ORIGINAL RESEARCH

Validation of an ICD-Based Algorithm to Identify Sepsis: A Retrospective Study

Shi-Tong Diao [ID], Run Dong, Jin-Min Peng, Yan Chen, Shan Li, Shu-Hua He, Yi-Fan Wang, Bin Du, Li Weng On behalf of the China Critical Care Clinical Trials Group (CCCCTG)

Medical Intensive Care Unit, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing, People's Republic of China

Correspondence: Li Weng, Medical Intensive Care Unit, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, 1 Shuai Fu Yuan, Beijing, 100730, People's Republic of China, Email wengli@gmail.com

Background: Sepsis surveillance was important for resources allocation, prevention, and development of health policy.

Objective: The aim of the study was to validate a modified International Classification of Diseases (ICD)-10 based algorithm for identifying hospitalized patients with sepsis.

Methods: We retrospectively analyzed a prospective, single-center cohort of adult patients who were consecutively admitted to one medical ICU ward and ten non-ICU wards with suspected or confirmed infections during a 6-month period. A modified ICD-10 based algorithm was validated against a reference standard of Sequential Organ Failure Assessment (SOFA) score based on Sepsis-3. Sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and areas under the receiver operating characteristic curves (AUROCs) were calculated for modified ICD-10 criteria, eSOFA criteria, Martin's criteria, and Angus's criteria.

Results: Of the 547 patients in the cohort, 332 (61%) patients met Sepsis-3 criteria and 274 (50%) met modified ICD-10 criteria. In the ICU setting, modified ICD-10 criteria had SE (84.47%), SP (88.57%), PPV (95.60), and NPV (65.96). In non-ICU settings, modified ICD-10 had SE (64.19%), SP (80.00%), PPV (80.33), and NPV (63.72). In the whole cohort, the AUROCs of modified ICD-10 criteria, eSOFA, Angus's criteria, and Martin's criteria were 0.76, 0.75, 0.62, and 0.62, respectively.

Conclusion: This study demonstrated that modified ICD-10 criteria had higher validity compared with Angus's criteria and Martin's criteria. Validity of the modified ICD-10 criteria was similar to eSOFA criteria. Modified ICD-10 algorithm can be used to provide an accurate estimate of population-based sepsis burden of China.

Keywords: sepsis, validation, International Classification of Diseases, administrative health data, identification

Background

As one of the leading causes of death and a major contribution to preventable mortality, sepsis was extrapolated to affect 48.9 million patients and contribute to 11 million deaths each year.[1] Although great efforts had been made to estimate the true burden of sepsis worldwide, most epidemiology studies of sepsis were conducted in high-income countries.[1–4] The scarcity of high-quality data in the rest of the world hampered the accuracy and generalizability of those estimates.[5] Meanwhile, it was estimated that 85.0% of incidences and 84.8% of case fatalities occurred in low- and middle-income countries (LMICs), which highlighted the priorities of sepsis epidemiology research in LMICs, particularly in China, the most populous country in the world.[1]

As per the most recent definition of Sepsis-3, sepsis was defined as acute change in Sequential Organ Failure Assessment (SOFA) score ≥2 points consequent to any infections.[6] However, the components of SOFA excluded the feasibility of sepsis surveillance, especially in LMICs.[7] In 2018, the Centers for Disease Control and Prevention (CDC) developed a hospital toolkit, known as eSOFA, using simplified clinical data obtained from the electronic health records (EHRs) for sepsis surveillance in hospital.[8] However, the limited availability and accessibility of EHRs in LMICs in

China remained a challenge.[9] Thereby, healthcare administrative claims data might be the most feasible way to estimate burden of sepsis in China.

As a common data source, the healthcare administrative claims were widely used for sepsis surveillance.[10] The most cited international classification of diseases (ICD) identification algorithm was the combination of ICD codes for both infection and organ dysfunction.[11] In the era of Sepsis-1, mainstream identification strategies including Angus's criteria and Martin's criteria focused mostly on narrow diagnoses of infection and organ failure. Validation of these algorithms showed a good specificity, but a suboptimal sensitivity using medical record reviews as a gold standard.[11–13] Similar findings were reported in the validation of the Sepsis-3 based algorithm.[7] Nevertheless, a retrospective medical record review was limited to several methodological mistakes,[14] inadequate data collection, inter-rater or intra-rater unreliability, and compromised data abstraction.

The Chinese version of the International Classification of Diseases, Tenth Revision (ICD-10) diagnosis codes was expanded to a 10-digit version from the 4-digit WHO version.[15] There were explicit sepsis cases that had an ICD-10 code that explicitly referenced sepsis, which differed from the traditional criteria.[16] For instance, the category code 008 (3-digit) refers to "complications following ectopic and molar pregnancy", and its secondary category code 008.2 (4-digit) pertains to the specific condition "embolism following ectopic and molar pregnancy", while its tertiary category code 008.200x006 (10-digit) pertains to the particular condition "embolism following abortion and ectopic and molar pregnancy (septicopyaemic)" and was thus labeled as an explicit sepsis code. However, 008.200x006 was not included in the traditional criteria.

Similar to previous studies (Supplementary Table S1), the aim of this study was, therefore, using a prospective cohort of hospitalized patients with SOFA score confirmed sepsis or non-sepsis, to validate a modified Sepsis-3 based ICD identification algorithm in ICU and non-ICU settings, hopefully to provide a validated ICD-based algorithm for sepsis surveillance in China.

Methods
Study Design and Data Collection
This is a retrospective analysis of a single-center, cohort study (NCT 02930070 registered in clinicaltrials.gov), which was designed to assess the diagnostic value of qSOFA for sepsis in a general ward.[17] During the study period from October 1, 2016, to March 31, 2017, all adult patients admitted to one medical ICU ward and ten non-ICU wards with suspected or confirmed infection were eligible for enrollment. Patients were excluded if their age were less than 18 years or hospital length of stay (LOS) less than 24 hours. All patients were followed up until hospital discharge, death, or end of 28-day period, whichever occurred first.

The retrospective analysis was performed based on two databases. We retrieved the following data from the original trial's dataset: (1) Demographic data, chronic comorbidities, information of infections; (2) systolic and diastolic blood pressure, heart rate, saturation of pulse oxygen, respiratory rate, body temperature, urinary output; (3) laboratory results required for SOFA scores, microbiology findings, arterial blood gases; and (4) requirement for critical care resources, ie, ICU admission, respiratory support, vasopressors, and renal replacement therapy. The ICD codes were extracted from the discharge abstract dataset containing ICD-10 diagnosis codes and ICD-9 procedure codes. The protocol was approved by the institutional review board of Peking Union Medical College Hospital, and informed consent was waived (ZS-1142).

Definitions and Sepsis Identification Criteria
Sepsis-3 Criteria
The Third International Consensus Definitions for Sepsis (Sepsis-3) were considered as the gold standard and sepsis was defined as a change in SOFA score ≥2 points consequent to an infection[6] (Supplementary Table S2). Infection was defined according to the predefined criteria (Supplementary Table S3). Baseline SOFA was defined as the estimated SOFA score on the hospital admission day. The baseline SOFA score was assumed to be zero if no history of organ dysfunction was present. The variables required for SOFA score were collected daily, including platelet count, serum bilirubin, serum creatinine, Glasgow Coma Score, saturation of pulse oxygen (SpO_2) or partial pressure of arterial oxygen (PaO_2), Fraction

of Inspired Oxygen (FIO$_2$), blood pressure, and vasopressors. We preferentially used the PaO$_2$ to FIO$_2$ ratio (PaO$_2$/FIO$_2$) when arterial blood gases were obtained. If not available, the PaO$_2$ was estimated from the SpO$_2$.[18]

Modified ICD-10 Criteria

Similar to previous studies,[11,13] our modified ICD-10 identification algorithm was based on a combination of ICD codes for both acute infection (Supplementary Table S4) and organ dysfunction. Organ dysfunction was defined based on organ dysfunction related ICD-10 diagnosis codes and ICD-9 procedure codes (Supplementary Tables S5 and S6). The identification algorithms of Angus and Martin were converted from ICD-9 to ICD-10 codes (Supplementary Tables S7–S10) according to the tool of Centers for Medicare & Medicaid Services.[19]

ESOFA Criteria

Based on the simplified organ dysfunction criteria (eSOFA) proposed by the US Centers for Disease Control and Prevention (CDC), sepsis was identified if there was presumed infection and concurrent eSOFA score ≥1 points.[7,8,20,21] The presumed infection was defined as a medical record of both blood culture and new administration of antibiotics for at least 4 consecutive days. The eSOFA criteria was optimized for electronic health records including: 1) initiation of vasopressors; 2) initiation of mechanical ventilation; 3) doubling of serum creatinine or decrease by 50% of estimated glomerular filtration rate (eGFR) relative to baseline; 4) total bilirubin ≥2.0 mg/dL and increase by 100% from baseline; 5) platelet count <100 cells/μL and ≥50% decline from baseline; and 6) serum lactate ≥2.0 mmol/L (Supplementary Table S2).

Statistical Analysis

Continuous variables were presented as medians (with interquartile range) and compared by Mann–Whitney U-test. Categorical variables were compared by Chi-square test or Fisher's exact test.

The Charlson Comorbidity Index was computed according to Supplementary Table S11. The accuracy of our ICD identification algorithm was evaluated by sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and Area Under the Receiver Operating Characteristic curve (AUROC). Confidence intervals (CI) were calculated by delta method. P-value <0.05 was considered as statistically significant. All statistical analyses were performed by statistical software IBM SPSS Statistics for Windows (IBM Corp. Released 2011, Version 20.0. Armonk, NY: IBM Corp.).

Results

A total of 547 patients were enrolled in the final cohort, including 138 patients in the ICU and 409 patients in the general wards (Figure 1). As shown in Table 1, in the ICU setting, there were 103 (75%) patients meeting Sepsis-3 criteria and 91 (66%) meeting modified ICD-10 criteria. In general wards, there were 229 (56%) patients meeting Sepsis-3 criteria and 183 (45%) meeting modified ICD-10 criteria. Clinical characteristics and outcomes were similar between patients meeting Sepsis-3 and modified ICD-10 criteria in both ICU and non-ICU settings. Compared with non-ICU sepsis patients, ICU patients had higher hospital acquired infections (54% vs 31%, $p<0.01$), blood stream infections (44% vs 20%, $p<0.01$), multiple infection rates (40% vs 24%, $p<0.01$), and expenses (142.06 vs 77.51 1000 CNY, $p<0.01$). Respiratory, blood stream and abdominal infection were the most common infection site.

Comparison of modified ICD-10, Angus's, Martin's, and eSOFA criteria for identification of sepsis were shown in Table 2. For both ICU and non-ICU setting, the sensitivity of modified ICD-10 (70.48%, 95% CI=65.20–75.27%) was higher than Angus's and Martin's criteria (30.12%, 95% CI=25.29–35.42% and 25.60%, 95% CI=21.07–30.72%, respectively), lower than eSOFA (80.12%, 95% CI=75.33–84.20%). On the contrary, the specificity of modified ICD-10 (81.39%, 95% CI=75.41–86.23%) was lower than Angus's and Martin's criteria (93.49%, 95% CI=89.09–96.26% and 98.14%, 95% CI=94.99–99.40%, respectively) and higher than eSOFA (69.77%, 95% CI=0.63–0.76). The corresponding AUROC value of modified ICD-10 criteria was the highest (0.76, 95% CI= 0.72–0.80). The modified ICD-10 criteria in the ICU setting had a higher sensitivity (84.47%, 95% CI=75.70–90.59% vs 64.19%, 95% CI=57.57–70.33%) and specificity (88.57%, 95% CI=72.32–96.27% vs 80.00%, 95% CI=73.67–85.68%) than th non-ICU setting. Positive and negative likelihood ratios of different identification strategies are shown in Supplementary Table S12.
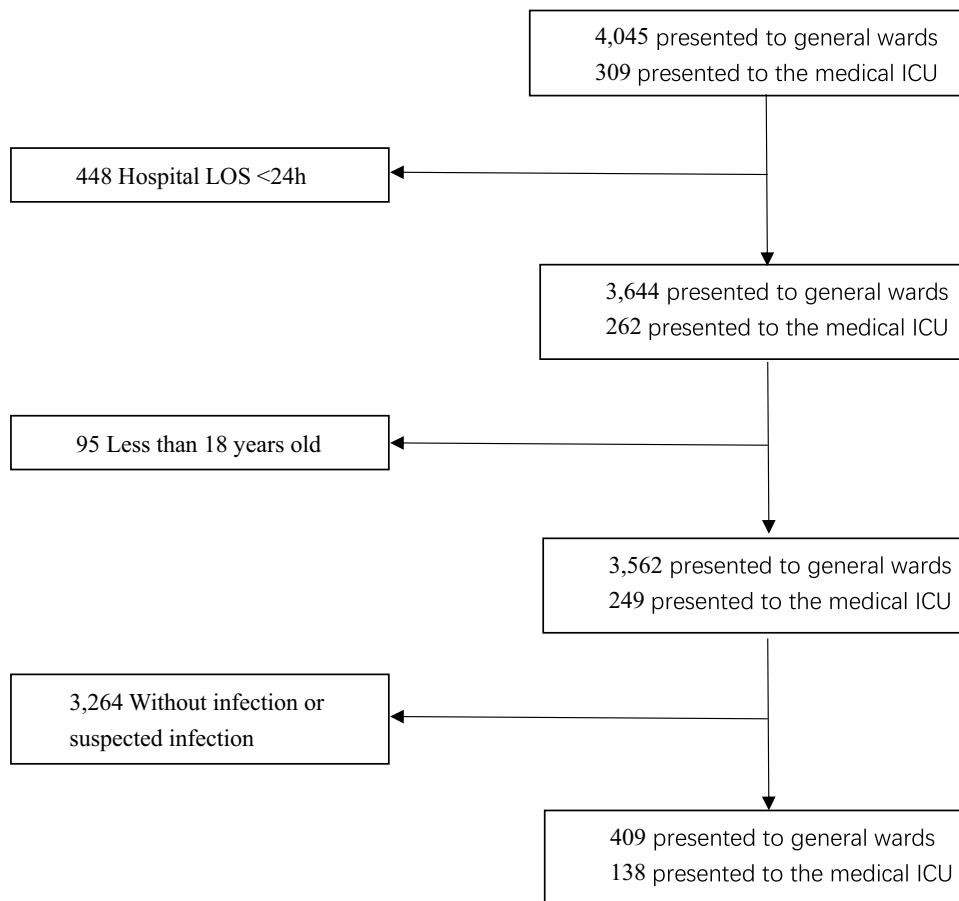
Figure 1 Flowchart of the patients admitted to the non-ICU and ICU.
Abbreviations: LOS, Length of stay; ICU, intensive care unit.

Characteristics Sepsis-3 (+)/ Modified ICD-10 (+), Sepsis-3 (+)/ Modified ICD-10 (-), Sepsis-3 (-)/ Modified ICD-10 (+), and Sepsis-3 (-)/ Modified ICD-10 (-) Patients

There were 234 (43%) patients who met both Sepsis-3 and modified ICD-10 criteria. In comparison, both the Sepsis-3 (+)/modified ICD-10(-) (n=98) and the Sepsis-3 (-)/modified ICD-10 (+) (N=40) (shown in Table 3) groups were less likely to have respiratory, blood stream, and two or more infections. Also, the SOFA score, qSOFA, eSOFA score, proportion of mechanical ventilation support and vasoactive agents, and mortality rates in both of the groups were significantly lower than the Sepsis-3+/modified ICD-10+ group. Among 98 Sepsis-3 (+)/modified ICD-10 (-) patients, 29 (30%) patients had abdominal infection and 43 (44%) patients had coagulation dysfunction. 82 of 98 Sepsis-3 (+)/modified ICD-10 (-) patients were not identified as sepsis by the modified ICD-10 algorithm because organ dysfunction ICD codes were not assigned.

Discussion

Reliable surveillance of sepsis was important for the estimation of disease burden, prevention initiatives. and national sepsis quality measures.[22] In this study, we validated a sepsis surveillance algorithm based on ICD-10 codes. Our findings showed that the modified ICD-10 based algorithm could identify more than 95% of sepsis patients in the ICU and more than 80% of sepsis patients outside the ICU in a single medical center of China. Considering the availability and accessibility of ICD codes in China,[23] a modified ICD-10 based algorithm was a feasible way for population-based sepsis surveillance in China.

Table 1 Clinical Characteristics and Outcomes of Patients Enrolled in ICU and Non-ICU Settings

	ICU (n=138)			Non-ICU (n=409)		
	Modified ICD-10 (n=91)	Sepsis 3.0 (n=103)	p-value	Modified ICD-10 (n=183)	Sepsis 3.0 (n=229)	p-value
Male, n (%)	44 (48)	51 (50)	0.872	100 (55)	124 (54)	0.920
Age, median (IQR)	58 (39–69)	60 (46–69)	0.475	57 (37–68)	58 (43–68)	0.580
Charlson comorbidity index, n (%)						
0	18 (20)	20 (19)	0.949	56 (30)	73 (32)	0.781
1	25 (27)	26 (25)	0.725	30 (16)	36 (16)	0.853
2	23 (25)	28 (27)	0.763	57 (31)	66 (29)	0.608
3 or more	25 (27)	29 (28)	0.916	40 (22)	54 (24)	0.679
Emergency admission, n (%)	37 (41)	42 (41)	0.987	60 (38)	76 (33)	0.931
Perioperative admission, n (%)	10 (11)	12 (12)	0.885	31 (17)	39 (17)	0.981
Infected site, n (%)						
Blood stream	41 (45)	45 (44)	0.848	34 (19)	46 (20)	0.701
Respiratory	70 (77)	80 (78)	0.901	118 (64)	145 (63)	0.922
Abdominal	18 (20)	17 (17)	0.554	33 (18)	53 (23)	0.205
Urinary	4 (4)	5 (5)	0.868	19 (10)	19 (8)	0.467
Skin and soft tissue	7 (8)	10 (1)	0.605	14 (8)	14 (6)	0.538
Neurological	4 (4)	3 (3)	0.590	7 (4)	11 (5)	0.629
Two or more infections	40 (44)	41 (40)	0.590	42 (23)	56 (24)	0.722
Nosocomial infection, n (%)	51 (56)	56 (54)	0.815	55 (30)	71 (31)	0.835
qSOFA≥2	73 (80)	82 (80)	0.996	100 (55)	122 (53)	0.782
SOFA, median (IQR)	8 (5–11)	7 (5–11)	0.835	4 (3–6)	4 (3–6)	0.890
eSOFA, median (IQR)	2 (1–3)	2 (1–3)	0.463	1 (0–2)	0 (1–2)	0.569
Organ support, n (%)						
Renal replacement therapy	16 (18)	20 (19)	0.743	7 (4)	7 (3)	0.669
Mechanical ventilation	63 (69)	68 (66)	0.634	24 (13)	30 (13)	0.997
Vasoactive agent	54 (59)	56 (54)	0.486	38 (21)	44 (19)	0.695
Outcomes						
Hospital LOS, median (IQR)	21 (11–37)	21 (11–37)	0.876	24 (14–33)	24 (13–32.5)	0.766
Hospital mortality, n (%)	22 (24)	23 (22)	0.761	31 (17)	32 (14)	0.406
28-day mortality, n (%)	15 (16)	15 (15)	0.712	29 (16)	30 (13)	0.429
Total cost, 1000 CNY (mean±SD)	149.98±150.19	142.06±144.75	0.711	82.26±91.25	77.51±87.28)	0.594

Abbreviations: SOFA, Sequential Organ Failure Assessment; eSOFA, Simplified Organ Dysfunction Criteria; IQR, inter quartile range; LOS, length of stay; CNY, Chinese yuan renminbi; SD, standard deviation.

Recent published studies have demonstrated that EHRs-based criteria were more accurate than administrative claims in identifying patients with sepsis.[24–27] EHRs-based criteria had several advantages such as objectivity, quantization, and real-time monitoring for hospitalized patients.[28] CDC developed a Toolkit based on clinical data directly obtained from the EHRs, eSOFA, to track hospital-level sepsis incidence and outcome.[8,21] In our study, a high eSOFA diagnosis proportion in the Sepsis-3 (+)/modified ICD-10 (-) group showed the validity of EHRs. Although our previous validation study confirmed the performance of the toolkit in a hospitalized Chinese population, the feasibility to use the CDC sepsis surveillance toolkit was limited in most resource-limited areas where EHRs were not available or accessible.[29] Administrative claims based modified ICD-10 criteria offered a feasible low-cost approach to extrapolate population-based estimates of sepsis burden in LMICs, where most of the septic patients were admitted in non-ICU settings.[30] In addition, our study showed the overall accuracy in sepsis identification was similar between the EHRs algorithm and modified ICD-10 algorithm. Even so, most patients were not identified by modified ICD-10 criteria because of the absence of the corresponding organ dysfunction codes, which could be explained by insufficient coding training. With improving coding practices and more standardized coding procedure,[7] modified ICD-10 criteria might have better performance in sepsis surveillance.

Table 2 Comparison of Different ICD-Coded Algorithms and eSOFA Criteria for Identification of Sepsis in ICU and Non-ICU Settings

	SP, % (95% CI)	SE, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	AUC (95% CI)
No-ICU (n=409)					
Modified ICD-10	80.00 (73.67–85.68)	64.19 (57.57–70.33)	80.33 (73.67–85.68)	63.72 (57.04–69.92)	0.72 (0.67–0.77)
Angus's criteria	92.22 (87.04–95.52)	31.44 (25.57–37.94)	83.72 (73.85–90.50)	51.39 (45.80–56.95)	0.62 (0.56–0.67)
Martin's criteria	97.78 (94.04–99.29)	21.40 (16.39–27.39)	92.45 (80.93–97.55)	49.44 (44.14–54.75)	0.60 (0.54–0.65)
eSOFA criteria	68.33 (60.93–74.94)	76.86 (70.75–82.05)	75.54 (69.41–80.81)	69.89 (62.45–76.44)	0.73 (0.68–0.78)
ICU (n=138)					
Modified ICD-10	88.57 (72.32–96.27)	84.47 (75.70–90.59)	95.60 (88.50–98.58)	65.96 (50.60–78.72)	0.87 (0.79–0.94)
Angus's criteria	100.00 (87.68–100.00)	27.18 (19.11–36.99)	100.00 (84.98–100.00)	31.82 (23.45–41.48)	0.64 (0.54–0.73)
Martin's criteria	100.00 (87.68–100.00)	34.95 (26.00–45.05)	100.00 (87.99–100.00)	34.31 (25.38–44.45)	0.67 (0.58–0.76)
eSOFA criteria	77.14 (59.44–88.96)	87.38 (79.03–92.84)	91.84 (84.08–96.15)	67.50 (50.76–80.93)	0.82 (0.73–0.91)
ICU and non-ICU (n=547)					
Modified ICD-10	81.39 (75.41–86.23)	70.48 (65.20–75.27)	85.40 (80.53–89.25)	64.10 (58.07–69.74)	0.76 (0.72–0.80)
Angus's criteria	93.49 (89.09–96.26)	30.12 (25.29–35.42)	87.72 (79.93–92.88)	46.42 (41.66–51.24)	0.6 2(0.57–0.66)
Martin's criteria	98.14 (94.99–99.40)	25.60 (21.07–30.72)	95.51 (88.26–98.56)	45.07 (41.45–50.76)	0.62 (0.57–0.66)
eSOFA criteria	69.77 (63.08–75.73)	80.12 (75.33–84.20)	80.36 (75.58–84.42)	69.44 (62.77–75.42)	0.75 (0.71–0.79)

Abbreviations: SP, specificity; SE, sensitivity; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; AUC, area under curve.

Table 3 Characteristics of Infected Patients With or Without Sepsis Defined by Sepsis-3 or Modified ICD-10 Criteria

	Sepsis 3.0 (+)/Modified ICD-10 (+) (N=234)	Sepsis 3.0 (–)/Modified ICD-10 (+) (N=40)	Sepsis 3.0 (+)/Modified ICD-10 (-) (N=98)	Sepsis 3.0 (–)/Modified ICD-10 (-) (N=175)
Male, n (%)	120 (51)	24 (60)	55 (56)	91 (52)
Age, median (IQR)	58 (38.75–68)	45.5 (34.5–68.5)**‡	60 (47.75–69)	54 (38–66)
Charlson comorbidity index, n (%)				
0	62 (27)	12 (30)†	31 (31)†	83 (47)
1	48 (21)	7 (18)	14 (14)	20 (11)
2	67 (29)	13 (33)	27 (25)	35 (20)
3 or more	57 (24)	8 (20)	26 (27)	48 (27)
Emergency admission, n (%)	100 (43)	18 (45)†	26 (27)**	46 (26)
Surgical treatment, n (%)	33 (14)	8 (20)	17 (17)	44 (25)
Infected site, n (%)				
Blood stream	72 (31)	3 (8) **	15 (15)**	15 (9)
Respiratory	169 (72)	19 (48) **	46 (47)**	81 (46)
Abdominal	40 (17)	11 (28)	29 (30)**	40 (23)
Urinary	17 (7)	6 (15)	6 (6)	11 (6)
Skin and soft tissue	17 (7)	4(10)	4 (4)	15 (9)
Neurological	10 (4)	1 (3)	5 (5)	5 (3)
Two or more infections	76 (32)	6 (15)*	20 (20)*†	19 (11)
Hospital LOS, median (IQR)	24 (12–37)	20.5 (15–30)	23 (13–30)	17 (10–27)
Healthcare-associated infection, n (%)	93 (40)	13 (33)	34 (35)†	40 (23)
qSOFA≥2, n (%)	164 (70)	7 (18)**	40 (41)**‡	24 (14)
SOFA, median (IQR)	6 (4–9)	1 (1–3)**	3 (2–5)**‡	1 (0–2)
eSOFA, median (IQR)	2 (1–3)	0 (0–1)**	1 (1–2)**‡	0(0–1)
Cardiovascular, n (%)	91 (39)	1 (3)**	9 (9)**†	5 (2)
Respiratory, n (%)	84 (36)	3 (8)**	14 (14)**	14 (8)
Renal, n (%)	56 (24)	2 (5)**	14 (14)†	10 (6)
Hepatic, n (%)	54 (23)	4 (10)*	26 (27)‡	18 (10)
Coagulation, n (%)	82 (35)	9 (23)	43 (44)‡	21 (12)
Lactate, n (%)	37 (16)	1 (3)*	3 (3)*	9 (5)

Notes: *p<0.05, **p<0.01, compared with Modified ICD-10+ Sepsis 3.0+. †p<0.05, ‡p<0.01, compared with Modified ICD-10- Sepsis 3.0-.
Abbreviations: IQR, interquartile range; LOS length of stay.

A major strength of our study was the use of a prospective cohort which could accurately identify patients with sepsis and collect all components of SOFA score and other detailed clinical data. In 2019, Rhee et al[26] evaluated the accuracy of sepsis identification strategy based on claims data and clinical data from EHRs. The sensitivity of administrative claims strategy ranged from 5% to 54% for explicit sepsis codes, and from 42% to 80% for implicit codes. However, the robustness of the findings in the study was limited to the retrospectively confirmed diagnosis of sepsis, especially in non-ICU settings. Also, previous studies[7,24–26] validating administrative data algorithms were based on retrospective cohorts which were inevitably influenced by missing data including Glasgow score, laboratory findings, and urinary output. Another strength of our study was the inclusion of patients with sepsis admitted to non-ICU settings,[31] allowing generalization of our findings to the hospitalized populations.

The study also had several limitations. First, our study was based on a cohort from a single center. Although the ICD coding practice might be different in hospitals across different regions, the data qualities of ICD coding were ensured by manual review (Supplementary Figure S1), quality control meetings, courses of staff training, and regulations for inspection in all public hospitals in China.[32] Further validation was warranted for generalization of our algorithms in other hospitals in China. Second, we could not differentiate sepsis-related or non-sepsis related organ dysfunctions based on the ICD codes, which would lead to an overestimation of the sepsis rate. Lastly, we only enrolled patients with infection or suspected infection. Further investigation was needed for population with extremely low incidence of infection.[24,33]

Conclusions

This study demonstrated that modified ICD-10 criteria had higher validity compared with Angus's criteria and Martin's criteria. Validity of the modified ICD-10 criteria was similar to eSOFA criteria. Modified ICD-10 algorithm can be used to provide an accurate estimate of population-based sepsis burden of China.

Data Sharing Statement

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Ethical Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the institutional review board of Peking Union Medical College Hospital (ZS-1142). The data accessed complied with relevant data protection and privacy regulations.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis, and interpretation; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet*. 2020;395(10219):200–211. doi:10.1016/S0140-6736(19)32989-7

2. Rhee C, Wang R, Zhang Z, Fram D, Kadri SS, Klompas M. Epidemiology of hospital-onset versus community-onset sepsis in US hospitals and association with mortality: a retrospective analysis using electronic clinical data. *Crit Care Med*. 2019;47(9):1169–1176. doi:10.1097/CCM.0000000000003817

3. Walkey AJ, Lagu T, Lindenauer PK. Trends in sepsis and infection sources in the United States. A population-based study. *Ann Am Thorac Soc*. 2015;12(2):216–220. doi:10.1513/AnnalsATS.201411-498BC

4. Wang H, Devereaux R, Yealy D, Safford M, Howard G. National variation in United States sepsis mortality: a descriptive study. *Int J Health Geogr*. 2010;9(1):1–9. doi:10.1186/1476-072X-9-9

5. World Health Organization. *Global Report on the Epidemiology and Burden of Sepsis: Current Evidence, Identifying Gaps and Future Directions*. Geneva: World Health Organization; 2020.

6. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315(8):801–810. doi:10.1001/jama.2016.0287

7. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. *JAMA*. 2017;318(13):1241–1249. doi:10.1001/jama.2017.13836

8. Centers for Disease Control and Prevention. Hospital toolkit for adult sepsis surveillance; 2018 [cited February 7, 2022]. Available from: https://www.cdc.gov/sepsis/pdfs/Sepsis-Surveillance-Toolkit-Mar-2018_508.pdf. Accessed October 28, 2023.

9. Lin H, Tang X, Shen P, et al. Using big data to improve cardiovascular care and outcomes in China: a protocol for the Chinese Electronic health Records Research in Yinzhou (CHERRY) Study. *BMJ Open*. 2018;8(2):e019698. doi:10.1136/bmjopen-2017-019698

10. Whittaker SA, Mikkelsen ME, Gaieski DF, Koshy S, Kean C, Fuchs BD. Severe sepsis cohorts derived from claims-based strategies appear to be biased toward a more severely ill patient population. *Crit Care Med*. 2013;41(4):945–953. doi:10.1097/CCM.0b013e31827466fl

11. 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. *Crit Care Med*. 2001;29(7):1303–1310. doi:10.1097/00003246-200107000-00002

12. Wang HE, Addis DR, Donnelly JP, et al. Discharge diagnoses versus medical record review in the identification of community-acquired sepsis. *Crit Care*. 2015;19(1):42. doi:10.1186/s13054-015-0771-6

13. Jolley RJ, Quan H, Jette N, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open*. 2015;5(12):e009487. doi:10.1136/bmjopen-2015-009487

14. Vassar M, Holzmann M. The retrospective chart review: important methodological considerations. *J Educ Eval Health Prof*. 2013;10:12. doi:10.3352/jeehp.2013.10.12

15. Wang Q, Ji Z, Wang J, et al. A study of entity-linking methods for normalizing Chinese diagnosis and procedure terms to ICD codes. *J Biomed Inform*. 2020;105:103418. doi:10.1016/j.jbi.2020.103418

16. Guo DC, Zhuang HZ, Lin J, Zhi DY, Duan ML. Epidemiology of sepsis in Beijing from 2012 to 2018: analysis of hospital homepage databases derived from the Beijing Public Health System. *BMC Public Health*. 2022;22(1):2237. doi:10.1186/s12889-022-14725-1

17. Luo J, Jiang W, Weng L, et al. Usefulness of qSOFA and SIRS scores for detection of incipient sepsis in general ward patients: a prospective cohort study. *J Crit Care*. 2019;51:13–18. doi:10.1016/j.jcrc.2019.01.012

18. Ellis RK. Determination of PO2 from saturation. *J Appl Physiol*. 1989;67(2):902. doi:10.1152/jappl.1989.67.2.902

19. Centers for Medicare & Medicaid Services. General equivalence mappings; 2016 [cited February 7, 2022]. Available from: https://www.cms.gov/Medicare/Coding/ICD10/2016-ICD-10-CM-and-GEMs. Accessed October 28, 2023.

20. Dong R, Tian H, Zhou J, et al. External validity of Adult Sepsis Event's simplified eSOFA criteria: a retrospective analysis of patients with confirmed infection in China. *Ann Intensive Care*. 2020;10(1):14. doi:10.1186/s13613-020-0629-1

21. Rhee C, Zhang Z, Kadri SS, et al. Sepsis surveillance using adult sepsis events simplified eSOFA criteria versus sepsis-3 sequential organ failure assessment criteria. *Crit Care Med*. 2019;47(3):307–314. doi:10.1097/CCM.0000000000003521

22. Curren E, Lutgring J, Kabbani S, et al. Advancing diagnostic stewardship for healthcare-associated infections, antibiotic resistance, and sepsis. *Clin Infect Dis*. 2022;74(4):723–728. doi:10.1093/cid/ciab672

23. Cao L, Gu D, Ni Y, Xie G. Automatic ICD code assignment based on ICD's Hierarchy Structure for Chinese Electronic Medical Records. *AMIA Jt Summits Transl Sci Proc*. 2019;2019:417-424. PMID: 31258995

24. Liu YZ, Chu R, Lee A, et al. A surveillance method to identify patients with sepsis from electronic health records in Hong Kong: a single centre retrospective study. *BMC Infect Dis*. 2020;20(1):652. doi:10.1186/s12879-020-05330-x

25. Valik JK, Ward L, Tanushi H, et al. Validation of automated sepsis surveillance based on the Sepsis-3 clinical criteria against physician record review in a general hospital population: observational study using electronic health records data. *BMJ Qual Saf*. 2020;29(9):735–745. doi:10.1136/bmjqs-2019-010123

26. Rhee C, Jentzsch MS, Kadri SS, et al. Variation in identifying sepsis and organ dysfunction using administrative versus electronic clinical data and impact on hospital outcome comparisons. *Crit Care Med*. 2019;47(4):493–500. doi:10.1097/CCM.0000000000003554

27. Umberger R, Indranoi CY, Simpson M, Jensen R, Shamiyeh J, Yende S. Enhanced screening and research data collection via automated EHR data capture and early identification of sepsis. *SAGE Open Nurs*. 2019;5:2377960819850972. doi:10.1177/2377960819850972

28. Shappell CN, Rhee C. Leveraging electronic health record data to improve sepsis surveillance. *BMJ Qual Saf*. 2020;29(9):706–710. doi:10.1136/bmjqs-2020-010847

29. Liang J, Li Y, Zhang Z, et al. Adoption of Electronic Health Records (EHRs) in China during the past 10 Years: consecutive Survey Data Analysis and Comparison of Sino-American Challenges and Experiences. *J Med Internet Res*. 2021;23(2):e24813. doi:10.2196/24813

30. Zhou J, Tian H, Du X, et al. Population-based epidemiology of sepsis in a subdistrict of Beijing. *Crit Care Med*. 2017;45(7):1168–1176. doi:10.1097/CCM.0000000000002414

31. Whittaker S-A, Fuchs BD, Gaieski DF, et al. Epidemiology and outcomes in patients with severe sepsis admitted to the hospital wards. *J Crit Care*. 2015;30(1):78–84. doi:10.1016/j.jcrc.2014.07.012

32. Tian DC, Zhang C, Yuan M, et al. Incidence of multiple sclerosis in China: a nationwide hospital-based study. *Lancet Region Health Western Pacific*. 2020;1:100010. doi:10.1016/j.lanwpc.2020.100010

33. 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. *Surg Infect*. 2014;15(5):513–516. doi:10.1089/sur.2013.089