Peer Review File

Article information: https://dx.doi.org/10.21037/tp-22-510

Reviewer A

In the manuscript "Effective diagnosis of sepsis in critically ill children using Probabilistic Graphical Model" the authors document their attempt to create a model capable of predicting pediatric sepsis. This computer modeling application has been used successfully to predict other medical conditions such as community-acquired pneumonia and adult sepsis. A successful predictive model in pediatric sepsis would be both novel and widely useful in clinical practice. The authors use a probabilistic graphical model method known as Tree Augmented Naïve Bayes to construct their predictive model. The model incorporates clinical selected variables from four domains: vital signs, clinical symptoms, laboratory data, and microbiological tests. Data from the publicly available Pediatric Intensive Care Dataset was used to build, test, and validate the models. The model using variables from all four domains performed the best but still demonstrated mixed results. All models demonstrated high levels of specificity and negative predictive value but rather low levels of sensitivity and positive predictive value. This has been a common finding with other reported predictive models for pediatric sepsis in the literature. Please see my comments, questions, and critiques below regarding the manuscript

Cover page:

Comment 1: ♣ Author contributions (lines 14-21)

o No author is listed to contributing to data analysis and interpretation. Who was responsible for the computer modeling, interpretation of the results, and final model selection?

Reply 1: The first author was responsible for the computer modelling. The results were interpreted and analyzed by all authors. We have edited the author contribution for data analysis and interpretation to "All authors" (see Cover page, page 2, line 32).

Changes in the text: Data analysis and interpretation: All authors (see Cover page, page 2, line 32).

Introduction:

Comment 2: ♣ Lines 75-78.

o Only the first aim is sufficiently described and discussed in the manuscript. No results are provided for aim #2 with relation to the biomarkers. Information related to aim #3 is not presented in either the results or discussion sections. More information should be included in the manuscript related to these aims or they should be removed from the introduction.

Reply 2: We have removed aims #2 and #3 in the Introduction and in the Background of Abstract.

Changes in the text: Changes were made in the Background of Abstract (see page 3, line 38-39) and in the Introduction (see page 5, line 77-79). The aim of the study

was amended to "**We employed Tree Augmented Naive Bayes (TAN), a PGM method, to develop diagnosis models to test our hypothesis and investigate the effectiveness of PGM in pediatric sepsis diagnosis.**"

Methods:

Comment 3: ♣ Study Design (line 83)

o The study used data from the Pediatric Intensive Care Dataset. This would then imply that only patients within the PICU were used in the creation of the model. Sepsis is not a diagnosis confined to the intensive care unit though. How applicable would this model be to patients on acute care (regular) floors or in intermediate care units?

Reply 3: As rightly pointed out by the reviewer, we only considered children in the PICU (based on the data available in the dataset), and the developed models were trained only on the data of patients admitted to the PICU. Therefore, it may not be applicable to other clinical settings such as acute care, intermediate care. We have added this point as a limitation in the discussion.

Changes in the text: The following sentences were inserted to the Discussion (see page 14, line 292-297): "**Furthermore, our study only extracted the children in the ICUs and the diagnostic models were trained solely on their clinical characteristics. Therefore, our models may not be applicable to other clinical settings outside of the ICU (e.g., acute care floors or intermediate care units). To apply the model in these settings, additional amendment and re-training would be required.**"

Comment 4: ♣ Sepsis Definition (lines 94-99)

o ICD-10 codes alone have been shown to have limited utility when identifying sepsis patients from a dataset and will miss certain patients. (Lindell, et al. Comparison of methods for identification of pediatric severe sepsis and septic shock in the virtual pediatric systems database. Crit Care Med. 2019;47(2): e129-135). Could you comment on what impact on not including these patients would have on the validity of the model?

Reply 4: We agree with the reviewer that the ICD-10 has limited utility to identify sepsis patients from a dataset. Using ICD-10 potentially misses patients with bacteremia and other subgroups of sepsis (e.g., viral and fungal sepsis) (as described in line 279). The main impact on not including these patients would be the limited size of sepsis cohort, causing the imbalance problem in supervising learning, where one class dominates the other (lines 274-277). As a result, the diagnosis model will suffer from poor predictive capability for the minority class as there is not enough information of it to learn from. Moreover, there is a high likelihood that model will not be able to recognize these missing sepsis cases as the model has never been exposed to them before.

Changes in the text: We have made amendment to page 13, line 278-281 in the discussion to elaborate the limitation of using ICD-10 to identify sepsis patients from our dataset: "**First, by using the ICD-10 code, we may have missed cases of patients with bacteremia (ICD-10: R78.81) and other subgroups of sepsis (e.g., viral and fungal sepsis). This may have led to a small number of sepsis cases, causing the problem of imbalanced dataset.**".

♣ Data extraction

Comment 5:
o **Line 109 –** If each hospitalization was treated as an independent event, how many subjects were included in the dataset more than once?
Reply 5: There were 52 patients included in the study cohort more than once with 108 different hospitalizations.
Changes in the text: We have included this information in the results section, page 9, line 167-168: "**Of these, 52 patients were admitted more than once (with 108 unique admissions).**"

Comment 6:
o **Line 114 –** Earlier versions of Microsoft Excel had significant flaws noted in statistical algorithms. Please include what version of Excel was used for analyses.
Reply 6: We used Microsoft Excel 16.55 for statistical analysis.
Changes in the text: We have included Microsoft Excel version in the text, at line 117, page 7: "**All analyses were performed in Microsoft Excel (version 16.55) with a statistical significance taken as p <0.05.**"

Comment 7:
o Were other medical diagnoses known to be risk factors for sepsis such as oncologic disorders or immunodeficiencies given any additional weight in the modeling?
Reply 7: We did not include comorbidities as a variable in this study. However, we will consider the risk factors and comorbidities in the next study. We have included this point as a limitation in our discussion.
Changes in the text: We have added the following limitation in the discussion, page 14, line 290-292: "**We did not consider comorbidities (e.g., oncological disorders, immunodeficiencies) as a variable in this study. Therefore, we could not investigate the effect of this variable in our study.**"

<u>Results:</u>
Comment 8: ♣ Lines 167-169 – Almost 10,000 admissions were excluded from the study for lack of clinical data or missing vital signs in the first 24 hours of the PICU admission. This accounts for almost 75% of the admissions in the database. Why such a large number? Could the authors comment on quality control methods incorporated by the operators of the database? What is the inter-rater reliability? Are objective third-party audits of the data performed?
Reply 8: The following admissions were excluded (line 164- 166): "492 that were not the first PICU admissions, 609 that did not have PICU clinical data, and 9,334 that did not have vital signs within 24 hours of the PICU stay". It was documented in the Pediatric Intensive Care Dataset (PICD) that vital signs were collected intermittently and manually by nurses and not all vital signs were measured at all time points. Therefore, we observed a lot of missing data, especially in vital sign data. As of the current version, the authors are facing several challenges in integrating and processing the data and are working on releasing better quality data in the next version. There is no information on the inter-rater reliability and third-party audits from the authors. We addressed this point as a limitation in our revised discussion.
Changes in the text: The following sentences were inserted to page 14, line 298-301: "**As of the current version, the dataset owners are facing several**

challenges in integrating and processing the data and are working on releasing better quality data in the next version. There is no information on the inter-rater reliability and third-party audits from them."

Comment 9: ♣ Lines 172-178 – the authors focus on bacterial sources of sepsis in the manuscript, but sepsis is not synonymous with a bacterial infection. What about other microbiological causes of sepsis such as viruses, fungus, etc.
Reply 9: There were 55 cases of viral sepsis labeled under unspecified sepsis and no case of fungal sepsis (ICD-10: B37.7, Candida sepsis). The cause of majority of viral sepsis cases is Pneumonia (n=43, ICD-10: J18.9, J18.0, P23.9, P23.5). Nevertheless, bacteria caused most cases of sepsis in this dataset. We included all these cases in our cohort of patients with sepsis. To clarify this, we have added the cases of viral and fungal sepsis to the result section.
Changes in the text: We have made amendment to page 9, line 169-173: "**Of these unspecified sepsis cases**, **there were 55 cases of viral and no case of fungal sepsis found (ICD-10: B37.7, Candida sepsis). The most common source of viral sepsis was pneumonia (43, 32%, ICD-10: J18.0, J18.9, P23.5, P23.9)**".

<u>Discussion:</u>
Comment 10: ♣ Lines 296-298 – As the authors note, ICD-10 codes are only assigned at the time of hospital discharge. Only data from the first 24 hours of the PICU admission was used to create and validate the models. This point in time may not be the period in which the patient experienced sepsis during their hospitalization. How do the authors account for this?
Reply 10: Our sepsis cohort consisted of sepsis patients identified by the ICD-10 code (n=112), suspected infection with SIRS (n=8) and both (n=14) (Supplementary A, Figure 1). We considered patients with suspected infections as those who had microbiological cultures sampled followed by antibiotic administration within 72 hours, or antibiotic administration followed by cultures taken within 24 hours (Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):762.).
We reviewed 112 patients in ICD10 diagnosis of sepsis to ensure that timing of sepsis is close to PICU admission. Of these, 88 patients satisfied our sepsis onset definition; the remaining 24 patients had either PICU admission diagnosis related to sepsis (e.g., bacterial sepsis, unspecified sepsis, pneumonia), microbiological cultures taken, and/or antibiotics taken within 24 hours of PICU admission. We have added this information to the methods section.
Changes in the text: The following sentences were added to the methods, sepsis definition section, page 6, line 101-106: "**We reviewed patients with ICD-10 diagnosis of sepsis to ensure that timing of sepsis is close to PICU admission. If the patient did not meet the sepsis onset definition, we examined the dataset to ensure that these patients had an admission diagnosis related to sepsis (e.g., bacterial sepsis, unspecified sepsis, pneumonia), microbiological cultures taken, and/or antibiotics taken within 24 hours of PICU admission.**"

<u>Conclusion:</u>

Comment 11:♣ Line 320-21 - I would argue that microbiological tests are NOT the gold standard in diagnosing sepsis. As discussed above, sepsis can be due to other microbiological causes not captured routinely by culture. Sepsis is a heterogenous clinical syndrome with no gold standard in diagnosis which is what makes it so difficult to identify in research and clinical practice.

Reply 11: We agree with the reviewer that microbiological tests are not considered as the gold standard in diagnosing sepsis anymore like it used to be several years ago. However, using microbiological tests is still a common practice to confirm sepsis in children as other symptoms such as clinical appearances, vital signs are non-specific in children. The International Pediatric Sepsis Consensus Definition in 2005 remains a popular criterion to identify sepsis children until recently. As there are diverse opinion about this, we have removed the "gold standard" from our conclusion.

Changes in the text: We have amended the following sentences in conclusion, page 15, line 313: "**Microbiological tests were unreliable due to the high negative incidence~~, despite being the gold standard in diagnosing sepsis~~.**"

Table B.2

Comment 12: ♣ It is unclear how certain clinical symptom variables are related to sepsis: moan, scream, quiet, regurgitation, heart murmur, chest tightness or pain.

Reply 12: The clinical symptoms were collected from different sepsis literature. For example, Mahallei et al. (2018) has describe some common clinical symptoms in sepsis children as following: Moaning, lethargy, and feeding problems, Fever, poor feeding, neonatal icterus, tachypnea, and respiratory distress. Table B.2 has been moved to main text as Table 1, and we have added the references for the symptoms at the footnote of the table.

Changes in the text: The following references have been added:

(1) Mahallei M, Rezaee MA, Mehramuz B, Beheshtirooy S, Abdinia B. Clinical symptoms, laboratory, and microbial patterns of suspected neonatal sepsis cases in a children's referral hospital in northwestern Iran. Medicine (Baltimore). 2018 Jun;97(25):e10630. doi: 10.1097/MD.0000000000010630. PMID: 29923969; PMCID: PMC6024470.

(2) Launay, Elise MD; Gras-Le Guen, Christèle MD, PhD; Martinot, Alain MD, PhD; Assathiany, Rémy MD; Blanchais, Thomas MD; Mourdi, Nadjette MPH; Aouba, Albertine MD; Bouvier-Colle, Marie-Hélène PhD; Rozé, Jean-Christophe MD, PhD; Chalumeau, Martin MD, PhD. Suboptimal care in the initial management of children who died from severe bacterial infection: A population-based confidential inquiry*. Pediatric Critical Care Medicine: July 2010 - Volume 11 - Issue 4 - p 469-474 doi: 10.1097/PCC.0b013e3181ce752e

(3) Santos Silva, E., Moreira Silva, H., Catarino, C. et al. Neonatal cholestasis: development of a diagnostic decision algorithm from multivariate predictive models. Eur J Pediatr 180, 1477–1486 (2021). https://doi-org.libproxy1.nus.edu.sg/10.1007/s00431-020-03886-z

(4) Hammett, E. Can you spot the signs and symptoms of sepsis? BDJ Team 6, 8–10 (2019). https://doi.org/10.1038/s41407-019-0123-5

(5) Ka Hong Chan, MD, Shubhayan Sanatani, MD, James E Potts, PhD, Kevin C Harris, MD MHSc, The relative incidence of cardiogenic and septic shock in neonates, Paediatrics & Child Health, Volume 25, Issue 6, October 2020, Pages

372–377, https://doi-org.libproxy1.nus.edu.sg/10.1093/pch/pxz078
(6) Riley, C., Basu, R.K., Kissoon, N. et al. Pediatric Sepsis: Preparing for the Future Against a Global Scourge. Curr Infect Dis Rep 14, 503–511 (2012). https://doi-org.libproxy1.nus.edu.sg/10.1007/s11908-012-0281-5

Comment 13: ♣ Tachycardia and bradycardia are not respiratory symptoms
Reply 13: We meant to classify them as symptoms of respiratory distress, where the abnormal heartbeat indicates respiratory distress. However, we agree with the reviewer that these are better classified under cardiovascular symptoms. As such, we have re-classified them to the cardiovascular symptoms and updated our results. Generally, we observed little changes in the results and our results/discussion still holds.
Changes in the text: We updated the results in Table B.2, Table B.4, Table 2, Figure 2, Figure 3. Table B.2, B.4 were moved to main text as Table 1, Table 4 and Table 2 is amended as Table 5.

Table 1: List of clinical variables for sepsis diagnosis.

Variable groups	**Variables**
Respiratory symptoms	Apnea, cyanosis, respiratory distress, cough, phlegm, sputum, wheezing, dyspnea, expectoration, anhelation, asphyxia
Cardiovascular symptoms	Heart failure, heart murmur, pericarditis, endocarditis, myocarditis, ventricular, chest tightness, chest pain, **tachycardia, brachycardia**

Table 4 Distribution of clinical variables within 24 hours of PICU admission

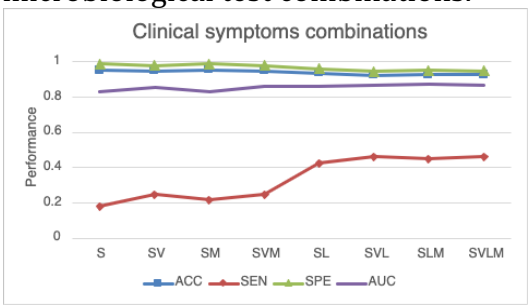
Variables	**Overall (N=3,014)**	**Sepsis Cohort (N = 134)**	**Non-Sepsis Cohort (N = 2,880)**	**P-value**
Presence of respiratory symptoms, *n (%)*	577 (19.1)	82 (61.2)	495 (17.2)	<0.001
Presence of cardiovascular symptoms, *n (%)*	431 (14.3)	49 (36.6)	377 (13.1)	<0.001

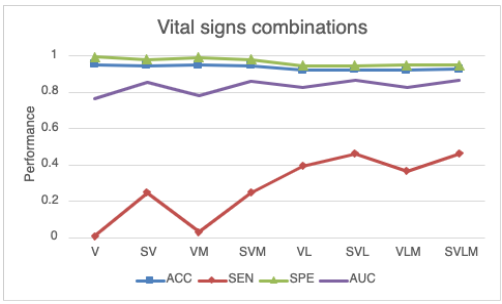
Table 5: Performance of the SLVM model in different sub-groups of the cohort

Group/Performance	**ACC (95% CI)**	**SEN (95% CI)**	**SPE (95% CI)**	**AUC (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Premature infants	0.766 (0.665-0.844)	0.483 (0.299-0.671)	0.892 (0.785-0.952)	0.736 (0.622-0.520)	0.667 (0.431-0.845)	0.795 (0.681-0.877)
Term infants	0.873 (0.840-0.899)	0.412 (0.293-0.550)	0.932 (0.904-0.952)	0.817 (0.750-0.884)	0.446 (0.316-0.584)	0.924 (0.895-0.946)

Age < 30 days	0.889 (0.860-0.913)	0.338 (0.231-0.464)	0.961 (0.940-0.976)	0.783 (0.716-0.850)	0.534 (0.378-0.685)	0.917 (0.890-0.938)
Age between 1m to 1yr	0.960 (0.944-0.971)	0.281 (0.144-0.470)	0.986 (0.974-0.992)	0.889 (0.814-0.964)	0.429 (0.226-0.656)	0.973 (0.959-0.982)
Age > 1yr	0.967 (0.957-0.975)	0.235 (0.114-0.416)	0.983 (0.976-0.989)	0.883 (0.809-0.957)	0.242 (0.117-0.426)	0.983 (0.975-0.989)
General wards	0.973 (0.965-0.979)	0.213 (0.112-0.361)	0.989 (0.983-0.993)	0.853 (0.785-0.921)	0.286 (0.152-0.465)	0.984 (0.977-0.988)
Emergency Units (ICU, PICU, NICU, SICU)	0.851 (0.822-0.876)	0.241 (0.159-0.347)	0.936 (0.913-0.953)	0.789 (0.731-0.847)	0.344 (0.230-0.478)	0.899 (0.872-0.920)
24 hours data cut-off	0.930 (0.916-0.936)	0.463 (0.376-0.550)	0.949 (0.940-0.956)	0.866 (0.826-0.906)	0.295 (0.235-0.362)	0.974 (0.968-0.980)
48 hours data cut-off	0.946 (0.918-0.936)	0.469 (0.377-0.550)	0.953 (0.941-0.957)	0.867 (0.828-0.906)	0.301 (0.239-0.368)	0.974 (0.968-0.980)
TAN	0.930 (0.916-0.936)	0.463 (0.376-0.550)	0.949 (0.940-0.956)	0.866 (0.826-0.906)	0.295 (0.235-0.362)	0.974 (0.968-0.980)
LR	0.950 (0.941-0.967)	0.130 (0.05-0.321)	0.990 (0.982-0.998)	0.560 (0.509-0.611)	0.450 (0.250-0.661)	0.960 (0.949-0.977)

Figure 2: Performance of TAN diagnosis models: (a) clinical symptoms combinations, (b) vital signs combinations, (c) laboratory test combinations, (d) microbiological test combinations.





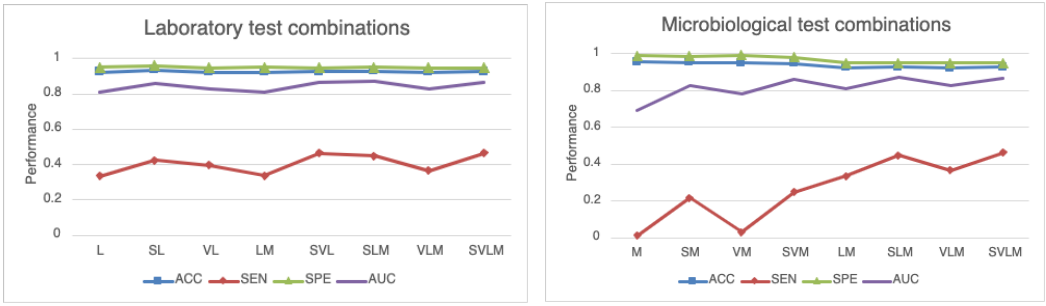
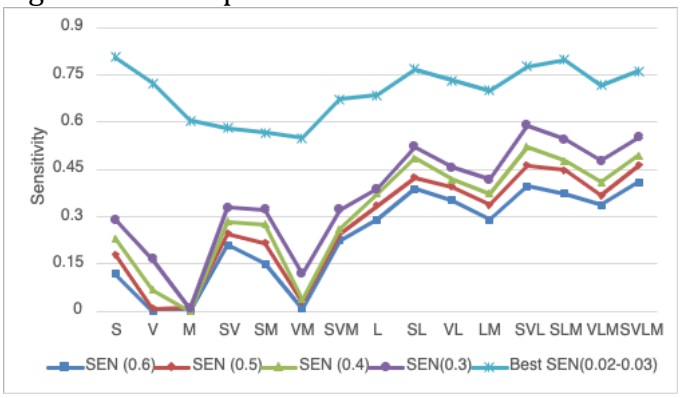


Figure 4: Model performance with different decision thresholds.



Comment 14:♣ Microbiological culture tests – Were any tests for viruses such as pcr included in the analyses?
Reply 14: We found no information on the PCR procedure recorded for Microbiological culture test in the dataset.
Changes in the text: No change

<u>Table B.4</u>
Comment 15: ♣ The vital sign data is difficult to interpret because it is not grouped by ages. It would be better to list the percentages of individuals that fell into the "high", "low", and "normal" categories as this is how the variable was used in the modeling.
Reply 15: As suggested, we have included the percentages of individuals that fell into the "high", "low", and "normal" categories for vital sign variables.
Changes in the text: We have added the information to Table 4 as follows.

Variables	**Overall (N=3,014)**	**Sepsis Cohort (N = 134)**	**Non-Sepsis Cohort (N = 2,880)**	**P-value**
Median heart rate, bpm (IQR)	140 (121-156) High:807 Low: 23 Normal: 2,184	158 (144-169.5) High: 49 Low: 1 Normal: 84	138 (120-155) High:758 Low:22 Normal: 2,100	<0.001
Median respiratory rate, /min (IQR)	34 (28-56) High: 576 Low: 352 Normal: 2,086	52 (42-56) High:40 Low:1 Normal: 93	34 (28-43) High:536 Low: 351 Normal:	<0.001

			1,993	
Median temperature, °C (IQR)	37.3 (37-37.7) High: 209 Low: 33 Normal: 2,772	37.1 (36.9-37.7) High:21 Low: 2 Normal: 111	37.3 (37-37.7) High:188 Low: 31 Normal: 2,661	0.068
Median oxygen saturation, % (IQR)	99 (98-100) Normal:2,413 Low: 601	96 (92-100) Normal:94 Low: 40	99 (98-100) Normal: 2,319 Low: 561	<0.001
Median SBP, mmHg (IQR)	98 (85-110) High:797 Low: 557 Normal: 1,660	76 (58.5-95.5) High: 18 Low: 62 Normal: 54	99 (86-95.5) High:779 Low: 495 Normal: 1,606	<0.001
Median DBP, mmHg (IQR)	56 (46-66) High:1,049 Low: 539 Normal: 1,426	43 (32-58) High:40 Low: 53 Normal: 41	56 (46-66) High:1,009 Low:486 Normal: 1,385	<0.001

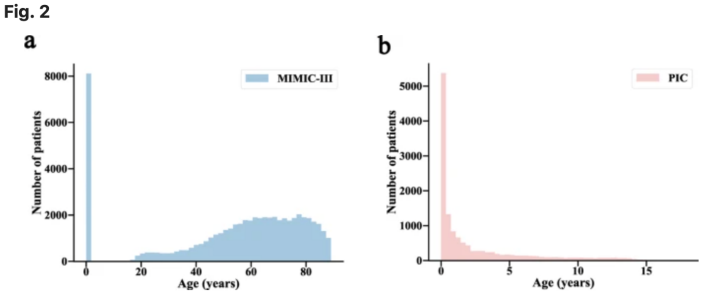
* Vital signs were categorized into "high", "low", and "normal" based on age group following cut-offs provided from (14) (Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics*: Pediatric Critical Care Medicine. 2005 Jan;6(1):2–8)

Table 1

Comment 16: ♣ Median age – If the study included subjects 0-18 years of age, why did the included the subjects skewed so low in age. The 75th percentile for age is only 4.3 years of age in all subjects and just 1.02 in subjects with sepsis.

Reply 16: As reported by the PICD, majority of the patients were below 1 years old. The mean age of the patients was 2.5 years (Q1–Q3: 0.1–3.3). Please see the below picture, cited from Zeng, X., Yu, G., Lu, Y. et al. PIC, a paediatric-specific intensive care database. Sci Data 7, 14 (2020). https://doi.org/10.1038/s41597-020-0355-4

Fig. 2



Age distributions of the MIMIC-III and the PIC databases. (**a**) MIMIC-III database. (**b**) PIC database.

Full size image >

Figure is cited from Zeng et al. (2020), (Zeng, X., Yu, G., Lu, Y. et al. PIC, a paediatric-specific intensive care database. Sci Data 7, 14 (2020). https://doi.org/10.1038/s41597-020-0355-4)
Changes in the text: No change.

Overall summary: This manuscript represents a worthwhile pursuit to create a predictive model for pediatric sepsis using probabilistic graphical modeling, but it is limited by significant limitations in the source data set used to create and validate the model. At best the model generated is a tool for predicting severe sepsis or septic shock from bacteria in a PICU population of infants, but the sensitivity of the model is low even in this narrow population.

Reviewer B

This is an excellent paper. Thanks for the opportunity to review it.
I suggest some modifications to make text clear to reader.

Comment 1:
1) insertion of Figure 2 - Tan model with full variables from four categories, supplemental material A
Reply 1: Thanks for the suggestion, we have moved the figure of TAN model with full variables from four categories to the main text as Figure 3 and amended the references accordingly in the manuscript.
Changes in the text: The figure of TAN model was moved to main text as Figure 3.

Comment 2:
2) insertion of table B2, B3 and B4 of supplemental material B
Reply 2: As suggested, we have moved table B2, B3, and B4 of supplemental material B to the main text and amended the references accordingly in the manuscript.
Changes in the text: Table B2, B3, B4 were moved to main text as Table 1, Table 2, and Table 4.
