, increasing prematurity was strongly associated with increased incidence of bacterial sepsis, with the rate of bacterial sepsis in the extremely preterm group (<28 weeks, 233.2 per 1000 preterm live births, 95% CI, 214.9-231.6) approximately 13.7 times that in the moderate to late preterm group (32-36 weeks, 17.0 per 1000 preterm live births, 95% CI, 16.3-17.7) in the study cohort. The infants below 24 weeks' gestation had few numbers of cases and were not included in trend analysis. From 25 to 36 weeks' gestation, the incidence rates of bacterial sepsis experience a drastic drop with the increase of gestational age. These findings are likely related to some neonatal factors, such as survival rate and prematurity.

Firstly, gestational age is strongly associated with outcome in terms of survival and survival rates among live-born infants below 24 weeks' gestation are relatively low. (95, 96) According to a population-based study of all infants born at 22 through 26 weeks' gestation in Norway in 2013-2014, the survival rate was 18% at 22 weeks, 29% at 23 weeks, 56% at 24 weeks, 84% at 25 weeks and 90% at 26 weeks. (97) Before 24 weeks' gestation, most deaths occurred in utero after admission to an obstetric unit or because of omission of resuscitation at birth. (98) They are likely to be died within 24 hours due to respiratory distress syndrome due to an immature respiratory system and/or congenital anomalies. (99) Consequently, they may don't survive long enough to

even get sepsis, or be coded with sepsis. For instance, at Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network Centers, infants who were born alive in 2003-2007 died at ≤ 12 hours, with most early deaths occurring at 22 and 23 weeks (85% and 43%, respectively). (29) In addition, other factors (e.g., death within a very short time after birth, termination of pregnancy) may be attributable to the lower rates of neonatal bacterial sepsis under 25 weeks' gestation.

Secondly, prematurity is one of the most important neonatal risk factors predisposing to infection that could result in sepsis. Our observation was consistent with those in other studies. (29, 91,100) For example, a nationwide Swedish prospective cohort study between 2004 and 2007 reported 66% of the 497 extremely preterm infants had at least one episode of sepsis. (101) In addition, a 36% of neonates born before 28 weeks suffered at least one episode of blood stream infection during their birth hospitalization in the US. (102) In our study, a 22.3% of neonates in extremely preterm group (<28 weeks) in Ontario between 2003 and 2017 had bacterial sepsis. There are many reasons why neonates born at preterm gestation face a great risk of acquiring sepsis. Compared with the immune system of the term infants, those of the preterm infants are less developed in several respects. (103) Immune dysfunction and an absence of transplacentally acquired maternal IgG antibodies in preterm infants may contribute to increased susceptibility to neonatal infections. (104) In addition, preterm infants have increased risk for hospital-acquired infections, since they often need prolonged intravenous access, endotracheal intubation or other invasive procedures that provide a portal of entry for pathogens or impair barrier and clearance mechanisms. (16)

5.3 Study Limitations

There are several limitations of our validation study in Objective 1. First, we defined the reference cohorts using the electronic health record which included microbiology test results and pharmacy data by data linkage at the patient level. It is a concern that misdiagnosis or incomplete documentation of free text clinical data on the electronic health record may have impacted our ability to evaluate level 2 of Brighton Collaboration Case Definition. For example, we may have missed capturing some true cases of bacterial sepsis and therefore increased the number of false negatives or false positives. Further, one of our diagnostic criteria was a course of antibiotics (≥ 7 days). If an infant died during administration of antibiotics and, therefore, did not complete a full course of antibiotics, they would not have reached the necessary criteria to be identified as a case in our study. As a next step from this study, a more extensive chart review is underway for a subset of false negative and false positive charts to check free text clinical notes which were not available in the electronic chart review at the time this study was conducted. Second, the study population for developing our coding algorithms was selected from births occurring at one hospital (TOH) in Ottawa, an urban tertiary academic teaching hospital that provides obstetrical care, and a level 3 neonatal intensive care unit (NICU) for preterm and/or unwell infants. Therefore, it is possible that the algorithm could perform differently in different clinical situations, such as in hospitals with lower level of obstetrical/neonatal care designation, which might lead to some misclassification if applying this TOH-based algorithm to other settings. Additionally, no term infants appeared in our true positive reference cohort. The distribution of preterm births at TOH, as a tertiary care centre, might be different from other hospitals in the province. If the incidence of sepsis was higher in TOH than in the general population, then the algorithm might perform worse (in terms of PPV) in the general population assuming sensitivity

and specificity remain constant. Furthermore, although we initially intended to include data from both TOH and CHEO in our validation study, in order to be able to identify all the true-positive neonatal sepsis cases, the complexities of accessing chart data from two organizations was beyond the financial and time feasibility of this Master's thesis project; therefore, we were only able to access data from TOH. Neonates who were discharged from TOH with no diagnosis of sepsis recorded, could have presented later within the neonatal period with sepsis at CHEO or another hospital, and this diagnosis would have been missed by our study. Additionally, we were only able to validate the coding algorithm in preterm infants, which restricted its application. Though rates of bacterial sepsis would be very low in term infants, since most infants are born at term, they still account for the majority of actual bacterial sepsis cases.

In Objective 2, with the suboptimal performances of the coding algorithm identified in Objective 1, the algorithm was not of sufficient accuracy to produce reliable population estimates of neonatal sepsis, and we likely overestimated the incidence. In addition, the derivation time frame of the TOH reference cohort used to develop the algorithm was 2012 to 2017, while the time frame for the sepsis incidence estimation was from 2003 to 2017. We were not able to test the performance of the algorithms against medical chart data in the decade prior to the derivation time period (2003 to 2011). There is no way to know for certain whether there were any temporal issues that could have affected the estimates. However, the coding system (ICD-10-CA) was consistent from 2003 onward. Nevertheless, the estimates of incidence rates and temporal trends were derived from administrative datasets, which relied on correct data coding. Accordingly, misclassification of sepsis could be possible and may vary by factors such as disease severity and over time. We did not assess the accuracy of the algorithm according to the position of the sepsis codes in the diagnostic code fields in the administrative data. It is possible that true sepsis cases

were rarely recorded as anything but a type M or 1 diagnosis. Therefore, requiring a type M or 1 diagnosis type could reduce false positives without significantly impacting sensitivity. Further, we were also unable to study the incidence rates in term infants, therefore, these results are only generalizable to preterm infants.

5.4 Conclusion

In conclusion, this study could serve as a caution for researchers who are interested in studying neonatal bacterial sepsis using health administrative data. The currently available health administrative data is not suitable for identifying neonatal bacterial sepsis. All potential coding algorithms suffered mainly from low PPVs. We observe the overall incidence rates of bacterial sepsis in the preterm infants in Ontario has declined over time and sex and gestational age show to be associated with the incidence rates of illness. Since the availability of electronic medical records is increasing, the availability of large administrative and clinical databases for programmatic monitoring and evaluation, as well as for research purposes is likely to expand. However, there is still much to learn about appropriate utilization of health administrative data for research, since it is a relatively new source of research data. For the next step, there is a need for a deeper clinical chart review including physician and nursing narrative notes for the cases that partially met the case definition, but couldn't be fully ascertained in the more limited electronic health chart, in which we did not have access to those bedside clinical notes. This deeper chart review of "probable" cases of sepsis from this validation study is in the early stages, and will hopefully result in a clearer picture of true positive and true negatives in the validation cohort. Future research priorities also include expanding bacterial sepsis to include viral and fungal sepsis, multicenter evaluation.

REFERENCES

1. World Health Organization Executive Board (EB140/12). Improving the prevention, diagnosis and clinical management of sepsis. http://appswhoint/gb/ebwha/pdf_files/EB140/B140_12-enpdf (accessed June 20,2017). 2017.

2. Fleischmann-Struzek C, Goldfarb DM, Schlattmann P, Schlapbach LJ, Reinhart K, Kissoon N. The global burden of paediatric and neonatal sepsis: a systematic review. The Lancet Respiratory Medicine. 2018;6(3):223-30.

3. Shah BA, Padbury JF. Neonatal sepsis: an old problem with new insights. Virulence. 2014;5(1):170-8.

4. Benchimol EI, Guttmann A, Griffiths AM, Rabeneck L, Mack DR, Brill H, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. Gut. 2009;58(11):1490-7.

5. Pisesky A, Benchimol EI, Wong CA, Hui C, Crowe M, Belair M-A, et al. Incidence of Hospitalization for Respiratory Syncytial Virus Infection amongst Children in Ontario, Canada: A Population-Based Study Using Validated Health Administrative Data. PloS one. 2016;11(3):e0150416.

6. Benchimol EI, Guttmann A, Mack DR, Nguyen GC, Marshall JK, Gregor JC, et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. Journal of clinical epidemiology. 2014;67(8):887-96.

7. Chan J, Mack DR, Manuel DG, Mojaverian N, de Nanassy J, Benchimol EI. Validation of an algorithm to identify children with biopsy-proven celiac disease from within health administrative data: An assessment of health services utilization patterns in Ontario, Canada. PLoS One. 2017;12(6):e0180338.

8. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. Journal of clinical epidemiology. 2011;64(8):821-9.

9. Jolley RJ, Quan H, Jetté N, Sawka KJ, Diep L, Goliath J, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. BMJ open. 2015;5(12):e009487.

10. Bouza C, Lopez-Cuadrado T, Amate-Blanco J. Use of explicit ICD9-CM codes to identify adult severe sepsis: impacts on epidemiological estimates. Critical Care. 2016;20(1):313.

11. Iwashyna TJ, Odden A, Rohde J, Bonham C, Kuhn L, Malani P. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. Med Care. 2014;52.

12. Fell DB, Hawken S, Wong CA, Wilson LA, Murphy MS, Chakraborty P, et al. Using newborn screening analytes to identify cases of neonatal sepsis. Scientific reports. 2017;7(1):18020.

13. Jolley RJ, Sawka KJ, Yergens DW, Quan H, Jette N, Doig CJ. Validity of administrative data in recording sepsis: a systematic review. Crit Care. 2015;19.

14. Thatte N, Kalter HD, Baqui A, Williams E, Darmstadt G. Ascertaining causes of neonatal deaths using verbal autopsy: current methods and challenges. Journal of Perinatology. 2009;29(3):187-94.

15. Singer M, Deutschman CS, Seymour C, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA. 2016;315(8):801-10.

16. Shane AL, Sánchez PJ, Stoll BJ. Neonatal sepsis. The Lancet. 2017:11.

17. Klinger G, Levy I, Sirota L, Boyko V, Reichman B, Lerner-Geva L, et al. Epidemiology and risk factors for early onset sepsis among very-low-birthweight infants. American journal of obstetrics and gynecology. 2009;201(1):38. e1-. e6.

18. Wynn JL, Wong HR, Shanley TP, Bizzarro MJ, Saiman L, Polin RA. Time for a neonatal–specific consensus definition for sepsis. Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies. 2014;15(6):523.

19. Blatt S, Schroth M. Neonatal Sepsis: Clinical Considerations. Journal of Child Science. 2017;1(01):e54-e9.

20. Van Den Hoogen A, Gerards LJ, Verboon-Maciolek MA, Fleer A, Krediet TG. Long-term trends in the epidemiology of neonatal sepsis and antibiotic susceptibility of causative agents. Neonatology. 2009;97(1):22-8.

21. Stoll BJ, Hansen N, Fanaroff AA, Wright LL, Carlo WA, Ehrenkranz RA, et al. Changes in pathogens causing early-onset sepsis in very-low-birth-weight infants. New England Journal of Medicine. 2002;347(4):240-7.

22. Camacho-Gonzalez A, Spearman PW, Stoll BJ. Neonatal infectious diseases: evaluation of neonatal sepsis. Pediatric Clinics of North America. 2013;60(2):367.

23. Mathur S, Li G, Folgori L, Sharland M, Heath PT. DeNIS collaboration: setting the future research agenda. Lancet Global Health. 2017;5(1).

24. Verma P, Berwal PK, Nagaraj N, Swami S, Jivaji P, Narayan S. Neonatal sepsis: epidemiology, clinical spectrum, recent antimicrobial agents and their antibiotic susceptibility pattern. International Journal of Contemporary Pediatrics. 2015;2(3):176-80.

25. Lawn JE, Cousens S, Zupan J, Team LNSS. 4 million neonatal deaths: when? Where? Why? The lancet. 2005;365(9462):891-900.

26. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The Lancet. 2012;379(9832):2151-61.

27. Group IC. Treatment of neonatal sepsis with intravenous immune globulin. New England Journal of Medicine. 2011;365(13):1201-11.

28. Thompson GC, Kissoon N. Sepsis in Canadian children: A national analysis using administrative data. Clinical epidemiology. 2014;6:461.

29. Stoll BJ, Hansen NI, Bell EF, Shankaran S, Laptook AR, Walsh MC, et al. Neonatal outcomes of extremely preterm infants from the NICHD Neonatal Research Network. Pediatrics. 2010:peds. 2009-959.

30. Cohen-Wolkowiez M, Moran C, Benjamin DK, Cotten CM, Clark RH, Benjamin Jr DK, et al. Early and late onset sepsis in late preterm infants. The Pediatric infectious disease journal. 2009;28(12):1052.

31. Haque KN, Khan MA, Kerry S, Stephenson J, Woods G. Pattern of culture-proven neonatal sepsis in a district general hospital in the United Kingdom. Infection Control & Hospital Epidemiology. 2004;25(9):759-64.

32. Martinot A, Leclerc F, Cremer R, Leteurtre S, Fourier C, Hue V. Sepsis in neonates and children: definitions, epidemiology, and outcome. Pediatric emergency care. 1997;13(4):277-81.

33. Stoll BJ, Hansen NI, Adams-Chapman I, Fanaroff AA, Hintz SR, Vohr B, et al. Neurodevelopmental and growth impairment among extremely low-birth-weight infants with neonatal infection. Jama. 2004;292(19):2357-65.

34. Wynn JL. Defining neonatal sepsis. Current opinion in pediatrics. 2016;28(2):135-40.

35. Committee ASCC. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit Care Med. 1992;20:864-74.

36. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 sccm/esicm/accp/ats/sis international sepsis definitions conference. Intensive care medicine. 2003;29(4):530-8.

37. Vergnano S, Buttery J, Cailes B, Chandrasekaran R, Chiappini E, Clark E, et al. Neonatal infections: Case definition and guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine. 2016;34(49):6038-46.

38. Verma P, Berwal PK, Nagaraj N, Swami S, Jivaji P, Narayan S. Neonatal sepsis: epidemiology, clinical spectrum, recent antimicrobial agents and their antibiotic susceptibility pattern. International Journal of Contemporary Pediatrics. 2017;2(3):176-80.

39. Benitz WE. Adjunct laboratory tests in the diagnosis of early-onset neonatal sepsis. Clinics in perinatology. 2010;37(2):421-38.

40. Srinivasan L, Harris MC. New technologies for the rapid diagnosis of neonatal sepsis. Current opinion in pediatrics. 2012;24(2):165-71.

41. Benitz WE, Wynn JL, Polin RA. Reappraisal of guidelines for management of neonates with suspected early-onset sepsis. The Journal of pediatrics. 2015;166(4):1070-4.

42. Hornik CP, Fort P, Clark RH, Watt K, Benjamin Jr DK, Smith PB, et al. Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units. Early human development. 2012;88:S69-S74.

43. Mtitimila EI, Cooke RW. Antibiotic regimens for suspected early neonatal sepsis. Cochrane Database of Systematic Reviews. 2004(4).

44. Hinds A, Lix LM, Smith M, Quan H, Sanmartin C. Quality of administrative health databases in Canada: A scoping review. Can J Public Health. 2016;107(1):56-61.

45. McChesney-Corbeil J, Barlow K, Quan H, Chen G, Wiebe S, Jetté N. Validation of a case definition for pediatric brain injury using administrative data. Canadian journal of neurological sciences. 2017;44(2):161-9.

46. World Health Organization. International Classification of Diseases (ICD). Available online at: http://wwwwhoint/classifications/icd/en/. Accessed May 3, 2014.

47. Jetté N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, et al. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. Medical care. 2010;48(12):1105-10.

48. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. Journal of clinical epidemiology. 2011;64(10):1054-9.

49. Carnahan RM, Herman RA, Moores KG. A systematic review of validated methods for identifying transfusion-related sepsis using administrative and claims data. pharmacoepidemiology and drug safety. 2012;21:222-9.

50. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Critical care medicine. 2001;29(7):1303-10.

51. Cevasco M, Borzecki AM, Chen Q, Zrelak PA, Shin M, Romano PS, et al. Positive predictive value of the AHRQ patient safety indicator “postoperative sepsis”: implications for practice and policy. Journal of the American College of Surgeons. 2011;212(6):954-61.

52. 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;15(5):513-6.

53. Madsen KM, Schønheyder HC, Kristensen B, Nielsen GL, Sørensen HT. Can hospital discharge diagnosis be used for surveillance of bacteremia? A data quality study of a Danish hospital discharge registry. Infection Control & Hospital Epidemiology. 1998;19(3):175-80.

54. 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.

55. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. Journal of clinical epidemiology. 2007;60(4):397-409.

56. Gaieski DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. Critical care medicine. 2013;41(5):1167-74.

57. Whittaker S-A, Mikkelsen ME, Gaieski DF, Koshy S, Kean C, Fuchs BD. Severe sepsis cohorts derived from claims-based strategies appear to be biased towards a more severely ill patient population. Critical care medicine. 2013;41(4).

58. Ibrahim I, Jacobs IG, Webb SA, Finn J. Accuracy of International classification of diseases, 10th revision codes for identifying severe sepsis in patients admitted from the emergency department. Critical Care and Resuscitation. 2012;14(2):112.

59. Gedeborg R, Furebring M, Michaëlsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. Journal of clinical epidemiology. 2007;60(2):155. e1-. e11.

60. Henriksen DP, Laursen CB, Jensen TG, Hallas J, Lassen AT. Incidence rate of community-acquired sepsis among hospitalized acute medical patients—a population-based survey. Critical care medicine. 2015;43(1):13-21.

61. World Health Organization. . Available online at: http://wwwwhoint/news-room/fact-sheets/detail/preterm-birth. Accessed February, 2018.

62. Ash A, Shwartz M. R2: a useful measure of model performance when predicting a dichotomous outcome. Statistics in medicine. 1999;18(4):375-84.

63. Born & Growing Annual Report. Born Ontario. 2012-2014.

64. Discharge Abstract Database (DAD) Metadata. Available online at: https://wwwcihica/en/types-of-care/hospital-care/acute-care/dad-metadata. 2017.

65. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. American journal of epidemiology. 1978;107(1):71-6.

66. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological bulletin. 1995;118(3):392.

67. Little R. Generalized linear models for cross-classified data from the WFS. 1978.

68. Myers RH, Myers RH. Classical and modern regression with applications: Duxbury press Belmont, CA; 1990.

69. Berk R, MacDonald JM. Overdispersion and Poisson regression. Journal of Quantitative Criminology. 2008;24(3):269-84.

70. Cameron AC, Trivedi PK. Regression analysis of count data: Cambridge university press; 2013.

71. McCullagh P, Nelder JA. Generalized linear models: CRC press; 1989.

72. Madhu G, SB SS. A study of urinary tract infection in neonatal sepsis. Journal of Evolution of Medical and Dental Sciences. 2014;3(5):1235-40.

73. Provincial Council for Maternal and Child Health Level of Care Designation. https://wwwpcmchonca/wp-content/uploads/2018/06/LOC-Website-List-June-2018-Updatepdf. 2018.

74. Perinatal Health Indicators for Ontario. Born Ontario. 2012;https://www.bornontario.ca/assets/documents/specialreports/Perinatal Health Indicators for Ontario 2012.pdf.

75. 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. Clinical infectious diseases. 2013;58(5):688-96.

76. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. Medical care. 1994;32(1):81-90.

77. 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. Annals of Internal medicine. 2014;160(6):380-8.

78. Yo CH, Hsu TC, Gabriel Lee MT, Porta L, Tsou PY, Wang YH, et al. Trend and outcome of sepsis in children: A nationwide cohort study. Journal of paediatrics and child health. 2018.

79. Rhee C, Murphy MV, Li L, Platt R, Klompas M. Comparison of trends in sepsis incidence and coding using administrative claims versus objective clinical data. Clinical Infectious Diseases. 2014;60(1):88-95.

80. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. American journal of respiratory and critical care medicine. 2016;193(3):259-72.

81. Warren DK, Zack JE, Cox MJ, Cohen MM, Fraser VJ. An educational intervention to prevent catheter-associated bloodstream infections in a nonteaching, community medical center. Critical care medicine. 2003;31(7):1959-63.

82. Coopersmith CM, Rebmann TL, Zack JE, Ward MR, Corcoran RM, Schallom ME, et al. Effect of an education program on decreasing catheter-related bloodstream infections in the surgical intensive care unit. Critical care medicine. 2002;30(1):59-64.

83. Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. Intensive care medicine. 2004;30(4):536-55.

84. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive care medicine. 2013;39(2):165-228.

85. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. Intensive care medicine. 2008;34(1):17-60.

86. Fee N, Hartigan L, McAuliffe FM, Higgins MF. Education in Sepsis: a review for the clinician of what works, for whom, and in what circumstances. Journal of Obstetrics and Gynaecology Canada. 2017;39(9):772-80.

87. Jefferies AL. Management of term infants at increased risk for early-onset bacterial sepsis. Paediatrics & child health. 2017;22(4):223-8.

88. Diseases CoI, Fetus Co, Newborn. Revised guidelines for prevention of early-onset group B streptococcal (GBS) infection. Pediatrics. 1997;99(3):489-96.

89. Schrag S, Gorwitz R, Fultz-Butts K, Schuchat A. Prevention of perinatal group B streptococcal disease. MMWR Recomm Rep. 2002;51(11):1-22.

90. Verani JR, McGee L, Schrag SJ. Prevention of perinatal group B streptococcal disease. Morbidity and Mortality Weekly Report (MMWR), Revised Guidelines from CDC, Recommendations and Reports. 2010;59(RR10):1-32.

91. Money D, Allen VM. No. 298-The Prevention of Early-Onset Neonatal Group B Streptococcal Disease. Journal of Obstetrics and Gynaecology Canada. 2018;40(8):e665-e74.

92. Watson RS, Carcillo JA. Scope and epidemiology of pediatric sepsis. Pediatric Critical Care Medicine. 2005;6(3):S3-S5.

93. Ghuman AK, Newth CJ, Khemani RG. Impact of gender on sepsis mortality and severity of illness for prepubertal and postpubertal children. The Journal of pediatrics. 2013;163(3):835-40. e1.

94. Angele MK, Pratschke S, Hubbard WJ, Chaudry IH. Gender differences in sepsis: cardiovascular and immunological aspects. Virulence. 2014;5(1):12-9.

95. Lefebvre F, Glorieux J, St-Laurent-Gagnon T. Neonatal survival and disability rate at age 18 months for infants born between 23 and 28 weeks of gestation. American journal of obstetrics and gynecology. 1996;174(3):833-8.

96. Hansen BM, Greisen G. Preterm delivery and calculation of survival rate below 28 weeks of gestation. Acta Paediatrica. 2003;92(11):1335-8.

97. Stensvold HJ, Klingenberg C, Stoen R, Moster D, Braekke K, Guthe HJ, et al. Neonatal morbidity and 1-year survival of extremely preterm infants. Pediatrics. 2017:e20161821.

98. Markestad T, Kaaresen PI, Rønnestad A, Reigstad H, Lossius K, Medbø S, et al. Early death, morbidity, and need of treatment among extremely premature infants. Pediatrics. 2005;115(5):1289-98.

99. Patel RM, Kandefer S, Walsh MC, Bell EF, Carlo WA, Laptook AR, et al. Causes and timing of death in extremely premature infants from 2000 through 2011. New England Journal of Medicine. 2015;372(4):331-40.

100. Wynn JL. Defining neonatal sepsis. Current opinion in pediatrics. 2016;28(2):135.

101. Ohlin A, Björkman L, Serenius F, Schollin J, Källén K. Sepsis as a risk factor for neonatal morbidity in extremely preterm infants. Acta Paediatrica. 2015;104(11):1070-6.

102. Barton L, Hodgman JE, Pavlova Z. Causes of death in the extremely low birth weight infant. Pediatrics. 1999;103(2):446-51.

103. Sandberg K, Fasth A, Berger A, Eibl M, Isacson K, Lischka A, et al. Preterm infants with low immunoglobulin G levels have increased risk of neonatal sepsis but do not benefit from prophylactic immunoglobulin G. The Journal of pediatrics. 2000;137(5):623-8.

104. Hobbs J, Davis J. Serum γG-globulin levels and gestational age in premature babies. The Lancet. 1967;289(7493):757-9.

APPENDIX

Appendix 1. Recognized pathogens (bacteria)

Acinetobacter spp:	Escherichia coli	Providencia sp
Acinetobacter baumannii		Providencia rettgeri
Acinetobacter lwoffii	Haemophilus sp.	Providencia stuartii
	Haemophilusinfluenzae	
Bacillus cereus	Haemophilusparainfluenzae	Pseudomonas sp
		Pseudomonas aeruginosa
Bordetella sp	Klebsiella sp.	Pseudomonas cepacia
Bordetella bronchiseptica	Klebsiellaaerogenes	Pseudomonas stutzeri
Bordetella parapertussis	Klebsiellaoxytoca	
Bordetella pertussis	Klebsiellapneumonia	Salmonella sp.
		Shigella sp.
Burkholderiacepacia	Lactobacillus sp.	Shigelladysenteriae
		Shigellaflexneri
Citrobacter sp.	Listeria monocytogenes	Shigellasonnei
CitrobacterKoseri		Serratia sp.
Citrobacterdiversus	Mycobacterium sp.	Serratialiquefaciens
Citrobacterfreundii	Mycobacterium tuberculosis	Serratiamarcescens
Clostridium sp.	Morganellamorganii	Staphylococcus aureus (Methicillin-sensitive and -resistant MSSA or MRSA)
Clostridium difficile		
Clostridium perfringens	Neisseria sp.	
	Neisseria meningitidis	Streptococcus agalactiae or group B streptococcus
Eikenellacorrodens	Neisseria gonorrhoeae	
	Nocardiasp	
Enterococcus spp.	Nocardiaasteroides	Streptococcus pneumoniae
Enterococcus faecalis		Streptococcus pyogenes
Enterococcus faecium	Pantoeasp	
Enterococcus gallinarum	Plesiomonasshigelloides	Stenotrophomonasmaltophilia
Enterobacter sp.	Proteus sp.	
Enterobacter aerogenes	Proteus mirabilis	Ureaplasma sp.
Enterobacter agglomerans	Proteus vulgaris	
Enterobacter cloacae		

Appendix 2. Non-pathogenic bacteria

Bacillus sp. (other than B. cereus)	Streptococcus sp.
	Streptococcus acidominimus
Bacteroides sp.	Streptococcus anginosus
Bacteroidesfragilis	Streptococcus bovis
	Staphylococcus capitis
Coagulase-negative staphylococci	Streptococcus constellatus
Coagulase-negative staphylococci (mixed)	Staphylococcus epidermidis
	Streptococcus equinus
Corynebacterium sp.	Staphylococcus haemolyticus
	Staphylococcus hominis
Diphtheroids	Streptococcus mitis
	Streptococcus mutans
Micrococcus sp.	Streptococcus oralis
	Streptococcus salivarius
Propionibacterium sp.	Streptococcus sanguis
Propionibacterium acnae	Staphylococcus saprophyticus
	Streptococcus suis
Peptococcus sp.	Streptococcus viridans
	Staphylococcus warneri
Peptostreptococcus sp.	
Peptostreptococcusmagnus	Stomatococcusmucilaginosus
Peptostreptococcus micros	

Appendix 3. All therapeutic categories used by the study population in Objective 1.

Code	Therapeutic Category*
81202	AMINOGLYCOSIDES
81206	CEPHALOSPORINS
81207	MISC. B-LACTAM ANTIBIOTICS
81212	ERYTHROMYCINS
81216	PENICILLINS
81220	SULFONAMIDES (SYSTEMIC)
81228	MISC. ANTIBIOTICS
81408	AZOLES
81428	POLYENES
81604	ANTITUBERCULOSIS AGENTS
81808	ANTI-RETRO VIRAL AGENTS
81824	MONOCLONAL ANTIBODIES
81832	NUCLEOSIDES AND NUCLEOTIDES
83092	MISCELLANEOUS ANTIPROTOZOALS
83600	URINARY ANTI-INFECTIVES
120808	ANTIMUSCARINICS/ANTISPASMODICS
121208	BETA-ADRENERGIC AGONISTS
121212	ALPHA AND BETA ADRENERGIC
122000	SKELETAL MUSCLE RELAXANTS
200404	IRON PREPARATIONS
201204	ANTICOAGULANTS
201600	HEMATOPOIETIC AGENTS
240408	CARDIOTONIC AGENTS
240816	CENTRAL ALPHA-AGONISTS
240820	DIRECT VASODILATORS
241292	MISCELLANEOUS VASODILATING AGENTS
242400	BETA-ADRENERGIC BLOCKING AGENTS
242808	DIHYDROPYRIDINES
243220	MINERALOCORTICOID (ALDOSTERONE) ANTAGNTS
280804	NONSTEROIDAL ANTI-INFLAMMATORY
280808	OPIATE AGONISTS
280892	MISC ANALGESICS & ANTIPYRETICS
281212	HYDANTOINS
282092	ANOREXIGENICS; RESPIR; CERE MISC
282404	BARBITURATES
282408	BENZODIAZEPINES

360400	ADRENOCORTICAL INSUFFICIENCY
368400	TUBERCULOSIS
400800	ALKALINIZING AGENTS
401200	REPLACEMENT PREPARATIONS
402000	CALORIC AGENTS
402808	LOOP DIURETICS
402820	THIAZIDE DIURETICS
521000	CARBONIC ANHYDRASE INHIBITORS
560400	ANTACIDS AND ADSORBENTS
561000	ANTIFLATULENTS
561400	CHOLELITHOLYTIC AGENTS
562812	HISTAMINE H2-ANTAGONISTS
562836	PROTON-PUMP INHIBITORS
569200	MISCELLANEOUS GI DRUGS
680400	ADRENALS
682008	INSULINS
682212	GLYCOGENOLYTIC AGENTS
682800	PITUITARY
683604	THYROID AGENTS
683608	ANTITHYROID AGENTS
880800	VITAMIN B COMPLEX
882400	VITAMIN K ACTIVITY
882800	MULTIVITAMIN PREPARTIONS

* Based on AHFS Pharmacologic-Therapeutic Classification system.

Appendix 4. ICD-10-CA Codes for neonatal and adult sepsis

ICD-10_CA Code	Nomenclature
Neonatal Sepsis	
P360	Sepsis of newborn due to streptococcus, group B
P361	Sepsis of newborn due to other and unspecified streptococci
P362	Sepsis of newborn due to Staphylococcus aureus
P363	Sepsis of newborn due to other and unspecified staphylococci
P364	Sepsis of newborn due to Escherichia coli
P365	Sepsis of newborn due to anaerobes
P368	Other bacterial sepsis of newborn
P369	Bacterial sepsis of newborn, unspecified
Adult Sepsis	
A021	Salmonella sepsis
A227	Anthrax sepsis
A267	Erysipelothrix sepsis
A327	Listerial sepsis
A400	Sepsis due to streptococcus, group A
A401	Sepsis due to streptococcus, group B
A402	Sepsis due to streptococcus, group D
A403	Sepsis due to Streptococcus pneumoniae
A408	Other streptococcal sepsis
A409	Streptococcal sepsis, unspecified
A410	Sepsis due to Staphylococcus aureus
A411	Sepsis due to other specified staphylococcus
A412	Sepsis due to unspecified staphylococcus
A413	Sepsis due to Haemophilus influenzae
A414	Sepsis due to anaerobes
A4150	Sepsis due to Escherichia coli [E.coli]
A4151	Sepsis due to Pseudomonas
A4152	Sepsis due to Serratia
A4158	Sepsis due to other Gram-negative organisms
A4180	Sepsis due to Enterococcus
A4188	Other specified sepsis
A419	Sepsis, unspecified
A427	Actinomycotic sepsis
A5486	Gonococcal sepsis
B377	Candidal sepsis
O85002	Puerperal sepsis, delivered, with mention of postpartum complication

O85004	Puerperal sepsis, postpartum condition or complication
O85009	Puerperal sepsis, unspecified as to episode of care, or not applicable

Appendix 5. The Searching strategy for sepsis validation studies.

Searching strategy of sepsis literatures was as follows:

Firstly, 'humans' and 'English language' filters were applied.

Secondly, in order to identify studies assessing the diagnostic accuracy of ICD codes for identifying sepsis, 'AND' was used to combine three search concepts: sepsis, coding and validation.

a. Articles concerning sepsis were sought using 'OR' to combine the term 'sepsis' and the terms relevant to the condition of sepsis, including 'severe sepsis' and 'septic shock'.
b. Articles concerning the concept of coding were sought using 'OR' to combine keyword searches for the following terms: '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' and 'codes'.
c. Articles concerning validity were sought using 'OR' to combine the keyword searches for the terms 'validity', 'validation', 'case definition', 'algorithm', 'accuracy', 'sensitivity', 'specificity', 'positive predictive value', 'negative predictive value', 'positive likelihood ratio'

To be eligible for inclusion, articles had to compare the accuracy of ICD-9 or ICD-10 codes for sepsis in an administrative database to a reference standard and report at least one of diagnostic accuracy measures (i.e., sensitivity, specificity, positive predictive value or negative predictive value). Twelve articles met all eligibility criteria and were worth summarizing. None of above twelve publications through the above searching strategy were conducted specifically for validation of algorithms for identifying neonatal sepsis.