BMJ Open

Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data

Rachel J Jolley,[1] Hude Quan,[1,2] Nathalie Jetté,[1,2,3,4] Keri Jo Sawka,[1] Lucy Diep,[1] Jade Goliath,[1] Derek J Roberts,[1,5,6] Bryan G Yipp,[6,7] Christopher J Doig[1,6,7]

For numbered affiliations see end of article.

Correspondence to
Dr Christopher J Doig;
cdoig@ucalgary.ca

ABSTRACT

Objective: Administrative health data are important for health services and outcomes research. We optimised and validated in intensive care unit (ICU) patients an International Classification of Disease (ICD)-coded case definition for sepsis, and compared this with an existing definition. We also assessed the definition's performance in non-ICU (ward) patients.

Setting and participants: All adults (aged ≥18 years) admitted to a multisystem ICU with general medicosurgical ICU care from one of three tertiary care centres in the Calgary region in Alberta, Canada, between 1 January 2009 and 31 December 2012 were included.

Research design: Patient medical records were randomly selected and linked to the discharge abstract database. In ICU patients, we validated the Canadian Institute for Health Information (CIHI) ICD-10-CA (Canadian Revision)-coded definition for sepsis and severe sepsis against a reference standard medical chart review, and optimised this algorithm through examination of other conditions apparent in sepsis.

Measures: Sensitivity (Sn), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV) were calculated.

Results: Sepsis was present in 604 of 1001 ICU patients (60.4%). The CIHI ICD-10-CA-coded definition for sepsis had Sn (46.4%), Sp (98.7%), PPV (98.2%) and NPV (54.7%); and for severe sepsis had Sn (47.2%), Sp (97.5%), PPV (95.3%) and NPV (63.2%). The optimised ICD-coded algorithm for sepsis increased Sn by 25.5% and NPV by 11.9% with slightly lowered Sp (85.4%) and PPV (88.2%). For severe sepsis both Sn (65.1%) and NPV (70.1%) increased, while Sp (88.2%) and PPV (85.6%) decreased slightly.

Conclusions: This study demonstrates that sepsis is highly undercoded in administrative data, thus under-ascertaining the true incidence of sepsis. The optimised ICD-coded definition has a higher validity with higher Sn and should be preferentially considered if used for surveillance purposes.

Strengths and limitations of this study

- This study examined the validity of an optimised International Classification of Disease (ICD)-10-CA-coded case definition to identify sepsis and severe sepsis in an inpatient administrative database for both ICU and non-ICU patients.
- Sepsis is undercoded in administrative data. Although sepsis is undercoded, our algorithm identifies with confidence a cohort of patients with sepsis (a minimum number of false-positive cases). This algorithm is optimal for studies where identifying a cohort of true sepsis cases is important.
- We also report an algorithm that optimises the identification of patients with an increased case-capture rate for sepsis (although a slight increase in the number of false positives): this algorithm may be optimal for surveillance studies.
- Sepsis is a hard-to-define condition. A validated algorithm to identify patients with sepsis from administrative data may facilitate health services research into this expensive and high morbidity and mortality condition.

INTRODUCTION

Sepsis is a life-threatening condition with a high rate of occurrence in the intensive care unit (ICU).[1 2] It is one of the most costly diseases to treat[3 4] leaving long-term physical and cognitive effects on its survivors.[5] Historically, sepsis has been difficult to define, diagnose and treat.[6] In 1992, the American College of Chest Physicians and Society for Critical Care Medicine (ACCP/SCCM) published the first consensus clinical definitions of sepsis outlining the terminology and clinical characteristics of the spectrum of illness.[7] In 2001, these clinical definitions were updated to provide more clarification on the signs and symptoms of the disease, and to identify methodologies to increase the accuracy and reliability of the diagnosis of sepsis.[8] Since the consensus conference clinical definitions were published, most studies use these clinical definitions

regardless of study type (ie, clinical trial or health services research) and/or data source (ie, administrative data, or prospective clinical record).

Administrative health data are widely collected, and are a generally cost-effective way of studying multiple outcomes, health service usage and resource allocation in large populations.[9] Administrative data typically use WHO's International Classification of Diseases (ICD)[10] codes, an alphanumeric classification system including a core code category made up of the first three characters that are mandatory reporting to facilitate international comparisons, with the most recent update, ICD-10, released in 1994. A major advantage of ICD-10 is that it contains almost twice the number of codes (12 420 codes in ICD-10 compared against 6882 in ICD-9) permitting richer and more precise capture of clinical information, allowing for improved international comparability.[11 12]

However, irrespective of coding systems, it may be difficult to recognise and translate complex conditions, such as sepsis, into a single code. Therefore, often for complex conditions such as sepsis, multiple codes may exist. Some studies have used infection codes,[13] or a more limited number of codes, for sepsis.[14] Reported sensitivities in validation studies have ranged from 5.9% to 82.3%.[15–25] These studies varied significantly in the number and types of codes applied, and the methods in developing the ICD coding algorithms.

The Canadian Institute for Health Information (CIHI) created an ICD-10-CA (Canadian Revision)-coded case definition to define sepsis in administrative data.[26] This particular definition uses 49 ICD codes to define sepsis (in adult and neonate populations) and 28 codes specific to organ dysfunction for severe sepsis. The Canadian Revision which includes more detailed subcodes, however, remains true to the original ICD-10 implementation. The CIHI administrative data-coded definition, although using the enhanced capability of ICD-10, has not been validated. An accurate and validated ICD-coded case definition is important, as healthcare resource allocation and other healthcare delivery system decisions can be and have been determined from these data.[9]

We therefore examined the validity of the CIHI ICD-10-coded case definition in ICU and non-ICU settings, and determined if it could be improved to increase the accuracy of case capture for a diagnosis of sepsis.

METHODS
Data sources and study population
This study used two databases, the inpatient discharge abstract database (DAD), which has detailed information including demographic, administrative and procedural data on inpatient hospital visits, with each inpatient visit record containing up to 50 ICD-10-CA diagnosis coding fields recorded per hospital encounter. Of these, 25 are

released to researchers. In prior research in acutely ill patient populations including diagnoses of catheter-related blood-stream infections and postoperative sepsis, the minimum number of diagnostic coding fields needed to capture at least 90% of secondary diagnosis cases was 15 fields.[27] Clinical data were also abstracted from an ICU-specific clinical database (TRACER—details described elsewhere)[28] containing ICU-specific clinical and demographic characteristics including APACHE (acute physiology and chronic health evaluation) II[29] and SOFA[30] (sequential organ dysfunction assessment) scores. Medical charts were also reviewed. All data were linked using the Alberta personal health number, which is a unique lifetime identifier.

Our study population comprised two separate validation cohorts. The first cohort included all adult patients (aged 18 years and older) admitted to an ICU in one of three hospitals in the Calgary region in Alberta, Canada, between 1 January 2009 and 31 December 2012. All three hospitals contain a multisystem ICU with general medicosurgical ICU care; Foothills Medical Centre (FMC) includes a regional specialty programme of burns, trauma surgery, neurosciences, thoracic surgery and transplant surgery (renal, pancreas, bone marrow); Peter Lougheed Centre includes a vascular surgery programme and the Rockyview General Hospital includes the regional urological and ENT programme. The second cohort included a random selection of all non-ICU, or general medical and surgical inpatient medical records from the FMC in Alberta, Canada, between 1 January 2009 and 31 December 2012.

Defining sepsis in medical chart and data abstraction
Sepsis was defined in the medical record review using a checklist criteria tool (table 1) developed based on the ACCP/SCCM 2001 Consensus Conference updated definitions[8] and consensus of clinical experts. The tool was tested through a consensus review completed by two independent physicians, one trained in intensive care medicine and the other in surgery (BGY and DJR). Each physician was given the same 10 randomly selected health records, with health record coding masked, and using the tool, determined if sepsis was present or absent for each case. If sepsis was present, the classification of severity (sepsis, severe sepsis and septic shock) was indicated. These results were compared and discussed to ensure full consensus. A full consensus agreement (κ statistic=1.00) occurred after the first round of 20 medical charts, validating the tool for use in the subsequent part of the study.

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

Table 1 Diagnostic criteria used to determine a diagnosis of sepsis, severe sepsis or septic shock

Infection (infection defined as a pathological process induced by a microorganism)—documented or suspected, through microbiological or other equivalent diagnostic confirmation, and some of the following				
Suspected infection symptoms/signs (patient symptoms/ physical findings)	▶ Chills ▶ Fever ▶ Rigors ▶ Rash ▶ Dysuria	▶ Dyspnoea ▶ Confusion (delerium, encephalopathy) ▶ Stiff neck ▶ New heart murmurs	▶ Bronchial breath sounds ▶ Pleuritic chest pain ▶ Peritoneal findings (acute or rigid abdomen, rebound) ▶ Abdominal pain	▶ Pain out of proportion ▶ Purulent wound ▶ Cellulitis ▶ Skin changes of necrotising facitis
SIRS criteria At least 2 of the following	Fever: temperature >38.3°C Hypothermia: temperature <36°C Tachycardia: heart rate >90/min Tachypnoea (respiratory rate of more than 20 breaths/min) Leucocytosis (WCC count >12×10^9/L) Leukopenia (WCC count <4×10^9/L) WCC count with >10% immature granulocytes (bands+myelocytes+metamyelocytes)			
Organ dysfunction variables At least 1 of the following	PaO_2/FiO_2 <300 or <200 in patients with lung injury PaO_2/FiO_2 ratio <250 in patients with bilateral pulmonary infiltrates Any FiO_2; SaO_2 <90% w/FiO_2 ≥50% Decreased $SSVCO_2$ ≤70% Need for non-elective invasive or non-invasive mechanical ventilation Creatinine >176.8 μmol/L >50% increase in SCr from baseline Decreased urine output <0.5 mL/kg h for >2 or 45 mL/h for at least 2 h, despite adequate fluid resuscitation Total bilirubin >34.2 μmol/L Asparta transaminase >80 IU/L Alanine transaminase >80 IU/L Decreased consciousness or GCS ≤11 Low platelet count (<100×10^9/L) Prolonged capillary refill time (>3 s) INR >1.5 Lactate (arterial) >2.2 mmol/L			
Shock/hypotension variables	SBP<90 mm Hg MAP<65 mm Hg SBP decrease of >40 mm Hg from baseline Vasopressors—any continuous infusion, any dose or otherwise indicated: epinephrine or norepinephrine, vasopressin >0.02 u/min, dobutamine, dopamine >6 μg/kg/min			

Sepsis is defined as infection plus at least 2 SIRS criteria, severe sepsis defined as sepsis plus at least one organ dysfunction variable, and septic shock was defined as severe sepsis with one of the shock/hypotension variables.
Other equivalent diagnostic confirmation includes: laparotomy or surgical findings of infection, surgical debridement, chest X-ray consistent with pneumonia, chest X-ray consistent with ARDS/ALI. Any X-ray/CT consistent with ischaemia, any X-ray/CT consistent with infection, any X-ray/CT consistent with abscess, abdominal X-ray with free air.
ALI, acute lung injury; ARDS, acute respiratory distress syndrome; FiO_2, fraction of inspired oxygen; GCS, Glasgow Coma Scale; INR, international normalised ratio; MAP, mean arterial blood pressure; PaO_2, symbol for partial pressure of oxygen in arterial blood; SaO_2, saturated arterial oxygen; SBP, systolic blood pressure; SCr, serum creatinine; $SSVCO_2$, saturated venous gas; WCC, white cell count.

done until the strength of agreement achieved among all four reviewers was near perfect (κ statistic between 0.81 and 1.00).[31] Two rounds of review were performed; the κ score was calculated after each round until full consensus was reached; any remaining discrepancies were discussed and resolved through a third-party expert reviewer (CJD). Following the consensus review, data abstraction was completed independently. Cases with uncertainty were discussed to ensure consistency among all reviewers, and any major unresolved cases were brought to a third-party critical care physician (CJD) for resolution.

Defining sepsis in ICD administrative data
Administrative data from the DAD were obtained for each patient corresponding to the specified inpatient visit during the study period. Using the DAD, sepsis was defined as per CIHI's 2009 report[26] by searching through any 1 of the 25 diagnosis coding fields for any of the codes listed in table 2. Any neonate and paediatric-specific codes from the original definition remained in the algorithm, although we limited our study population to adults. Severe sepsis was indicated by the combination of a code of sepsis and at least one organ dysfunction code.

After the primary analysis, we revised the CIHI ICD-10-CA-coded case definition for sepsis informed by a systematic review of the existing literature.[32] We examined ICD-10-CA codes to determine if codes, which may indicate sepsis, were missing and should be included in the new definition based on clinical knowledge of the resulting diagnosis (see table 3 for a list of all ICD codes

Table 2 ICD-10-CA codes used to define sepsis and severe sepsis in administrative data by ICD-coded case definition

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

*ICD-10-CA (Canadian edition) specific codes.
CIHI, the Canadian Institute for Health Information; ICD, International Classification of Disease.

used with description). As well, we determined the codes in the primary diagnostic coding position that had a high frequency in the false-negative population. We performed an additive analysis in which each possible new code was added individually to the original CIHI definition (see online supplementary table S1), as well as the inverse in which all new codes were included in the original definition, with the removal of each individually to determine the changes in accuracy until the most optimal values of sensitivity (Sn), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV) were achieved.

Statistical analysis

A sample size calculation estimated that 409 charts were required using an estimated prevalence of 19%,[33] at a significance level of 5% and 99% confidence. In order to gain a representative sample of the population, a random sample of 1001 patients was selected spread across the three tertiary care hospitals. Descriptive statistics were used to describe the study populations acquired by each ICD-coded case definition. The Charlson comorbidity score was calculated using previously described methods.[34] Sn, Sp, PPV, NPV and their 95% CIs for the CIHI and optimised coding algorithm were calculated. Sn was calculated as the proportion of cases classified as positive by both the administrative data (DAD) and medical record review or 'true positives' (TP) compared with all cases positive by the reference standard (medical record review). Sp was calculated as the proportion of cases without sepsis identified by both the DAD and medical record review, or 'true negatives'

(TN), compared with all cases negative by the reference standard. PPV was calculated as the proportion of TP cases of sepsis compared with all the cases identified as sepsis by the DAD. NPV was calculated as the proportion of cases without sepsis (TN) compared with all the sepsis compared with all the cases identified as not sepsis by the DAD. All statistical analyses were performed using STATA V.12 (Stata Corp., College Station, Texas, USA).[35]

RESULTS
Patient characteristics for reference standard diagnosis

A total of 1001 patients admitted to the ICU were included and linked to the DAD and TRACER databases. Of these, 604 patients were classified as sepsis (86 (14.2%) with sepsis, 203 (33.6%) with severe sepsis, 315 (52.2%) with septic shock,) and 397 were classified as not sepsis. Of the sepsis patients included in the study, 59.3% were men, their median age was 61 years, 76.5% were admitted through the emergency department (ED), and 44.9% had two or more Charlson comorbidities (table 4). The mean APACHE II score within the first 24 h of admission was 20.8, and the admission SOFA score was 6.6. Median hospital length of stay (LOS) was 19 days, and median ICU LOS was 5.8 days. ICU mortality was 17.1% and hospital mortality was 24.0%.

Patient characteristics for the CIHI and optimised algorithm

There were 285 cases of sepsis identified by the CIHI algorithm, and 257 cases of severe sepsis. The optimised

Table 3 ICD-10-CA codes and descriptions

Diagnostic code	Code description
Sepsis	
A03.9	Shigellosis, unspecified
A02.1	*Salmonella* sepsis
A04.7	Enterocolitis due to *Clostridium difficile*
A20.7	Septicaemic plague
A21.7	Generalised tularaemia
A22.7	Anthrax sepsis
A23.9	Brucellosis, unspecified
A24.1	Acute and fulminating melioidosis
A26.7	Erysipelothrix sepsis
A28.0	Pasteurellosis
A28.2	Extraintestinal yersiniosis
A32.7	Listerial sepsis
A39.2	Acute meningococcaemia
A39.3	Chronic meningococcaemia
A39.4	Meningococcaemia, unspecified
A40	Streptococcal sepsis
A40.0	Sepsis due to *Streptococcus*, group A
A40.1	Sepsis due to *Streptococcus*, group B
A40.2	Sepsis due to *Streptococcus*, group D
A40.3	Sepsis due to *Streptococcus pneumoniae*
A40.8	Other streptococcal sepsis
A40.9	Streptococcal sepsis, unspecified
A41	Other sepsis
A41.0	Sepsis due to *Staphylococcus aureus*
A41.1	Other sepsis
A41.2	Sepsis due to unspecified *Staphylococcus*
A41.3	Sepsis due to *Haemophilus influenzae*
A41.4	Sepsis due to anaerobes
A41.5	Sepsis due to other Gram-negative organisms
A41.50*	Sepsis due to *Escherichia coli*
A41.51*	Sepsis due to *Pseudomonas*
A41.52*	Sepsis due to *Serratia*
A41.58*	Sepsis due to other Gram-negative organisms, NOS
A41.8	Other specified sepsis
A41.80*	Sepsis due to *Enterococcus*
A41.88*	Other specified sepsis
A41.9	Sepsis, unspecified, includes: septicaemia
A42.7	Actinomycotic sepsis
B00.7	Disseminated herpes viral disease, includes: herpes viral sepsis
B37.7	Candidal sepsis
B95.48	Other *Streptococcus* as the cause of diseases classified to other chapters
B95.6	*S. aureus* as the cause of diseases classified elsewhere
B96.2	*E. coli* as the cause of diseases classified elsewhere
J18.9	Pneumonia, unspecified organism
J44.0	Chronic obstructive pulmonary disease with acute lower respiratory infection
N39.0	Urinary tract infection, site not specified
P360	Sepsis of newborn due to *Streptococcus*, group B

Continued

Table 3 Continued

Diagnostic code	Code description
P361	Sepsis of newborn due to other and unspecified streptococci
P362	Sepsis of newborn due to *S. aureus*
P363	Sepsis of newborn due to other and unspecified staphylococci
P364	Sepsis of newborn due to *E. coli*
P365	Sepsis of newborn due to anaerobes
P368	Other bacterial sepsis of newborn
P369	Bacterial sepsis of newborn, unspecified
P352	Congenital herpes viral (herpes simplex) infection
P372	Neonatal (disseminated) listeriosis
P375	Neonatal candidiasis
Respiratory	
J96.0	Acute respiratory failure
J96.9	Respiratory failure, unspecified
J80	Diseases of bronchus, not elsewhere classified
R09.2	Respiratory arrest
Cardiovascular	
R57.0	Cardiogenic shock
R57.1	Hypovolaemic shock
R57.8	Other shock
R57.9	Shock, unspecified
I95.1	Orthostatic hypotension
I95.9	Hypotension, unspecified
Renal	
N17.0	Acute renal failure with tubular necrosis
N17.1	Acute renal failure with acute cortical necrosis
N17.2	Acute renal failure with medullary necrosis
N17.8	Other acute renal failure
N17.9	Acute renal failure, unspecified
Neurological	
K72.0	Acute and subacute hepatic failure
K72.9	Hepatic failure, unspecified
K76.3	Infarction of liver
F05.0	Delirium not superimposed on dementia, so described
F05.9	Delirium, unspecified
G93.1	Anoxic brain damage, not elsewhere classified
G93.4	Encephalopathy, unspecified
G93.80	Metabolic encephalopathy
Haematological	
D69.5	Secondary thrombocytopenia
D69.6	Thrombocytopenia, unspecified
D65	Disseminated intravascular coagulation (defibrination syndrome)

*ICD-10-CA (Canadian edition) specific codes.
ICD, International Classification of Disease; NOS, not otherwise specified.

ICD-coded case definition increased the number of cases of sepsis identified by 207 (n=492), and 138 for severe sepsis (n=395). The optimised definition had similar

Table 4 Patient clinical characteristics and demographics of the study population by ICD-coded algorithm and reference standard definition (n=1001)

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

*Emergent surgery refers to surgery needed within 24–48 h since admission with no prior indication of surgery needed before the present admission.
†APACHE II score recorded within the first 24 h of admission by physician.
‡Admission SOFA score—the maximum score recorded within the first 24 h of admission to the ICU.
APACHE, acute physiology and chronic health evaluation; CIHI, the Canadian Institute for Health Information; COPD, chronic obstructive pulmonary disease; ER, emergency room; ICD, International Classification of Disease; ICU, intensive care unit; LOS, length of stay; SOFA, sequential organ dysfunction assessment.

cohort characteristics in the sepsis and severe sepsis populations compared with the CIHI definition, however, the CIHI definition patients had higher mean APACHE II scores for both sepsis (22.9 vs 20.9) and severe sepsis (23.6 vs 22.4), and higher admission SOFA scores for sepsis (7.5 vs 6.6) and severe sepsis (7.7 vs 7.0) (see table 4). Median ICU LOS was higher in the patients identified with the optimised severe sepsis ICD-coded case definition at 6.3 vs 5.9 days in the CIHI definition, while overall hospital LOS was similar among each cohort. ICU mortality was 6.6%

higher in patients with sepsis, and 4.4% higher in patients with severe sepsis classified based on the CIHI coding definition. Hospital mortality was 7.0% higher in patients with sepsis, and 4.2% higher in patients with severe sepsis identified by the CIHI coding definition.

Performance of ICD-coded case definitions for sepsis classification in ICU patients
The results of the performance of each ICD-coded case definition are shown in table 5. The CIHI ICD-10-CA

Table 5 Validity by administrative data definition/coding algorithm

Definition/coding algorithm	TP (n)	FN (n)	FP (n)	TN (n)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
ICU population (n=1001)								
CIHI								
Sepsis	280	324	5	392	46.4 (42.3 to 50.4)	98.7 (97.0 to 99.6)	98.2 (96.0 to 99.4)	54.7 (51.0 to 58.4)
Severe sepsis	245	274	12	470	47.2 (42.8 to 51.6)	97.5 (95.7 to 98.7)	95.3 (92.0 to 97.6)	63.2 (59.6 to 66.6)
Optimised								
Sepsis	434	170	58	339	71.9 (68.1 to 75.4)	85.4 (81.5 to 88.7)	88.2 (85.0 to 90.9)	66.6 (62.3 to 70.7)
Severe sepsis	338	181	57	425	65.1 (60.9 to 69.2)	88.2 (85.0 to 90.9)	85.6 (81.7 to 88.9)	70.1 (66.3 to 73.8)
Non-ICU population (n=202)								
CIHI								
Sepsis	1	14	0	187	6.7 (0.1 to 31.9)	100	100	93.0 (88.6 to 96.1)
Severe sepsis	1	3	0	198	25 (0.6 to 80.6)	100	100	98.5 (95.7 to 99.7)
Optimised								
Sepsis	9	6	10	177	60 (32.2 to 83.7)	94.7 (90.4 to 97.4)	52.6 (28.9 to 75.6)	96.7 (93.0 to 98.8)
Severe sepsis	1	3	1	197	25 (0.6 to 80.6)	99.5 (97.2 to 99.9)	50 (1.3 to 98.7)	98.5 (95.7 to 99.7)

CIHI, the Canadian Institute for Health Information; FN, false negative; FP, false positive; ICU, intensive care unit; NPV, negative predictive value; PPV, positive predictive value; TN, true negatives; TP, true positives.

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

Performance of ICD-coded case definitions for sepsis classification in non-ICU patients

A total of 202 non-ICU patient medical records were included and linked to the DAD. For the non-ICU population, the CIHI-coded case definition for a diagnosis of sepsis had extremely low Sn of 6.7%, and for severe sepsis it was slightly higher with Sn of 25%; however, both were highly specific at 100% and had high PPV and NPV (table 5). The optimised ICD-coded case definition improved the Sn for sepsis cases to 60%, while the Sn remained the same for severe sepsis at 25%, however, in both cases, the PPV was decreased substantially to 52.6% for sepsis and 50% for severe sepsis.

DISCUSSION

This study examined the validity of an optimised ICD-10-CA-coded case definition to identify sepsis and severe sepsis in an inpatient administrative database. We identified ICD codes that optimised the performance of the coded definitions, and our data show the new, optimised ICD-10-CA-coded definitions with added codes achieve a higher validity than the existing CIHI definition. We increased the Sn by over 25% in the ICU

population without losing much Sp by including codes for pneumonia (J189), enterocolitis due to *Clostridium difficile* (A047), chronic obstructive pulmonary disease with acute lower respiratory infection (J440), other *Streptococcus* as the cause of diseases classified elsewhere (B9548), *Staphylococcus aureus* as the cause of diseases classified elsewhere (B956) and *Escherichia coli* as the cause of diseases classified elsewhere (B962). The code for septic shock (R572) was missing from the original CIHI definition, and was also included in the new definition.

When sepsis is identified and coded, it is relatively accurate, as determined by the moderate to excellent Sp and high PPV in our results. This optimised ICD-based case definition, although capturing more cases, is still only moderately sensitive suggesting that sepsis is under-coded in administrative data. Our ICD case definition has Sn of 71.9%, similar to that of other hospital-acquired infections internationally,[36] and for non-communicable diseases, such as hypertension[37] and diabetes,[38] in Canadian data. The low NPV achieved by our definition for both sepsis and severe sepsis codes may be related to the high prevalence of sepsis in ICU patients.[39] In patients admitted to non-ICU settings, sepsis may not be detected well at any point during their hospital stay, as shown in our analysis of non-ICU patients. Although some studies have suggested that patients with severe sepsis are commonly admitted to non-ICU settings,[40 41] these studies have sometimes been based on administration of antibiotics in the ED as the criterion for suspected infection, or case identification using anecdotal screening rather than a developed objective instrument. In our anecdotal experience, most patients with an estimated mortality rate of 20% or higher at the time of hospital admission usually receive treatment in an ICU setting. Approximately 80% of our patients were admitted to the ICU directly from the ED,

whereas, the remaining patients were admitted from another hospital ward. It may be that severe sepsis is not highly prevalent in non-ICU settings, or it may be that coding for sepsis in non-ICU settings is often missed. Although our sample size in the non-ICU patients was smaller than the ICU, our results did demonstrate a high Sp and NPV indicating that when sepsis was coded as not present in the non-ICU population, it was accurate.

Undercoding may have important implications if used for surveillance of sepsis, or planning of resources, and allocation of services. Other conditions have also been found to be grossly undercoded, resulting in inaccurate assessments of prevalence, and thereby contributing to inadequate allocation of resources for monitoring and appropriate treatment.[42] For sepsis survivors, it is important to have an accurate way of capturing these patients for future planning as they are at a high risk for long-term neurocognitive and physical conditions.[43–45] Further, these coding definitions could be used for quality assessment surveillance monitoring studies, for example, to document the rapidity of administration of antibiotics.

The undercoding of sepsis could be due to a variety of other reasons including physician documentation in the medical record. Healthcare coders may not identify a diagnosis of sepsis based on the physician's documentation alone. Physicians may not explicitly state the term 'sepsis' within the medical chart, instead terms such as 'SIRS' or 'shock' are used, or identifying only the infection present. Rothberg et al[46] suggest that patients may be diagnosed with respiratory failure having the symptoms of pneumonia, and/or criteria of sepsis without identifying the specific condition or sepsis. As well, selective undercoding of a milder form of sepsis may occur, as coders may intentionally disregard coding sepsis if there are other more resource-intensive and very apparent diagnoses present, that is, any highly acute but mild cases of sepsis that clinically resolves quickly, but where a patient has an extended hospital stay for another reason complicating the episode of sepsis, sepsis may be missed as contributing to the hospital stay.[47] Although new clinical definitions for sepsis have been developed, and/or may be developed in the future, how these definitions are applied in research involving administrative data is uncertain. Definitions that rely on specific laboratory results such as pro-calcitonin levels, may not be captured by healthcare coders unfamiliar with the specific implications of these diagnostic results such as these.

Other studies that have examined the definition of sepsis in administrative data have also identified variations in reporting. Gaieski et al[48] examined four previously published methods of capturing cases of severe sepsis in administrative data using ICD-9 codes including the well known 'Angus' and 'Martin' implementations, and compared the incidence and mortality over a 6-year period. They identified up to a 3.5-fold variation among four sepsis case definitions in incidence, with a number of cases ranging from 894 013 to 3 110 630, and mortality ranging from 14.7% to 29.9% depending on the ICD-9 definition used. Iwashyna et al[49] validated the ICD-9 coding definitions for the Angus and Martin implementations and found these to have low sensitivities when identifying severe sepsis using administrative data. These studies along with our results suggest the need for linkages of administrative to other types of data, such as pharmacy data (eg, antibiotics or inotropic use), to enhance the ascertainment of sepsis for surveillance purposes.

Limitations
There are several limitations to this study. First, we defined our reference standard using medical record data extracted by reviewers to assess the validity of the ICD-10-CA data. The potential for misclassification of sepsis within the chart review may have occurred, however, we used a comprehensive process for training and validation to mitigate this possibility. The ICU patient population was selected from tertiary care centres in a large metropolitan area which may then influence the generalisability of case capture to data coming from smaller community hospitals. We also could not validate the optimised algorithm on a different patient sample due to feasibility of medical record review, which therefore, may also impact the generalisability of the case capture. However, we believe that based on the representativeness of the original sample, the optimised definition would still have performed better than the original CIHI case definition. We would encourage other investigators to examine the performance of our reported algorithm in other data sets.

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

Author affiliations
[1]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[2]O'Brien Institute for Public Health, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[3]Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[4]Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[5]Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary, Canada

[6]Department of Critical Care Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
[7]Snyder Institute of Chronic Diseases, University of Calgary, Calgary, Alberta, Canada

REFERENCES

1. Vincent JL, Sakr Y, Sprung CL, et al., Sepsis Occurrence in Acutely Ill Patients Investigators. Sepsis in European intensive care units: results of the SOAP study. *Crit Care Med* 2006;34: 344–53.
2. Alberti C, Brun-Buisson C, Burchardi H, et al. Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. *Intensive Care Med* 2002;28:108–21.
3. Weycker D, Akhras KS, Edelsberg J, et al. Long-term mortality and medical care charges in patients with severe sepsis. *Crit Care Med* 2003;31:2316–23.
4. Torio CM, Andrews RM. *National Inpatient Hospital Costs: the Most Expensive Conditions by Payer, 2011. HCUP Statistical Brief #160.* Rockville, MD: Agency for Healthcare Research and Quality, August 2013. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf (accessed 1 Oct 2014).
5. Iwashyna TJ, Ely EW, Smith DM, et al. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA* 2010;304:1787–94.
6. Bone RC. Let's agree on terminology: definitions of sepsis. *Crit Care Med* 1991;19:973–6.
7. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992;101:1644–55.
8. Levy MM, Fink MP, Marshall JC, et al. SCCM/ESICM/ACCP/ATS/ SIS: 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003;31:1250–6.
9. O'Malley KJ, Cook KF, Price MD, et al. Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005;40:1620–39.
10. World Health Organization. International Classification of Diseases (ICD). http://www.who.int/classifications/icd/en/ (accessed 3 Feb 2014).
11. Jette N, Quan H, Hemmelgarn B, et al. IMECCHI Investigators. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Med Care* 2010;48:1105–10.
12. De Coster C, Quan H, Finlayson A, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. *BMC Health Serv Res* 2006;6:77.
13. Angus DC, Linde-Zwirble WT, Lidicer J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001;29: 1303–10.
14. Martin GS, Mannino DM, Eaton S, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 2003;348:1546–54.
15. Cevasco M, Borzecki AM, Chen Q, et al. Positive predictive value of the AHRQ patient safety indicator "postoperative sepsis": implications for practice and policy. *J Am Coll Surg* 2011;212:954–61.
16. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. *Pharmacoepidemiol Drug Saf* 2008;17:890–5.
17. Gedeborg R, Furebring M, Michaelsson K, et al. Diagnosis-dependent misclassification of infections using administrative data variably affected the incidence and mortality estimates of ICU patients. *J Clin Epidemiol* 2007;60:155–62.
18. Ibrahim I, Jacobs IG, Webb SAR, et al. Accuracy of international classification of diseases, 10th revision codes for identifying severe sepsis in patients admitted from the emergency department. *Crit Care Resusc* 2012;14:112–18.
19. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann Surg* 2012;256:973–81.
20. Madsen KM, Schonhyder HC, Kristensen B, et al. Can hospital discharge diagnosis be used for surveillance of bateremia? A data quality study of a Danish hospital discharge registry. *Infect Control Hosp Epidemiol* 1998;19:175–80.
21. Ollendorf DA, Fendrick AM, Massey K, et al. Is sepsis accurately coded on hospital bills? *Value Health* 2002;5:79–81.
22. Quan H, Eastwood C, Cunningham CT, et al. IMECCHI investigators: Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). *BMJ Open* 2013;3:1–7.
23. Ramanathan R, Leavell P, Stockslager G, et al. Validity of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) Screening for Sepsis in Surgical Mortalities. *Surg Infect (Larchmt)* 2014;15:513–16.
24. Schneeweiss S, Robicsek A, Scranton R, et al. Veteran's affairs hospital discharge database coded serious bacterial infections accurately. *J Clin Epid* 2007;60:397–409.
25. Whittaker SA, Mikkelsen ME, Gaieski DF, et al. Severe sepsis cohorts derived from claims-based strategies appear to be biased toward a more severely ill patient population. *Crit Care Med* 2013;41:945–53.
26. Canadian Institute for Health Information. *Focus: a national look at sepsis.* Ottawa, OT: CIHI, 2009. https://secure.cihi.ca/free_products/ HSMR_Sepsis2009_e.pdf (accessed 2 May 2013).
27. Drösler SE, Romano PS, Sundararajan V, et al. World Health Organization Quality and Safety Topic Advisory Group. How many diagnosis fields are needed to capture safety events in administrative data? Findings and recommendations from the WHO ICD-11 Topic Advisory Group on Quality and Safety. *Int J Qual Health Care* 2014;26:16–25.
28. Doig CJ, Zygun DA, Fick GH, et al. Study of clinical course of organ dysfunction in intensive care. *Crit Care Med* 2004;32:384–90.
29. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–29.
30. Janssens U, Graf C, Graf J, et al. Evaluation of the SOFA score: a single-center experience of a medical intensive care unit in 303 consecutive patients with predominantly cardiovascular disorders. Sequential Organ Failure Assessment. *Intensive Care Med* 2000;26:1037–45.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
32. Jolley RJ, Sawka KJ, Yergens DW, et al. Validity of administrative data in recording sepsis: a systematic review. *Crit Care* 2015; 19:139.
33. Martin CM, Priestap F, Fisher H, et al. STAR Registry Investigators. A prospective, observational registry of patients with severe sepsis: the Canadian Sepsis Treatment and Response Registry. *Crit Care Med* 2009;37:81–8.

34. Quan H, Sundararajan V, Halfon P, *et al*. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.
35. StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.
36. Goto M, Ohl ME, Schweizer ML, *et al*. Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. *Clin Infect Dis* 2014;58:688–96.
37. Quan H, Khan N, Hemmelgarn BR, *et al*, Hypertension Outcome and Surveillance Team of the Canadian Hypertension Education Programs. Validation of a case definition to define hypertension using administrative data. *Hypertension* 2009;54:1423–8.
38. Southern DA, Roberts B, Edwards A, *et al*. Validity of administrative data claim-based methods for identifying individuals with diabetes at a population level. *Can J Public Health* 2010;101:61–4.
39. Padkin A, Goldfrad C, Brady AR, *et al*. Epidemiology of severe sepsis occurring in the first 24 hrs in intensive care units in England, Wales, and Northern Ireland. *Crit Care Med* 2003;31:2332–8.
40. Whittaker SA, Fuchs BD, Gaieski DF, *et al*. Epidemiology and outcomes in patients with severe sepsis admitted to the hospital wards. *J Crit Care* 2015;30:78–84.
41. Esteban A, Frutos-Vivar F, Ferguson ND, *et al*. Sepsis incidence and outcome: contrasting the intensive care unit with the hospital ward. *Crit Care Med* 2007;35:1284–9.
42. Fillit H, Geldmacher DS, Welter RT, *et al*. Optimizing coding and reimbursement to improve management of Alzheimer's disease and related dementias. *J Am Geriatr Soc* 2002;50:1871–8.
43. Griffiths JA, Gager M, Waldmann C. Follow-up after intensive care. *Contin Educ Anaesth Crit Care Pain* 2004;4:202–5.
44. Pandharipande PP, Girard TD, Jackson JC, *et al*, BRAIN-ICU Study Investigators. Long-term cognitive impairment after critical illness. *N Engl J Med* 2013;369:1306–16.
45. Zielske J, Bohne S, Brunkhorst FM, *et al*. Acute and long-term dysphagia in critically ill patients with severe sepsis: results of a prospective controlled observational study. *Eur Arch Otorhinolaryngol* 2014;271:3085–93.
46. Rothberg MB, Pekow PS, Priya A, *et al*. Variation in diagnostic coding of patients with pneumonia and its association with hospital risk-standardized mortality rates: a cross-sectional analysis. *Ann Intern Med* 2014;160:380–8.
47. Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care* 1994;32:81–90.
48. Gaieski DF, Edwards JM, Kallan MJ, *et al*. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med* 2013;41:1167–74.
49. Iwashyna TJ, Odden A, Rohde J, *et al*. 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:e39–43.