Impact of hourly serial SOFA score on signaling emerging sepsis

Bin-Shenq Ho, Yan-Hwa Wu Lee, Yi-Bing Lin *  
National Yang Ming Chiao Tung University, Hsinchu City, 300093, Taiwan, ROC

ARTICLE INFO

Keywords:  
Artificial intelligence (AI)  
Clinical information system  
Data intelligence  
Sepsis  
Sequential organ failure assessment (SOFA)

ABSTRACT

Background: Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. Under suspicion of infection, sepsis can be identified as an acute increase in Sequential Organ Failure Assessment (SOFA) score of 2 points or more. Professional critical care societies have called for early detection and treatment of sepsis; however, the fundamental tool to address the need remains unmet.

Objectives: The present study aims at exploring the possibility of a solution to bridging clinical information system with AI medicine and supporting decision-based medical tasks through integrated data-intensive intelligence.

Patients and methods: We extracted data from a well-recognized database and explored the feasibility of a real-time solution to sepsis identification for adult ICU patients under suspicion of infection. To analogize the requirement of a randomized controlled trial, we adopted propensity score matching to tackle the imbalance of baseline covariates between study groups frequently encountered in observational studies.

Results: Our study indicates that the hourly assessment protocol outperforms the 24-h assessment counterpart in terms of the timing of sepsis identification by a median of 14.5 h earlier and the change of total SOFA score by a median of 1.0 point lower.

Conclusions: We conclude that real-time SOFA score to signal emerging sepsis becomes feasible through the introduction of data-intensive intelligence and data processing technologies.

1. Introduction

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection (Sepsis-3) [1–3]. Organ dysfunction can be identified as an acute increase in Sequential Organ Failure Assessment (SOFA) score of 2 points or more consequent to infection, which is associated with mortality over 10%. In 2017, an estimated 48.9 million incident cases of sepsis were recorded worldwide and 11.0 million sepsis-related deaths were reported, representing 19.7% of all global deaths [4]. Sepsis is the primary cause of death from infection, especially if not recognized and treated promptly.

Lower risk-adjusted in-hospital mortality could be achieved by more rapid completion of a 3-h bundle of sepsis care and rapid administration of antibiotics [5]. According to a multicenter study, hourly delays in antibiotic administration were associated with increased odds of hospital mortality among patients with sepsis hospitalized through the emergency department, even in patients who received antibiotics within 6 h [6]. Moreover, each hour of delay in initiation of effective antimicrobial therapy was associated with a mean decrease in survival of 7.6% over the first 6 h after the onset of recurrent or persistent hypotension [7]. It was recommended that administration of intravenous antibiotics be initiated as soon as possible within 1 h upon recognition of sepsis [8]. Early identification and antibiotic treatment of sepsis improve the prognosis of patients.

Leveraging clinical data accrued in real time is promising in artificial intelligence (AI) medicine. However, diagnosis of sepsis is not always knowable in real time. During times of diagnostic uncertainty, the timing and broadness of initial antibiotics vary with the likelihood of bacterial infection and the illness severity, and the optimal treatment may differ across patients [9]. A change in total SOFA score of 2 or more is now a defining attribute of sepsis syndrome [10]. The requirement to detect modest serial changes in a patient’s SOFA score was therefore advocated. However, although professional critical care societies have proposed new clinical criteria that aid sepsis recognition by Sepsis-3, the essential need for early detection and treatment remains challenging [11]. Furthermore, the methodologies involved in a retrospective clinical study usually face systematic differences in baseline covariates, which could potentially bias research results.

With the advent of the digital revolution, the ever-growing data collected during day-to-day healthcare services provide a challenging
opportunity for the development of real-time intelligence. AI could leverage these data and provide solutions to assist physicians in treating their patients. This paper aims at supporting decision-based medical tasks through data-intensive intelligence. We integrate demographic, clinical, laboratory, prescription, microbiological, and timing factors to investigate early recognition of sepsis. Through exploitation of heterogeneous data sources and critical adoption of data processing methodologies, our study explores the possibility of a solution to bridging clinical information system with AI medicine, notably in terms of data dimension transformation and real-time resolution, and to tackling distribution balance issues of baseline covariates inevitably encountered in an observational study between study groups.

2. Patients and methods

2.1. Database

Data were extracted from the Medical Information Mart for Intensive Care database (MIMIC-III, version 1.4), which is supported by grants from the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health and is freely available to the medical research community [12-14]. MIMIC-III comprises deidentified health data associated with totally 46,476 patients and 61,532 intensive care unit (ICU) admissions at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, USA, between 2001 and 2012. The installation of MIMIC-III was approved by the Institutional Review Boards of the Beth Israel Deaconess Medical Center (Boston, MA) and Massachusetts Institute of Technology (Cambridge, MA). The first author (BIN-SHENQ HO, certification record ID 42420948) has passed a recognized course and been approved for accessing data from MIMIC-III for research since May 11, 2021.

2.2. Cohort selection

This is a retrospective cohort study. We selected all adult ICU patients aged between 18 and 89 years old on admission. To track the serial SOFA scores for at least 3 days, patients with ICU length of stay less than 72 h were excluded. Suspicion of infection was defined by using both prescription and microbiology events, i.e., culture done within 72 h before or 24 h after antibiotic was given [2]. As an inclusion criterion, a change in total SOFA score of less than 2 over the first 24 h after ICU admission was required. Additionally, to ensure the independence of data, two patients were excluded because sepsis and non-sepsis were respectively recognized based on ICD-9-CM diagnosis codes during different ICU admissions. Finally, for the patient who had multiple ICU admissions, only the leading admission was adopted for analysis. Fig. 1 shows the patient selection flowchart. Totally, 45,301 patients and 59,674 ICU admissions were excluded, and 1166 patients with one ICU admission for each were included for the present study. According to the way to identify sepsis, the present study comprised two parts, namely Study A and Study B, as below.

![Patient inclusion flowchart](image)

- Suspicion of infection: culture done within 72 h before or 24 h after antibiotic given.
- Four ICU admissions: sepsis and non-sepsis respectively recognized based on ICD-9-CM diagnosis codes during different ICU admissions of the same patients; eight ICU admissions: not eligible due to not the leading admissions.
- Sepsis identified as an increase in SOFA score of 2 points or more under suspicion of infection based on assessments at 24-h intervals for Study A; sepsis identified as an increase in SOFA score of 2 points or more under suspicion of infection based on hourly assessment over a sliding 24-h window for Study B.
2.3. Sepsis identification protocols

For Study A, sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on assessments at 24-h intervals. For Study B, sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on hourly assessment over a sliding 24-h window. SOFA score consists of six sub-scores, namely respiration (PaO2/FIO2), coagulation (platelets), liver (bilirubin), cardiovascular (hypotension), central nervous system (Glasgow Coma Scale), and renal (creatinine or urine output) [15]. The time-stamped SOFA score was calculated based on the most severe value of each sub-score for the period spanning the preceding 24 h. For a single sub-score value, if the underlying data was missing, it was assumed null, i.e. 0, for the calculation [16,17].

During day-to-day clinical practice, the components of each sub-score and subsequently total SOFA score are not necessarily measured hourly as well as at the same time point. Here we were faced with data asynchrony. To break through the predicament, we aligned multiple time series pertaining to SOFA score calculation of a patient since ICU admission and constructed a sliding 24-h window. As the prevailing clinical practice required, the score is calculated on admission and every 24–48 h thereafter using the most severe values measured during the prior 24 h. Study A therefore adopted total SOFA score based on assessments at 24-h intervals to feature the state-of-the-art clinical practice in terms of sepsis identification. Furthermore, to explore the feasibility of potentially embedded intelligence, Study B calculated total SOFA score based on hourly assessment for the period spanning the preceding 24 h.

2.4. Variable extraction

Clinical known confounders were extracted from the database for the initial 24 h as baseline characteristics, which included patient demographics, race, body mass index, vital signs, laboratory results, and total SOFA score. To address the disease severity on the first day, Simplified Acute Physiology Score II (SAPS II) was queried from the database [18]. To quantify the level of the inflammatory response of the body on the first day, the score of Systemic Inflammatory Response Syndrome (SIRS) criteria was calculated [19].

To balance the comorbidities during the hospitalization, Quan Enhanced ICD-9-CM Elixhauser Comorbidity measures were further derived from the database [20]. We prespecified the cutoff for an acceptable data missing percentage [21–23], and only variables with a missing rate of 5% or less were included for analysis. We calculated SAPS II, SIRS, and Elixhauser Comorbidity measures using the PostgreSQL codes that were available in the MIMIC code repository [17].

2.5. Propensity score matching

As patients do not randomly develop sepsis, there was a significant imbalance in baseline covariate distributions between patients in the sepsis group and the non-sepsis group. Propensity analysis then proceeded to estimate the propensity score of every patient [24]. The propensity score is defined as the conditional probability of assignment to a group given a vector of observed covariates [25,26]. We constructed models to estimate the propensity score of every patient to be identified as sepsis with observed covariates. Missing data were imputed by the k-nearest neighbor (KNN) algorithm [27]. Kolmogorov-Smirnov test for goodness of fit justified that the KNN-imputed distribution was nearly identical to the original sample distribution in the present study.

Several approaches, including logistic regression, artificial neural network, recursive partitioning and regression trees, and random forest, were experimented to build the propensity score model [28,29]. For the present study, the propensity score model was constructed to analogize the requirement of a randomized controlled trial. Accordingly, based on standardized mean difference (SMD) [26,30], we prespecified its threshold at 0.1 to examine how many covariates could be well balanced in a candidate model. With a series of hyperparameters tested, we explored candidate algorithms, and the representative results were summarized in Table 1. By virtue of the number of covariates to be well balanced for Study A and Study B, respectively, logistic regression and nearest neighbor matching with replacement by a ratio of 1:2 between the sepsis and non-sepsis groups outperformed all the other models that we probed for the present study and was consequently adopted to build the propensity score model. Propensity score estimation and nearest-neighbor matching were performed with Matchit package [31], Breiman’s random forest algorithm was implemented using randomForest package [32], and covariate balance was assessed with cobalt package [33].

2.6. Survival analysis and sepsis identification analysis

We conducted the Kaplan-Meier survival analysis of the non-sepsis and sepsis groups for Study A and Study B, respectively. Survival analysis was performed before and after propensity score matching. For adult patients admitted to ICU, the 30-day mortality was predominantly determined by the characteristics of the acute illness, and the 90-day mortality was mainly determined by age and comorbidity [34]. Therefore, we followed the survival status of the study cohort up to 90 days after ICU admission.

A major goal of the present study is to explore the possibility of identifying emerging sepsis in time. Consequently, we extracted the time of sepsis identification and the concomitant total SOFA score change according to the 24-h assessment protocol in Study A and the hourly assessment protocol in Study B, respectively. The measurement difference between the two protocols was calculated and tested for statistical significance.

2.7. Statistical analysis

For the univariate analysis, as assumption of normality checked by Kolmogorov-Smirnov test showed that continuous variables were not normally distributed, medians with interquartile ranges were used to summarize distributions. Categorical variables were shown as proportions. The Wilcoxon rank-sum test and chi-squared test were used to statistically assess the differences in continuous and categorical variables, respectively. The Wilcoxon signed-rank test was applied for paired difference test. The Kaplan-Meier method was carried out to conduct survival analysis before and after propensity score matching [35,36]. The log-rank test was conducted to compare the survival distributions of the sepsis and non-sepsis groups. The generalized linear model was constructed to predict the odds of 90-day mortality after ICU admission.

Covariate distributions between the sepsis and non-sepsis groups were checked with SMD as well as variance ratio (VR) [26,30]. We prespecified the threshold for SMD at 0.1 [37]. For a covariate with SMD between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation. VR close to 1 would indicate good balance between 0.1 and 0.25 [38], VR was used to rule out extreme imbalance for the equivocal situation.
whether the two assessment protocols differentially detected sepsis in adult patients once admitted to ICU. To address external validity, we lifted the restrictions that the ICU length of stay was at least 72 h and the change of total SOFA score was less than 2 over the first 24 h. Otherwise, the cohort selection followed the same process as illustrated in Fig. 1 and all the other methodologies stated in the previous subsections were applied accordingly.

### 2.9. AIoT and IoT computation environment

Fig. 2 illustrates the real-time Internet of Things (IoT) and Artificial Intelligence of Things (AIoT) computation environment based on the guidelines provided by Chang and Lin [41]. For AI training, the data of the MMIC-III database are downloaded and accessed as an instance running on PostgreSQL (version 13.4, https://www.postgresql.org/). The data intelligence platform on a local computer can conduct the training and testing through the algorithms stated in the previous subsections. We use R (version 4.0.5, https://www.r-project.org/) and Python (version 3.8.5, https://www.python.org/) for statistical analysis [42,43], both running on Jupyter Notebook (server version 6.1.4, https://www.jupyter.org/), distributed by Anaconda (Anaconda3-2020.11).

For AI inference, the real-time AIoT components include three major components. The medical IoT devices in the ICU collect patients’ information, including biophysiological signals, laboratory profiles, and medication records, which are sent to the IoT platform called IoTtalk [44]. The IoTtalk server can be installed in a local Wi-Fi AP. In this scenario, the CPU utilization in the local server is about 40% for each core under a dual-core processor (Intel i3-4005U) and the memory utilization is 689 MB of 1.99 GB. The IoTtalk server can also be installed in a VM in a commercial cloud at Chunghwa Telecom (CHT), and its location is remote from the ICU. The CPU utilization in the CHT cloud is about 50% for each core under a quad-core processor VM and the memory utilization is 795 MB of 1.95 GB. The IoTtalk server can preprocess the collected data as described above and then send them to the QOCA AI Platform [41]. The trained AI and R models at the data intelligence platform can be rebuilt in the QOCA AI platform that auto-mates the AI tools just like DataRobot, RapidMiner, H2O.ai, AWS SageMaker, and Azure AutoML.

Through measurement experiments [44], the histograms for the delay from an IoT device to the QOCA AI platform are illustrated in Fig. 3. In the local scenario, the expected delay is $E[t_L] = 31.68$ ms. In the remote scenario, the expected delay is $E[t_R] = 59.14$ ms. It is clear that our AIoT platform can receive the real-time data and provide the inference results immediately.

We note that the training and the testing phases of AI modeling can be conducted directly in the QOCA AI platform that automates the AI tools just like DataRobot, RapidMiner, H2O.ai, AWS SageMaker, and Azure AutoML.

### Table 1

| Model | Replacement | Ratio (Sepsis to Non-sepsis) | Study A Cohorta | Study B Cohortb |
|-------|-------------|------------------------------|------------------|------------------|
|       |             | n = 1166                     | v = 23c          | n = 1166         |
|       |             | Matched                      | SMD < 0.1        | Matched          | SMD < 0.1        |
| LR    | no          | 1:1                          | 1000             | 8                | 814              |
| LR    | yes         | 1:1                          | 965              | 17               | 1034             |
| LR    | yes         | 1:2                          | 1061             | 21               | 1115             |
| ANN   | no          | 1:1                          | 1000            | 10c              | 814              |
| ANN   | yes         | 1:1                          | 784             | 6c               | 762              |
| ANN   | yes         | 1:2                          | 851             | 11c              | 765              |
| RPART | no          | 1:1                          | 1000             | 12               | 814              |
| RPART | yes         | 1:1                          | 677              | 3                | 770              |
| RPART | yes         | 1:2                          | 688             | 6                | 781              |
| RF    | no          | 1:1                          | 1000             | 11               | 814              |
| RF    | yes         | 1:1                          | 972              | 14               | 1042             |
| RF    | yes         | 1:2                          | 1080             | 20               | 1121             |

Abbreviations: LR = logistic regression, ANN = artificial neural network, RPART = recursive partitioning and regression trees, RF = random forest.

a Nearest neighbor matching was adopted.

b Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on assessments at 24-h intervals.

c Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on hourly assessment over a sliding 24-h window.

d Number of variables.

e ANN stopped after 100 iterations.

---

Fig. 2. The real-time AIoT and IoT computation environment.

Fig. 3. The histograms for the delay from an IoT device to the QOCA platform.
3. Results

3.1. Patient characteristics

Of the 61,532 MIMIC-III ICU admissions reviewed, charted data were available for 60,840 ICU admissions. Through the cohort selection process, 1166 ICU patients met the selection criteria (see Fig. 1). In Study A, sepsis was identified in 666 patients, constituting 57.12% of the study cohort. In study B, sepsis was identified in 759 patients, constituting 57.09% of the study cohort. Of the 30 available potential clinical confounders extracted from the database, 23 items with a missing rate of 5% or less were included for analysis. Table 2 presents the patient characteristics of the sepsis and non-sepsis groups in Study A and Study B, respectively. Univariate analyses showed that significant differences were reported in ICU type and the medians of peripheral oxygen saturation, temperature, hemoglobin, hematocrit, glucose, potassium, sodium, creatinine, bilirubin, base deficit, blood urea nitrogen, SOFA, SAPS II, and comorbidity measures between the two groups on admission in Study A. In Study B, the analyses showed that significant differences were reported in ICU type and the medians of mean arterial pressure, peripheral oxygen saturation, temperature, hemoglobin, hematocrit, glucose, SAPS II, and comorbidity measures between the two groups on admission. There was a significant imbalance in the ICU admission status between patients in the sepsis group and the non-sepsis group in Study A and Study B, respectively.

3.2. Propensity score matching

Table 3 presents balance diagnostics of the baseline covariates between the sepsis and non-sepsis groups in Study A and Study B. Propensity scores, between the sepsis and non-sepsis groups was 0.552 in Study A and 0.584 in Study B. After matching, the SMD of the distance between the groups was 0.005 and 0.006 in Study A and Study B, respectively, which meant a good balance in the distance (SMD < 0.1). Furthermore, according to the SMD, all the covariates but hemoglobin and hematocrit were balanced after matching in Study A (SMD < 0.1). As to these two covariates, the VR was close to 1 and below the threshold of concern. In Study B, all the covariates but bicarbonate were balanced after matching (SMD < 0.1), and the VR later relieved the concern of imbalance for bicarbonate. In both studies, all the covariates were balanced between the sepsis and non-sepsis groups after propensity score matching (see Fig. 4).

3.3. Survival analysis

Overall, 255 in 1166 (21.87%) patients died within 90 days after ICU admission in the present study (see Fig. 5). In Study A, the Kaplan-Meier estimates reported that the 90-day survival probability of the sepsis group was 0.742 (CI, 0.709–0.776). As compared with the unadjusted 90-day survival probability of the non-sepsis group (0.834; CI, 0.802–0.867), 0.092 higher in the 90-day mortality was noted in the sepsis group. The crude odds ratio (OR) of the 90-day mortality between the sepsis and non-sepsis groups was 1.749 (P value < 0.001). When compared with the weighted 90-day survival probability of the non-sepsis group (0.797; CI, 0.743–0.854), 0.055 higher in the 90-day mortality was noted in the sepsis group. The adjusted OR of the 90-day mortality between the sepsis and non-sepsis groups was 1.725, i.e. exponential of the coefficient (P value = 0.002) (see Table 4).

Table 2

| Variable | Study A Cohort n = 1166 | P Value | Study B Cohort n = 1166 | P Value |
|----------|------------------------|---------|------------------------|---------|
| Age (year) | 65.0 (52.7–6.76.4) | 0.825 | 66.1 (51.9–7.74) | 0.323 |
| Male | 283 (56.6%) | 0.514 | 223 (54.7%) | 0.758 |
| ICU type | | | | |
| CCU | 79 (15.8%) | 0.001 | 75 (18.43%) | 0.978 |
| CSRU | 76 (15.2%) | | 58 (14.25%) | |
| MICU | 152 (30.4%) | | 116 (28.5%) | |
| TISCU | 84 (16.8%) | | 94 (23.1%) | |
| HR (beat/min) | 85.3 (74.5–96.0) | | 85.1 (74.9–94.0) | |
| MAP (mmHg) | 78.3 (71.7–86.8) | | 78.2 (71.7–86.2) | |
| RR (breath/min) | 18.2 (16.2–20.6) | | 17.8 (16.0–20.1) | |
| SpO2 (%) | 97.8 (96.4–99.1) | | 97.8 (96.4–99.0) | |
| Temperature (°C) | 36.9 (36.5–37.3) | | 36.8 (36.5–37.3) | |
| Hemoglobin (g/dL) | 10.7 (9.6–12.0) | | 10.6 (9.7–12.0) | |
| Hematocrit (%) | 31.2 (28.5–34.9) | | 31.2 (28.5–34.9) | |
| WBC (K/µL) | 11.1 (8.4–14.7) | | 11.3 (8.4–14.5) | |
| Platelet (K/µL) | 210.0 (159.0–286.0) | | 217.0 (166.5–287.0) | |
| Glucose (mg/dL) | 126.8 (108.0–151.8) | | 127.8 (108.0–149.8) | |
| BUN (mg/dL) | 18.2 (12.0–32.0) | | 18.2 (12.0–32.0) | |
| Creatinine (mg/dL) | 0.9 (0.7–1.4) | | 0.9 (0.7–1.4) | |
| Sodium (mEq/L) | 139.0 (137.0–141.0) | | 139.0 (137.0–141.0) | |
| Potassium (mEq/L) | 4.1 (3.8–4.4) | | 4.1 (3.8–4.4) | |
| Chloride (mEq/L) | 106.0 (102.0–109.0) | | 106.0 (102.0–109.0) | |
| Bicarbonate (mEq/L) | 24.0 (21.5–26.0) | | 24.0 (21.5–26.0) | |
| Maximum SOFA score | 5.0 (4.0–7.0) | | 5.0 (4.0–6.0) | |
| SAPS II | 32.0 (24.0–40.0) | | 31.0 (24.0–39.0) | |
| SIRS criteria | 3.0 (2.0–3.0) | | 3.0 (2.0–4.0) | |
| ECMvW | 7.0 (1.0–13.0) | | 7.0 (1.0–13.0) | |

Abbreviations: CCU = coronary care unit, CSRU = cardiac surgery recovery unit, MICU = medical intensive care unit, SICU = surgical intensive care unit, TSICU = trauma surgical intensive care unit, HR = heart rate, MAP = mean arterial pressure, RR = respiratory rate, SpO2 = peripheral oxygen saturation, WBC = white blood cell, BUN = blood urea nitrogen, SOFA = sequential organ failure assessment, SAPS II = simplified acute physiology score II, SIRS = systemic inflammatory response syndrome, ECMvW = Elixhauser comorbidity measures, van Walraven.

* Data are presented as median (interquartile range) for continuous variables and as number (proportion) for categorical variables.

1 Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on assessments at 24-h intervals.

2 Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on hourly assessment over a sliding 24-h window.
The SOFA score change detected by the hourly assessment protocol was 1.0 points upon identification of sepsis. Nevertheless, the concomitant total SOFA score change detected by the hourly assessment protocol was 1.0 point lower than that detected by the 24-hourly counterpart. The statistical test implied that the former could detect subler changes of total SOFA score than the latter (P value < 0.001).

3.5. Sensitivity analysis

After the restrictions on the ICU length of stay and the first 24-h change of total SOFA score were lifted, 9715 ICU patients met the selection criteria. The 24-h assessment protocol identified 6660 sepsis patients, constituting 68.55% of the selected cohort; the hourly assessment protocol identified 6222 patients, constituting 64.05% of the selected cohort. The 24-h assessment protocol identified sepsis at a median time of 24 h after ICU admission while the hourly assessment protocol identified sepsis at a median time of 10 h after ICU admission. The median difference of the sepsis identification times was 14.5 h, for which the statistical test significantly suggested that the hourly assessment protocol outmatched the 24-h counterpart (P value < 0.001) (see Table 5).

The total SOFA score change at the time of sepsis identification was also investigated. As presented in Table 5, the 24-h and hourly assessment protocols detected a median total score change of 4 and 3 points, respectively, upon identification of sepsis. Besides, the concomitant total SOFA score change detected by the hourly assessment protocol was 2.5 points lower than that detected by the 24-hourly counterpart. The statistical test implied that the former could detect subler changes of total SOFA score than the latter (P value < 0.001).

4. Discussion

In the present study, we explored the possibility of a solution to bridging clinical information system with AI medicine, focusing on data asynchrony and data dimensionality. After all, for the healthcare system,
clinical services usually come before clinical research. MIMIC-III is based on data accrued during day-to-day clinical services rather than originally designed for clinical research. Likewise, issues of data asynchrony and data dimensionality have been inevitable in contemporary clinical information systems. Nevertheless, we look forward to an end-to-end approach for paradigm transformation to AI-powered health-care system in the near future. On the basis of our exploratory research, real-time SOFA score to signal emerging sepsis in critical care can be realized. Furthermore, it may signify a promising example for developing real-time clinical information system.

Based on the results, we address three points regarding the early identification of sepsis. First, a real-time assessment protocol can earlier identify patients with sepsis under suspicion of infection. As shown in Table 5, the paired analysis reported that the hourly assessment protocol identified patients with sepsis by a median of 14.5 h earlier than the 24-h counterpart. During the clinical practice, the day-to-day intense routines keep the system functioning to assure a certain quality of healthcare. However, an hourly assessment protocol would be very demanding, and there are limits to human power and attention span. In the age of the

Fig. 4. Distribution and covariate balance between the non-sepsis and sepsis groups before and after propensity score matching. Abbreviations: HR = heart rate, MAP = mean arterial pressure, RR = respiratory rate, SpO₂ = peripheral oxygen saturation, WBC = white blood cell, SOFA = sequential organ failure assessment, SAPS II = simplified acute physiology score II, SIRS = systemic inflammatory response syndrome, ECM_vW = Elixhauser comorbidity measures_van Walraven.
digital revolution, real-time intelligence could be accomplished by new data generation technologies and new hardware accelerators and assist physicians in treating patients. Second, a real-time assessment protocol can detect a seemingly inconspicuous change of total SOFA score indicating the timing of sepsis identification. As shown in Table 5, the hourly assessment protocol tracked down the total SOFA score change by a median of 1.0 point lower than the 24-h counterpart. Although the difference appeared subtle, it addressed how sensitive the assessment protocols could earlier identify patients of sepsis. Accordingly, it would be the potential niche of real-time intelligence. At the level of clinical translation, the wealth of medical data provides the soil for AI medicine. Nevertheless, the sophistication required by clinical translation remains challenging. The combination of AI technologies with the Internet of Things (IoT) to simulate a human body as well as down to the subsystem levels can be encouraging. The QOCA AiTo platform can be a solution (see Fig. 2).

Third, real-time SOFA score can signal emerging sepsis and be integrated into a range of other applications. As recognized in the medical society, there is no gold standard for diagnosing sepsis but the Sepsis-3 definition. Under suspicion of infection, serial SOFA scores provide a dynamic representation of the disease course, and the change of total SOFA score plays a crucial role in identifying sepsis. In terms of the timing advantages, the real-time SOFA score deserves further investigation to develop a new definition. Under suspicion of infection, serial SOFA scores provide a dynamic representation of the disease course, and the change of total SOFA score plays a crucial role in identifying sepsis. In terms of the timing advantages, the real-time SOFA score deserves further investigation to develop a new definition.

| Variable | Coef | Study A | Coef | Study B |
|----------|------|---------|------|---------|
| Sepsis   | 0.545 | 0.009 | 0.577 | 0.009 |
| Age      | 0.025 | 0.039 | 0.027 | 0.039 |
| Male     | -0.015 | -0.356 | -0.135 | 0.552 |
| ICU type | 0.222 | 0.354 | 0.062 | 0.325 |
| HR       | -0.006 | 0.007 | -0.017 | 0.009 |
| MAP      | -0.002 | 0.014 | -0.019 | 0.013 |
| RR       | 0.071 | 0.119 | 0.036 | 0.129 |
| SpO2 (%) | -0.004 | -0.103 | -0.089 | 0.108 |
| Temperature (°C) | -0.273 | -0.545 | -0.006 | -0.293 |
| Hematocrit (%) | 0.006 | 0.043 | 0.046 | 0.029 |
| WBC (K/µL) | -0.012 | 0.020 | -0.031 | 0.035 |
| Platelet (K/µL) | 0.001 | 0.002 | -0.002 | 0.002 |
| Glucose (mg/dL) | 0.001 | 0.004 | 1.049 | <0.001 |
| Creatinine (mg/dL) | -0.083 | 0.022 | 1.408 | <0.009 |
| Sodium (mEq/L) | -0.018 | 0.022 | 1.108 | <0.005 |
| Potassium (mEq/L) | -0.046 | 0.279 | 1.301 | <0.216 |
| Bicarbonate (mEq/L) | -0.022 | 0.018 | 1.227 | <0.038 |
| Maximum SOFA score | -0.055 | 0.015 | 1.565 | <0.054 |
| SAPS II  | 0.059 | 0.079 | 1.922 | 0.043 |
| SIRS criteria | 0.157 | 0.367 | 1.488 | 0.036 |
| ECM_vW   | 0.047 | 0.027 | 0.036 | 0.078 |

Abbreviations: HR = heart rate, MAP = mean arterial pressure, RR = respiratory rate, SpO2 = peripheral oxygen saturation, WBC = white blood cell, SOFA = sequential organ failure assessment, SAPS II = simplified acute physiology score II, SIRS = systemic inflammatory response syndrome, ECM_vW = Elixhauser comorbidity measures, van Walraven, Coef = coefficient, VIF = variance inflation factor.

* Hemoglobin, chloride, and blood urea nitrogen were removed owing to multicollinearity.
* Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on assessments at 24-h intervals.
* Logit function was applied in the model and 95% confidence intervals of coefficients were reported.

Randomized controlled trials (RCTs) remain the current gold standard in clinical studies. However, among others, demanding logistics, ethical issues, cost, and feasibility usually limit RCTs in practice. The present study is a retrospective observational study, which could take advantage of the wealth of existing medical data and precede a pivotal RCT if indicated. To tackle the common disadvantage concerning distribution imbalance of the baseline covariates between study groups, we conducted propensity score matching and predicted an adjusted OR of the 90-day mortality between the sepsis and non-sepsis groups by a generalized linear model. As presented in Table 4, the adjusted ORs were supported that the hourly assessment protocol excelled the 24-h counterpart regarding the strength of OR, i.e., 2.578 vs. 1.725. Propensity score methodology can be used to help design observational studies in a way analogous to the way randomized experiments are designed [38, 39, 49, 50].

In the present study, we constructed the research based on those patients who were under suspicion of infection. We intended not only to comply with the Sepsis-3 definition but also to foster systematic comparability between study groups. To deal with the potential disadvantage that non-sepsis patients were not enough to match sepsis patients in the propensity score matching stage, we adopted the strategy of matching with replacement by a ratio of 2:1 to extract as much information as applicable. Another potential disadvantage was the sample size of the study cohort as compared with the samples available in the database. We then lifted the restrictions on the ICU length of stay and the size of the study cohort as compared with the samples available in the database. We then lifted the restrictions on the ICU length of stay and the size of the study cohort as compared with the samples available in the database. We then lifted the restrictions on the ICU length of stay and the size of the study cohort as compared with the samples available in the database. We then lifted the restrictions on the ICU length of stay and the size of the study cohort as compared with the samples available in the database. We then lifted the restrictions on the ICU length of stay and the size of the study cohort as compared with the samples available in the database.
healthcare system more substantially if the huge amounts of data accumulated during the day-to-day clinical practice could be utilized more effectively and efficiently. Real-time intelligence will prevail in medicine. Concealed intelligence should be uncovered to facilitate the end-to-end approach for paradigm transformation to AI-powered healthcare system.

5. Conclusions

In this observational study of adult ICU patients under suspicion of infection, the hourly assessment protocol outperformed the 24-h assessment protocol in identifying patients of sepsis by the serial SOFA score change. Real-time SOFA score can signal emerging sepsis prospectively. Real-time intelligence can carry forward the transformation of healthcare system by embracing the notions and technologies of the digital revolution.

Declaration of competing interest

The authors declare that they have no known competing financial
interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by Taiwan Ministry of Science and Technology, ROC under Grant Number: MOST 110-2321-B-A49-003.

Table 5

| Table 5 Evaluation of sepsis identification between 24-h and hourly assessment protocols. |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| **Sepsis identification time**   | **Median (Minimum-Maximum)**    | **Median Difference**^d^          | 2.5th percentile  | 97.5th percentile | P Value  |
|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| 24-h assessment                   | 72 (48–1152)                    | 14.5                             | 13.5                             | 15.5                             | <0.001                           |
| Hourly assessment                 | 64 (1–1139)                     |                                  |                                  |                                  |                                  |
| Total SOFA score change           |                                  |                                  |                                  |                                  |                                  |
| 24-h assessment                   | 3 (2–11)                        | 1.0                              | 1.0                              | 1.0                              | <0.001                           |
| Hourly assessment                 | 3 (2–7)                         |                                  |                                  |                                  |                                  |
| Sepsis identification time - SA   |                                  |                                  |                                  |                                  |                                  |
| 24-h assessment                   | 24 (24–1152)                    | 15.0                             | 14.5                             | 15.5                             | <0.001                           |
| Hourly assessment                 | 10 (1–1139)                     |                                  |                                  |                                  |                                  |
| Total SOFA score change - SA      |                                  |                                  |                                  |                                  |                                  |
| 24-h assessment                   | 4 (2–20)                        | 2.5                              | 2.0                              | 2.5                              | <0.001                           |
| Hourly assessment                 | 3 (2–13)                        |                                  |                                  |                                  |                                  |

^a^ Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on assessments at 24-h intervals.

^b^ Sepsis was identified as an increase in SOFA score of 2 points or more under suspicion of infection based on hourly assessment over a sliding 24-h window.

^c^ Sensitivity analysis: We lifted the restrictions on the ICU length of stay and the first 24-h change of total SOFA score in the sensitivity analysis.

^d^ Wilcoxon signed-rank test estimated the median of the difference between 24-h and hourly assessments with 95% confidence interval.
Kassambara A, Kosinski M, Biecek P. survminer: drawing survival curves using ‘ggplot2’ (R package version 0.4.9). Available: https://CRAN.R-project.org/package=survminer; 2021.

Normand S-LT, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol 2001;54(4):387–98. https://doi.org/10.1016/S0895-4356(00)00321-8.

Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci 2010;25(1):1–21. https://doi.org/10.1214/09-STS313.

Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. Health Serv Outcome Res Methodol 2001;2(3):169–88. https://doi.org/10.1023/A:1020363010465.

Fox J, Monette G. Generalized collinearity diagnostics. J Am Stat Assoc 1992;87(417):178–83. https://doi.org/10.1080/01621459.1992.10475190.

Chang T, Lin Y-B. AIoT for digital transformation of healthcare. IEEE CTSoc-NCT NEWS on Consumer Technology; 2021.

Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available: https://www.R-project.org/.

Van Rossum G, Drake F. The Python language reference (version 3.8). Available: https://docs.python.org/3.8/reference/; 2021.

Lin Y-W, Lin Y-B, Hsiao C-Y, Wang Y-Y. IoTalk-RC: sensors as universal remote control for aftermarket home appliances. IEEE Internet Things 2017;4(4):1104–12. https://doi.org/10.1109/JIOT.2017.2715859.

Paoli CJ, Reynolds MA, Sinha M, Gitlin M, Crouser E. Epidemiology and costs of sepsis in the United States - an analysis based on timing of diagnosis and severity level. Crit Care Med 2018;46(12):1889–97. https://doi.org/10.1097/CCM.0000000000003342.

Medicine Agency European. Guideline on clinical investigation of medicinal products for the treatment of sepsis. Available: https://www.ema.europa.eu/en/clinical-investigation-medicinal-products-treatment-sepsis; 2006.

Mehazza A, Laterre PF, Russell JA, Bergmann A, Gattinoni L, Gayat E, et al. Designing phase 3 sepsis trials: application of learned experiences from critical care trials in acute heart failure. J Intensive Care 2016;4(1):1–11. https://doi.org/10.1186/s40560-016-0151-6.

Combi C, Pozzi G. Clinical information systems and artificial intelligence: recent research trends. Yearb Med Inform 2019;28(1):83–94. https://doi.org/10.1055/s-0039-1677915.

Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011;46(3):399–424. https://doi.org/10.1080/00273171.2011.568786.

Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit Anal 2007;15(3):199–236. https://doi.org/10.1093/polanal/pml013.