Early prediction of sepsis in intensive care patients using the machine learning algorithm NAVOY® Sepsis, a prospective randomized clinical validation study

Inger Persson, PhD a,*, Andreas Macura, MSc b, David Becedas, MSc b, Fredrik Sjövall, MD, PhD c

a Department of Statistics, Uppsala University, Uppsala, Sweden, AlgoDx AB, Stockholm, Sweden
b AlgoDx AB, Stockholm, Sweden
c Department of Intensive- and Perioperative Medicine, Skåne University Hospital, Malmö, Sweden

ARTICLE INFO

Keywords:
Sepsis
Prediction
Early detection
Machine learning
Software as a medical device
Intensive care unit

ABSTRACT

Purpose: To prospectively validate, in an ICU setting, the prognostic accuracy of the sepsis prediction algorithm NAVOY® Sepsis which uses 4 h of input for routinely collected vital parameters, blood gas values, and lab values.

Materials and methods: Patients 18 years or older admitted to the ICU at Skåne University Hospital Malmö from December 2020 to September 2021 were recruited in the study. A total of 304 patients were randomized into one of two groups: Algorithm group with active sepsis alerts, or Standard of care. NAVOY® Sepsis made silent predictions in the Standard of care group, in order to evaluate its performance without disturbing the outcome. The study was blinded, i.e., study personnel did not know to which group patients were randomized. The healthcare provider followed standard practices in assessing possible development of sepsis and intervening accordingly. The patients were followed-up in the study until ICU discharge.

Results: NAVOY® Sepsis could predict the development of sepsis, according to the Sepsis-3 criteria, three hours before sepsis onset with high performance: accuracy 0.79; sensitivity 0.80; and specificity 0.78.

Conclusions: The accuracy, sensitivity, and specificity were all high, validating the prognostic accuracy of NAVOY® Sepsis in an ICU setting, including Covid-19 patients.

1. Introduction

Sepsis is a potentially life-threatening condition where initial signs can be difficult to interpret for healthcare professionals. Sepsis arises when the host response to infection causes organ dysfunction and is associated with high mortality where approximately one patient in five dies [1]. Some of these patients will deteriorate further, with inability to maintain normal blood pressure i.e., septic shock with even higher mortality [2]. Globally, sepsis affected 49 million individuals in 2017 [3].

Timing is critical in sepsis treatment. Early diagnosis has been shown to reduce delays in treatment, increase appropriate care, and reduce mortality [4]. Despite being a potentially fatal condition, early intervention with antibiotics, antiviral treatment, oxygen therapy/mechanical ventilation and fluid control leads to dramatically improved outcomes.

Today, sepsis diagnosis is made by combining information from clinical examination, monitoring devices, and laboratory data. This procedure is both time-consuming and subjective, i.e., heavily dependent on the skills and experience of any given doctor or nurse. Diagnosis of sepsis is even more difficult in the ICU setting where many patients have other conditions with similar physiological and laboratory alterations as those encountered in patients with sepsis. Hundreds of biomarkers have been tested clinically as prognostic markers in sepsis, where procalcitonin (PCT) and C-reactive protein (CRP) have been the most widely used [5-7]. None have demonstrated sufficient specificity or sensitivity to be routinely used, as a definite marker of sepsis in clinical practice [5]. PCT has shown mean values of both sensitivity and specificity around 71% and an area under the summary receiver operator characteristic curve of 0.78 [8], and CRP has shown a sensitivity between 40 and 87%, with a specificity of 53–82% [9]. In this context, there exists a significant room for improvement.

The algorithm tested in this study, NAVOY® Sepsis, has retrospectively demonstrated the ability to detect ICU patients at high risk of...
developing sepsis within the coming hours based on routinely collected clinical variables [10].

In this study, we have evaluated the predictive performance of NAVOY® Sepsis in a clinical setting. This is, as far as the authors are aware, the largest randomized clinical trial ever conducted with a machine learning sepsis prediction algorithm, and the first one that clinically validates a sepsis prediction algorithm against the Sepsis-3 criteria.

2. Materials and methods

2.1. Study design and patients

This study (ClinicalTrials.gov, NCT04570618) was a randomized prospective study with wide inclusion criteria. The Clinical Investigation Plan (CIP) was designed in accordance with ISO 14155 and the Declaration of Helsinki. The study was designed to prospectively validate the sepsis prediction medical device software NAVOY® Sepsis in a general ICU setting. The CIP and relevant documentation were approved by the Swedish Ethical Review Authority (reference number 2020-00190). The study title “Prediktion av sepsis med en dataalgoritm hos patienter inlagda på Intensivvårdsavdelning, en randomiserad klinisk valideringsstudie (SEP-SE-02)” approved April 1, 2020 and the Swedish Medical Products Agency (reference number 5.1–2020-58,206, approved August 21, 2020).

Patients 18 years of age or older admitted to the ICU at Skåne University Hospital Malmö from December 2020 were eligible for recruitment to the study. Patients were excluded if they participated in another interventional clinical trial which could potentially impact variables used by NAVOY® Sepsis, if they were known to be pregnant, if death deemed imminent and inevitable, if the patient was incapable of making an informed decision due to chronically reduced mental capacity, or if the patient had previously been enrolled in the study. Due to the non-interventional nature of the study, deferred consent was approved by the Ethical Review Authority and the Medical Products Agency. If reachable, the patient’s next of kin or legally authorized representative was informed about the study and could waive participation. Written informed consent was sought directly from patients when recovered from the acute reduced mental capacity related to their illness.

Patients were randomized to either the Algorithm cohort, with sepsis alerts; or the Standard of care (SoC) cohort. The algorithm made predictions in both cohorts, but alerts were not displayed in the SoC cohort, to evaluate the algorithm’s performance without possible confounding of the outcome. The study was blinded in the sense that the study personnel did not know for which patients a sepsis alert, issued by the algorithm, was to be displayed or not, i.e., which patients were assigned to which study cohort. This information was however naturally revealed for any patients with a displayed sepsis alert. The CIP did not contain any interventions or actions following a sepsis alert, but to avoid patients possibly being treated differently in the Algorithm cohort and thus having an effect on the outcome, the primary outcome was evaluated only on the SoC cohort.

Patients remained in the study their entire ICU stay (from ICU admission until discharge or death) or until 30 days after last patient included (study closure), whichever occurred first. After ICU discharge, patient data regarding hospital length of stay and any in-hospital death were collected from other wards until hospital discharge. There was no diagnostic or medical intervention mandated by the CIP and ICU staff interacted with the patients during the ICU stay according to routine practices. This included standard practices in assessment of possible development of sepsis and subsequent interventions throughout the course of the study.

2.2. Procedures

NAVOY® Sepsis was deployed in tablets placed at each ICU bed. The randomization scheme was embedded in the device and eligible patients were randomized into the study in a ratio of 2:1 to the SoC cohort versus the Algorithm cohort, at the start of data entry. Each patient was assigned a unique study ID in ascending order. The randomization list was prepared by an external statistician (not otherwise involved in the study) and kept strictly confidential, accessible only to authorized persons until the time of unblinding.

The sepsis risk predictions were calculated by the NAVOY® Sepsis algorithm [10], based on the following 20 variables: age, sex, heart rate, respiratory rate, temperature, systolic blood pressure, diastolic blood pressure, vasopressor use, serum creatinine, glucose, lactate, platelets, white blood cell count (WBC), blood urea nitrogen (BUN), bilirubin, pH, oxygen saturation pulse oximetry (SpO₂), fraction of inspired oxygen (FiO₂), International Normalized Ratio, and Glasgow Coma Scale. Data were entered by bedside ICU staff when available. Age and sex were registered once on admission. Heart rate, respiratory rate, temperature, systolic and diastolic blood pressure, FiO₂, SpO₂ and vasopressor use were entered hourly. Data on serum creatinine, glucose lactate and pH were captured from blood gas analysis every 4–6 h and WBC, platelets, BUN, and bilirubin were taken from daily routine blood samples. If the predicted sepsis risk exceeded a predefined threshold in a patient assigned to the algorithm cohort, an alert was displayed. Alerts were logged in the study database, but not displayed, if the predicted sepsis risk score exceeded the threshold in patients in the SoC cohort, or any succeeding alerts in patients in the algorithm cohort. Succeeding alerts can be displayed in a real setting, the choice was made not to not display them in this study due to the intention to investigate any actions taken in the hours following an alert which would have been difficult to do if more than one alert would have been displayed.

2.3. Outcomes

As primary outcome, specificity, sensitivity and accuracy were used to clinically validate the sepsis prediction performance of the algorithm in the SoC cohort. A sepsis event was defined in accordance with the Sepsis-3 criteria [11], which requires a suspected infection and an increase in SOFA score of at least 2 points. Suspected infection was defined as present if a culture was ordered within 24 h after antibiotics were started, or antibiotics were prescribed within 72 h after a culture order, in accordance with Seymour et al. [12]. The ICU, in which the study was performed, does not use routine cultures so only cultures by indication was taken during the study and the time point of culture sample collection was used. All cultures were used; blood, urine, liquor, etc. The time point of suspected infection was determined as the earlier of these two. A patient was considered septic if the total SOFA score increased by at least 2 points within the time window from 48 h before to 24 h after the time of suspected infection. The time point of sepsis onset was defined as the time of the 2-point SOFA increase. Patients not fulfilling the Sepsis-3 criteria at any time were defined as non-septic.

Secondary outcomes were ICU length of stay and hospital length of stay (days from ICU admission to ICU discharge, hospital discharge, or death), and any treatments with fluids and/or antibiotics initiated at or after the time point when a sepsis alert was triggered by the algorithm.

Safety outcomes were Serious Adverse Events (SAEs), any new diagnoses/medical occurrences developed at ICU, treatments and interventions initiated at ICU, and in-hospital death. Only SAEs that were thought to be related to the use of NAVOY® Sepsis were reported.

2.4. Statistical analyses

The sample size for the SoC cohort was calculated to achieve an acceptable margin of error for a 95% confidence interval for the accuracy (primary outcome). The sample size for the Algorithm cohort was calculated to detect a difference in ICU length of stay (secondary objective). The planned sample size of 200 patients for the SoC cohort provided a 93% probability to achieve a 95% confidence interval with a
margin of error of ±0.06 (in most statistical literature a margin of error of 0.04–0.08 is considered desirable), and the planned sample size of 100 patients for the Algorithm cohort provided a power of 87%, based upon Gray's test [13] comparing the probability of being discharged from the ICU between the two cohorts, with a 5% significance level. This gave a total planned sample size of 300 patients, and a randomization ratio of 2:1.

All outcomes were evaluated by descriptive statistics. As death can shorten length of stay, the secondary outcomes relating to ICU and hospital length of stay were evaluated comparing the cohorts using the cumulative incidence function [14], giving the probability of being discharged from the ICU or hospital at or before a certain time point, taking the competing risk of death into account. All statistical analyses were conducted using SAS software, Version 9.4 of the SAS System for Windows.

Patients with missing data were included where possible, e.g., in the description of the patient population. Wherever the analysis required data on a variable, patients with missing data were excluded from the analysis. The analyses of the primary and secondary outcomes focused on the subpopulation excluding patients already septic at ICU admission and/or receiving antibiotics before/at ICU admission. In addition, the analysis of the primary outcome was repeated for both study cohorts combined.

3. Results

The first patient was enrolled in the study on December 3rd 2020 and the study was closed on November 1st 2021. Of 511 eligible patients, a total of 304 patients were randomized, 198 in the SoC cohort and 106 in the algorithm cohort (Fig. 1). All 304 patients were included in the analyses for secondary and safety outcomes. As the algorithm requires data on each of the 20 variables, risk scores cannot be produced during the first few hours of an ICU stay until all variables have been measured and entered. Five patients in the SoC cohort did not have all 20 variables measured during their ICU stay (Fig. 1). Further, the algorithm requires 4 h of input, starting from the time point when all 20 variables have been measured. As a consequence, risk scores cannot be produced during the first few hours of an ICU stay. There were 59 patients that had their sepsis onset earlier than the algorithm could make a prediction (including patients septic upon ICU admission) leaving 134 patients (Fig. 1). When excluding 49 patients already receiving antibiotics upon ICU admission (and thus not being able to be predicted for sepsis
development), the analyses for primary outcomes (SoC cohort only) was based on 85 patients (Fig. 1).

The distribution of patients was well balanced between cohorts (Table 1) and the demography representative of a Swedish ICU population [15,16]. The mean age for all patients was 63 years, 35% were female, and 14% had a Covid-19 diagnosis during the study.

Using the pre-defined threshold, the algorithm successfully identified patients at risk of developing sepsis with a sensitivity of 0.80, a specificity of 0.78, and a total accuracy of 0.79, three hours before sepsis onset (Table 2). When sepsis was predicted, the probability that the patient actually would become septic was 0.53 (positive predictive value), and when sepsis was not predicted the probability that this was a correct prediction (negative predictive value) was 0.93. Algorithm performance in subgroups of age, gender, SAPS III score, chronic health conditions, and length of stay are presented in the supplementary material. The area under the ROC curve (AUROC) for the algorithm was 0.80 on predicting sepsis three hours before onset (Fig. 2). The algorithm performance was similar in the Covid-19 patients included in the primary analysis (results provided in the supplementary material), and a sensitivity analysis where both study cohorts were combined (n = 126) showed acceptable results: accuracy 0.73; sensitivity 0.70, and specificity 0.74 (results provided in the supplementary material).

The secondary outcomes ICU length of stay and hospital length of stay showed no difference when comparing the Algorithm and SoC cohorts (results provided in the supplementary material). Very few patients had fluid resuscitation (6%), culture order (4%), or antibiotics administration (7%) initiated at or within 4 h after the time point when a sepsis alert was triggered by the algorithm in the Algorithm cohort (open alert), or when an alert was triggered by the algorithm in the SoC cohort but hidden to study personnel (silent alert) (results provided in the supplementary material).

No serious adverse events were reported during the study.

4. Discussion

In this study, NAVOY® Sepsis could identify patients with risk of developing sepsis with good performance three hours before sepsis onset, close to the demonstrated performance in the proof-of-concept study [10], which validates the prognostic accuracy of the algorithm in a prospective ICU setting. NAVOY® Sepsis has previously demonstrated a prediction performance superior to that of existing sepsis early warning scoring systems (e.g. qSOFA, SOFA, SIRS, MEWS, and NEWS2) and comparable with those of other prediction algorithms designed to predict sepsis onset [10]. NAVOY® Sepsis is a prediction algorithm

### Table 1
Patient demographics and baseline clinical characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Whole study</th>
<th>Primary analysis</th>
<th>Sensitivity analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Algorithm</td>
<td>SoC</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>(n = 106)</td>
<td>(n = 198)</td>
<td>(n = 304)</td>
</tr>
<tr>
<td>Age, mean (std), years</td>
<td>61.0 (15.1)</td>
<td>63.5 (14.1)</td>
<td>62.6 (14.5)</td>
</tr>
<tr>
<td>Age, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–29</td>
<td>5 (4.7)</td>
<td>5 (2.5)</td>
<td>10 (3.3)</td>
</tr>
<tr>
<td>30–39</td>
<td>9 (8.5)</td>
<td>11 (5.6)</td>
<td>20 (6.6)</td>
</tr>
<tr>
<td>40–49</td>
<td>9 (8.5)</td>
<td>16 (8.1)</td>
<td>25 (8.2)</td>
</tr>
<tr>
<td>50–59</td>
<td>22 (20.8)</td>
<td>30 (15.2)</td>
<td>52 (17.1)</td>
</tr>
<tr>
<td>60–69</td>
<td>23 (21.7)</td>
<td>63 (31.8)</td>
<td>86 (28.3)</td>
</tr>
<tr>
<td>70+</td>
<td>38 (35.9)</td>
<td>73 (26.9)</td>
<td>111 (36.5)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>32 (30.2)</td>
<td>74 (37.4)</td>
<td>106 (34.9)</td>
</tr>
<tr>
<td>Covid-19 diagnosis, n (%)</td>
<td>10 (9.4)</td>
<td>33 (16.7)</td>
<td>43 (14.1)</td>
</tr>
<tr>
<td>SAPS III score, mean (std)</td>
<td>64.0 (17.3)</td>
<td>59.2 (15.3)</td>
<td>60.9 (16.1)</td>
</tr>
<tr>
<td>Chronic health conditions, n (%)</td>
<td>56 (52.8)</td>
<td>127 (64.1)</td>
<td>183 (60.2)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal</td>
<td>13 (12.3)</td>
<td>30 (15.2)</td>
<td>43 (14.1)</td>
</tr>
<tr>
<td>Liver</td>
<td>11 (10.4)</td>
<td>13 (6.6)</td>
<td>24 (7.9)</td>
</tr>
<tr>
<td>Organ transplant</td>
<td>1 (0.9)</td>
<td>5 (2.5)</td>
<td>6 (2.0)</td>
</tr>
<tr>
<td>HIV positive</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Mental health disorder</td>
<td>23 (21.7)</td>
<td>33 (16.7)</td>
<td>56 (18.4)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>25 (23.6)</td>
<td>47 (23.7)</td>
<td>72 (23.7)</td>
</tr>
<tr>
<td>Chronic Obstructive</td>
<td>13 (12.3)</td>
<td>20 (10.1)</td>
<td>33 (10.9)</td>
</tr>
<tr>
<td>Pulmonary Disorder (COPD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>19 (17.9)</td>
<td>45 (22.7)</td>
<td>64 (21.1)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>24 (22.6)</td>
<td>27 (13.6)</td>
<td>51 (16.8)</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>2 (1.9)</td>
<td>6 (3.0)</td>
<td>8 (2.6)</td>
</tr>
<tr>
<td>Length of ICU stay, mean (std), days</td>
<td>4.3 (5.4)</td>
<td>4.1 (7.0)</td>
<td>4.2 (6.5)</td>
</tr>
<tr>
<td>Length of ICU stay, n (%)</td>
<td>0–4</td>
<td>71 (67.0)</td>
<td>152 (76.8)</td>
</tr>
<tr>
<td>5–9</td>
<td>23 (21.7)</td>
<td>27 (13.6)</td>
<td>50 (16.5)</td>
</tr>
<tr>
<td>10–14</td>
<td>6 (5.7)</td>
<td>6 (3.0)</td>
<td>12 (4.0)</td>
</tr>
<tr>
<td>15–19</td>
<td>3 (2.8)</td>
<td>3 (1.5)</td>
<td>6 (2.0)</td>
</tr>
<tr>
<td>20–24</td>
<td>2 (1.9)</td>
<td>4 (2.0)</td>
<td>6 (2.0)</td>
</tr>
<tr>
<td>25+</td>
<td>1 (0.9)</td>
<td>6 (3.0)</td>
<td>7 (2.3)</td>
</tr>
<tr>
<td>Death during ICU stay, n (%)</td>
<td>14 (13.2)</td>
<td>23 (11.6)</td>
<td>37 (12.2)</td>
</tr>
</tbody>
</table>

n = number of patients.

std = standard deviation.

* All patients in the study were included in the analysis of secondary outcomes.

### Table 2
Performance metrics for algorithm predicting sepsis 3 h in advance.

<table>
<thead>
<tr>
<th>Metric</th>
<th>n</th>
<th>Value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUROC</td>
<td>85</td>
<td>0.80</td>
</tr>
<tr>
<td>Accuracy</td>
<td>85</td>
<td>0.79 (0.70–0.88)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>20</td>
<td>0.80 (0.62–0.98)</td>
</tr>
<tr>
<td>Specificity</td>
<td>65</td>
<td>0.78 (0.68–0.88)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>30</td>
<td>0.53</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>55</td>
<td>0.92</td>
</tr>
</tbody>
</table>

CI = Confidence Interval.

AUROC = area under the receiver operating characteristic curve.
developed by the use of machine learning methods, optimized to make accurate predictions 3 h before sepsis onset. It uses 4 h of input for variables routinely collected in ICUs. Different prediction horizons were evaluated during algorithm development, between 1 and 24 h before sepsis onset, but at earlier time points there were considerably fewer ICU stays with data for all variables of interest and 3 h was chosen as the earliest time point with good availability of data. Further details of the algorithm development are found in the publication by Persson et al. [10].

The secondary and exploratory outcomes showed little or no difference when comparing the Algorithm cohort versus the SoC cohort where alerts were not displayed. This was unsurprising considering the design of the study as the CIP did not contain any interventions or actions following a sepsis alert issued by the investigational device, regardless of cohort. Both cohorts were treated according to the current standard of care. In the Algorithm cohort where the sepsis alert was displayed, bedside clinicians could have been influenced by the alert to initiate a work-up for sepsis, including cultures and laboratory as well as treatment, according to clinical standard. This seems however not to have been the case as there was a low rate (4%–7%) of cultures and initiation of treatment such as fluids and antibiotics in the 4 h following a sepsis alert. The similarity between the study cohorts provided a possibility to conduct a sensitivity analysis where the primary outcome was repeated on the whole group of patients (both cohorts combined), with acceptable results.

Another issue affecting the secondary outcomes was the ongoing Covid-19 pandemic, where ICU staff were experiencing a high workload during the entire study. For some patients, the manual data entry was done retrospectively (e.g., morning shift staff entering a backlog of data from the previous hours). This, in turn, caused a delay in algorithm predictions and sepsis alerts (if any), causing the alerts to be of less relevance for any clinical decisions. This, in addition to the limitations described above, made it difficult to find any differences between the cohorts in secondary outcomes such as ICU or hospital length of stay.

No serious adverse events were reported during the study. This is in line with what could be expected considering that the device is a clinical decision support and no interventions were prescribed in the CIP. The treatment and care of patients were provided according to the clinical staff’s discretion and current standard of care. Also, treatments and interventions initiated at ICU and in-hospital death, as described in the secondary objective, were similar in both cohorts.

Sepsis is a potentially life-threatening condition where initial signs of disease can be difficult to interpret for healthcare professionals, and current methods for detection of sepsis are incapable of early prediction. Many ICU patients display similar symptoms, so-called sepsis mimics [17], which add to the challenges of identifying true cases of sepsis. Results from this study further confirm the results of retrospective studies [10] indicating that the algorithm is able to predict sepsis three hours earlier compared to Sepsis-3 criteria [11].

As mentioned above, the present study was conducted during the Covid-19 pandemic. Patients with Covid-19 can be said to have a viral sepsis, while the algorithm was trained to detect bacterial sepsis. Even so, the algorithm was able to predict the onset of sepsis among patients with Covid-19 with seemingly similar accuracy as in the other ICU patients.

4.1. Relation to other studies

Fleuren et al. [18] and Moor et al. [19] reviewed previously developed sepsis prediction algorithms and found that very few had been prospectively evaluated in clinical practice. To our knowledge, very few ICU algorithms are currently available for clinical use, and only in the US so far [20-23]. Our results indicate that NAVOY® Sepsis outperforms all of these algorithms when comparing results from retrospective studies [10,20-23]. The Sepsis Sniffer [20] has comparable sensitivity and specificity, but a considerably lower positive predictive value than NAVOY® Sepsis. St. John Sepsis Surveillance Agent [21] and the TREWScore [22] have lower AUROC than that demonstrated for NAVOY® Sepsis on retrospective data, and the Epic Sepsis model [23] has a substantially lower AUROC. Only the Epic Sepsis Model has been evaluated clinically [23], showing poor discrimination, and all of these algorithms are only available with a specific electronic health record (EHR) system. Thus, NAVOY® Sepsis is the first stand-alone sepsis prediction algorithm to be clinically validated.

4.2. Strengths and limitations

To the best of our knowledge, this was the largest study at the time involving prospective use of a sepsis prediction algorithm in an ICU setting. The participating site was a general ICU which sees a wide range of patients in need of intensive care and with varying degrees of severity. This, together with the relatively open inclusion criteria, enabled the study population to resemble the normal ICU population at this site which improves the generalizability of the observed predictive performance. Furthermore, a considerable volume of patient data was collected in the study. The reliance on manual input did not result in any substantial loss of data or data quality, providing reliable results on the primary outcome. The availability of real-time data would likely be even greater if data transfer would occur automatically. Further, a portion of the patients were diagnosed with a positive SARS-CoV-2 infection and the algorithm performance was similar also on this subpopulation, noting however that the number of these patients is relatively small.

The study was conducted at a single hospital which limits the external validity. Secondly, the number of patients that could be used for the primary analysis was substantially lower than the planned number of patients. This led to wider confidence intervals than aimed for, but the results still support the findings from the retrospective proof-of-concept study [10]. Thirdly, the selection of Sepsis-3 diagnostic criteria for determining which patients are labelled as septic is a simplification of clinical practice, as the Sepsis-3 diagnostic criteria are used at the clinician’s discretion. Thus, in some proportion of cases where the observed

---

**Fig. 2.** Receiver operating characteristic (ROC) for algorithm predicting sepsis 3 h before sepsis onset (n = 88). True positives (sensitivity) are patients that develop sepsis that are accurately being predicted by NAVOY® Sepsis three hours before the onset of sepsis, and false positives (1 - specificity) are patients that don't develop sepsis that are wrongly being predicted by the algorithm.
clinical parameters and actions such as ordering cultures and initiation of antibiotics may correspond to sepsis as per Sepsis-3, the underlying cause could have been a sepsis mimic. Lastly, some of the collected bedside data were entered retrospectively with several hours’ delay, leading to a proportion of sepsis alerts appearing with delay, limiting their usefulness for clinical decisions and thus having an impact on the interpretability of secondary outcomes.

4.3. Future work

Future studies should investigate the impact of NAVOY® Sepsis on important clinical outcomes as well as ICU and hospital length of stay when implemented and used as a clinical decision support in a real-life setting, along with its prediction properties at various time points before sepsis onset. Also, the algorithm’s performance when used outside of the ICU remains unknown and further studies are needed in other healthcare settings.

5. Conclusions

The prognostic accuracy of NAVOY® Sepsis in a prospective ICU setting was validated in the current study, showing that it can serve as a relevant clinical decision support if integrated into routine care. This is the first sepsis prediction tool that has been clinically validated in the ICU.

Financial disclosure

The work was funded by the Swedish innovation agency Vinnova, and co-funded by AlgoDx AB and the Department of Statistics at Uppsala University.

CRediT authorship contribution statement

Inger Persson: Conceptualization, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. Andreas Macura: Software, Investigation, Writing – original draft, Writing – review & editing, Project administration. David Becedas: Conceptualization, Resources, Writing – review & editing, Supervision, Project administration. Fredrik Sjovall: Conceptualization, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

David Becedas and Inger Persson are shareholders of AlgoDx AB, and Andreas Macura has received stock options from AlgoDx AB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcrc.2023.154400.

References