JAMA Network Open

Invited Commentary | Infectious Diseases

# Extracting Symptoms of Sepsis From Admission Notes With LLM—Maybe AI Knows Something That We Do Not

Jonathan D. Baghdadi, MD, PhD; M. Cristina Vazquez-Guillamet, MD

+ Related article

Author affiliations and article information are listed at the end of this article.

Clinicians hate sepsis because it is common, associated with a high risk of mortality, and hard to define. Almost any symptom can be associated with sepsis,[1] which means no symptom can be used to rule the diagnosis in or out. Basically, if a clinician suspects a patient might have sepsis, then that is what they have.

At the same time, the presenting symptoms associated with sepsis have clinical implications.[2] By pointing toward the underlying site and source of infection, symptoms guide what imaging is performed, which antibiotics are started, and whether an invasive procedure might be required for source control. The easiest way to appreciate the significance of symptoms is to consider how sepsis is managed when the patient history is unavailable or unrevealing. Most patients with undifferentiated sepsis receive a combination of anti-MRSA (methicillin-resistant *Staphylococcus aureus*) and antipseudomonal antibiotics, such as intravenous vancomycin and piperacillin-tazobactam, which is almost never justified based on what eventually grows from culture.[3]

Thus, a tool that efficiently discerns the signs and symptoms of sepsis could be useful for understanding optimal approaches to early sepsis care. Pak et al[4] evaluated the feasibility of using LLaMA 3, a publicly available large language model (LLM), for this task. Their work, which provides a meticulous description of their methods, demonstrates that the LLM can detect and extract signs and symptoms of severe infection from an admission history of present illness with equivalent accuracy to a physician conducting medical record review. LLM classifications of signs and symptoms correlated with encounter-level *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification* diagnosis codes and presence of specific subtypes of resistant bacteria isolated in clinical cultures, further establishing the construct validity of the LLM output.

At present, applications of the methods explored in the study by Pak et al[4] seem best suited for research. Even the ACORN (Antibiotic Choice on Renal Outcomes) trial, which creatively used features of the electronic health record to enroll and randomize patients with sepsis, required a human to review the patient's medical record and identify the source of infection.[5] However, manual medical record review is time consuming and subjective. Automating that process with an LLM could enable similar studies to be performed at a much larger scale. Furthermore, sepsis is notoriously heterogeneous. Extracting key aspects of the patient history in real time, ideally prior to study enrollment, could enable prospective trials that set inclusion and exclusion criteria to align with specific sepsis phenotypes.

On the other hand, the potential clinical applications remain unclear. Predicting which antibiotic therapy a patient with sepsis needs is difficult.[3] Pak et al[4] successfully used the signs and symptoms extracted by the LLM to categorize patients into infectious syndromes, but recognizing that cough and dyspnea may represent pneumonia is likely within the capability of most clinicians. To guide more complicated treatment decisions, such as selecting empiric antibiotics based on risk of resistant pathogens, the LLM would also need to extract comorbidities, health care exposures, and prior treatments. Furthermore, validation in diverse patient populations would be required. Existing tools under development can estimate resistance reasonably well when built on cohorts with positive culture results. However, the positive predictive value of these tools decreases when applied to actual patients, many of whom have negative culture results. A model that overestimates the risk of resistance could perpetuate unnecessary use of broad-spectrum antibiotics.

Tools for clinical decision support are generally limited to using structured data elements from electronic health records, which can be inaccurate for relevant fields, such as prehospital origin.[6] Newer, sophisticated deep learning models may leverage the hierarchical organization and patterns of available structured data to overcome data quality issues and improve performance. With further development, LLM-based tools represent another compelling option. An LLM that uses all available medical information from clinical notes—including both unstructured data related to subjective symptoms and structured data such as medical history, vital signs, and laboratory results—may be superior to any tool that uses structured data alone.

Thus, we suspect that the use of LLM-based tools, both in research and clinical care, will eventually allow clinicians to improve clinical decision-making related to sepsis. But it is also important to acknowledge that overreliance on artificial intelligence (AI) could create problems. Based on tools currently available or in development, a nonhuman entity could guide or automate history taking, generation of a differential diagnosis, transcription of the encounter in a note, and subsequent clinical decision-making. At every step, we anticipate that AI could have a flattening effect, making patients and their stories appear more homogeneous than they really are. Clinicians and researchers want to believe that AI will reveal deep truths that would be otherwise imperceptible to the human mind, but a tool that summarizes and distills a nuanced patient narrative into a set of 5 to 10 symptoms is strictly focused on the surface level.

The most recent Surviving Sepsis Campaign guidelines[7] recommend initiation of antibiotic therapy, among other interventions, within 1 hour of recognition of possible septic shock or high likelihood of sepsis. To meet that time frame, clinical decision-making may be rushed and reflexive (eg, vancomycin and piperacillin-tazobactam). Thus, it is plausible to expect that an LLM might perform at or near the level of a human clinician when recognizing and responding to new-onset sepsis. However, despite strong performance on licensing examinations, LLMs in their current state may not be equipped to grapple with true diagnostic uncertainty. In simulated cases designed to mimic clinical scenarios, LLMs were unable to diagnose or follow guidelines for common presentations.[8] For now, AI seems better suited to automating simple tasks, such as the extraction of signs and symptoms, than participating in clinical decision-making.

## ARTICLE INFORMATION

**Published:** October 24, 2025. doi:10.1001/jamanetworkopen.2025.39275



**Corresponding Author:** Jonathan D. Baghdadi, MD, PhD, Infectious Diseases, University of Maryland School of Medicine, 10 S Pine St, Medical Student Teaching Facility, Ste 334, Baltimore, MD 21201 (jbaghdadi@som.umaryland.edu).

**Author Affiliations:** Infectious Diseases, University of Maryland School of Medicine, Baltimore (Baghdadi); Institute for Health Computing, University of Maryland, Baltimore (Baghdadi); Washington University in St Louis School of Medicine, St Louis, Missouri (Vazquez-Guillamet).

**Conflict of Interest Disclosures:** Dr Vazquez-Guillamet reported receiving personal fees from Merck outside the submitted work. No other disclosures were reported.

## REFERENCES

**1**. Yang JM, Tisherman SA, Leekha S, et al. What clinicians think about when they think about sepsis: results from a survey across the University of Maryland Medical System. *Crit Care Explor*. 2024;6(12):e1183. doi:10.1097/CCE.0000000000001183

**2**. Filbin MR, Lynch J, Gillingham TD, et al. Presenting symptoms independently predict mortality in septic shock: importance of a previously unmeasured confounder. *Crit Care Med*. 2018;46(10):1592-1599. doi:10.1097/CCM.0000000000003260

**3**. Vazquez Guillamet MC, Liu H, Atkinson A, Fraser VJ, Lu C, Kollef MH. Performance of risk models for antimicrobial resistance in adult patients with sepsis. *JAMA Netw Open*. 2024;7(11):e2443658-e2443658. doi:10.1001/jamanetworkopen.2024.43658

**4**. Pak TR, Kanjilal S, McKenna CS, Hoffner-Heinike A, Rhee C, Klompas M. Syndromic analysis of sepsis cohorts using large language models. *JAMA Netw Open*. 2025;8(10):e2539267. doi:10.1001/jamanetworkopen.2025.39267

**5**. Qian ET, Casey JD, Wright A, et al; Vanderbilt Center for Learning Healthcare and the Pragmatic Critical Care Research Group. Cefepime vs piperacillin-tazobactam in adults hospitalized with acute infection: the ACORN randomized clinical trial. *JAMA*. 2023;330(16):1557-1567. doi:10.1001/jama.2023.20583

**6**. Goodman KE, Taneja M, Magder LS, et al. A multi-center validation of the electronic health record admission source and discharge location fields against the clinical notes for identifying inpatients with long-term care facility exposure. *Infect Control Hosp Epidemiol*. 2024;45(8):979-984. doi:10.1017/ice.2024.37

**7**. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*. 2021;47(11):1181-1247. doi:10.1007/s00134-021-06506-y

**8**. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nat Med*. 2024;30(9):2613-2622. doi:10.1038/s41591-024-03097-1