Infection Control & Hospital Epidemiology (2026), **47**, 574–577
doi:10.1017/ice.2026.10429

SHEA

Original Article

Generative artificial intelligence for surgical site infection surveillance

Shatha Alshanqeeti MBBS[1], Aileen de Guzman RN[2], Mary K. Riley RN[2], K.C. Coffey MD, MPH[2,3], Kathrine E. Goodman PhD, JD[3,4], Anthony D. Harris MD, MPH[2,3,4], Jonathan Baghdadi MD, PhD[3,4], Lisa Pineles MA[3], Westyn Branch-Elliman MD, MMSc[5,6] and Daniel J. Morgan MD, MS[2,3]

[1]Institute for Human Virology, University of Maryland School of Medicine, Baltimore, MD, USA, [2]Division of Epidemiology, VA Maryland Health Care System, Baltimore, MD, USA, [3]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA, [4]University of Maryland Institute for Healthcare Computing, Bethesda, MD, USA, [5]Section of Infectious Disease, VA Greater Los Angeles HCS-West LA, Los Angeles, CA, USA and [6]Department of Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, USA

Abstract

Background: Surgical site infection (SSI) surveillance can be time consuming and resource intensive. This study investigates the potential of generative artificial intelligence (GenAI) to augment the detection and classification of SSIs.

Methods: A case control study of patients with SSI following spine surgery at one US hospital. SSIs were classified into superficial, deep, and organ space. All SSIs were confirmed by infection prevention (IP) experts as they occurred from October, 2023 to September, 2025 and matched 1:1 by year to surgeries deemed non-SSI. A secure GenAI was used to determine if patients had an SSI based on standardized prompts and clinical data. IP nurses used GenAI output to review cases with the ability to ask GenAI questions within the data provided or independently open the medical record. We compared GenAI determinations to initial IP nurses' determinations.

Results: A total of 555 patients had spine surgeries. All 16 SSIs were matched by year to 16 non-SSI. All SSIs were correctly identified by GenAI (sensitivity 100%, 16/16) and only 1 non-SSI was incorrectly identified as SSI (specificity 93.7%, 15/16). Although GenAI accurately identified all SSI cases, it was discordant with original review at classifying the level of infection in 37.5% (6/16) of cases. Upon final IP physician review, GenAI was correct in 66.7% (4/6) of discordant cases (often determining "organ space infections" rather than "deep"). Median time to complete GenAI assisted SSI reviews was 9 minutes (IQR 7–21).

Conclusion: GenAI is a promising tool to assist in SSI surveillance following spinal surgery that could improve efficiency.

(Received 12 November 2025; accepted 15 February 2026; electronically published 21 April 2026)

Background

Surgical site infections (SSI) account for 20% of healthcare-associated infections (HAI) and are associated with patient morbidity, mortality, and healthcare costs.[1] SSI following instrumented spinal surgery pose significant risks including need for unplanned surgical interventions, neurological injury, and prolonged hospitalization.[2–4] Surveillance and reporting are resource-intensive, with up to 45% of infection prevention (IP) expert time being spent on HAI surveillance.[5] SSI surveillance definitions are standardized but heavily rely on subjective criteria such as clinical signs and symptoms and clinician diagnosed infection making it difficult to automate compared to other HAI. Additionally, such subjectivity can result in inter-facility variation in reporting across different sites.[6] Generative artificial intelligence (GenAI) has shown potential to detect SSI after colon surgery in a set of abstracted clinical data.[7] To the best of our knowledge, there are no studies evaluating performance of GenAI for detecting SSI following instrumented spinal surgeries using clinical data. We performed a single center case control study evaluating GenAI compared to IP nurse review for instrumented spine SSI detection and classification.

Methods

We included all spinal surgeries performed at a 124-bed Veterans Affairs acute care hospital in Baltimore, Maryland from October 1st, 2023 to September 1st, 2025. We used IP surveillance database to identify all instrumented spine surgical procedures that occurred during the same period and were previously reviewed by IP experts. Each case of SSI was matched with a control patient who underwent a similar surgery without identification of an SSI. Patient data from each case was extracted and added to a single continuous document in the ".docx" format. Data extracted included clinical notes, radiology reports, all vital signs, and microbiological results during admission. Clinical notes were restricted to the following due to Chat GPT-4o data token limits: emergency department physician and triage notes, history and physical exam notes, all primary surgical and medical teams' physician and nursing notes, all physician and nursing operative notes, and consultation notes by infectious disease, orthopedics, and neurosurgery. We used previously described prompts[8,9] and summaries of United States Centers for Disease Control and Prevention (CDC) National Health and Safety Network

Corresponding author: Shatha Alshanqeeti; Email: salshanqeeti@ihv.umaryland.edu
Cite this article: Alshanqeeti S, de Guzman A, Riley MK, *et al.* Generative artificial intelligence for surgical site infection surveillance. *Infect Control Hosp Epidemiol* 2026. 47: 574–577, doi: 10.1017/ice.2026.10429

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Society for Healthcare Epidemiology of America.

https://doi.org/10.1017/ice.2026.10429 Published online by Cambridge University Press

(NHSN) definitions[1] to develop GenAI prompts (Appendix A). Definitions use a combination of objective criteria such as fever and positive cultures as well as subjective findings like clinical symptoms or a diagnosis by a physician. SSI are classified into superficial incisional, deep incisional, and organ space infections. Superficial infections occur within 30 days of surgery and do not require surgical intervention. Deep and organ space infections occur within 90 days of surgery and require a surgical intervention. Organ space SSI extends beyond the level of fascia and muscle while deep incisional SSI does not. OpenAI GPT 4o model was accessed through the VA Office of Digital Health (VA GPT). All work was performed using the secure VA GPT 4o model, which does not retain patient data and is compliant with protected healthcare information regulations. Temperature measures the creativity and randomness of GenAI output[10] and was set to 0 in VA GPT to limit random output and hallucinations. Two IP nurses entered prompts and data into VA GPT in a standardized fashion. GenAI was asked to provide a general summary of the case and if SSI criteria were met, then provide a final SSI determination. IP nurses reviewed GenAI output for accuracy and consulted the medical record if needed. We designed a data entry form where all GenAI responses were entered and IP nurses performing reviews answered questions about the case and their perceptions on GenAI tool performance. Sensitivity, specificity, and positive and negative predictive values were calculated using manual review as the reference standard. Confidence intervals (CI) of 95% were used. This was a quality improvement assessment and deemed non-human subjects research by the University of Maryland Baltimore Institutional Review Board (IRB) and the local VA Research & Development committee.

Results

A total of 555 instrumented spinal surgery cases had been reviewed from 10/01/2023 to 09/01/2025 for presence of an SSI. These reviews were performed in real time by IP nurses as surgeries occurred followed by hospital epidemiology physician confirmation. From these surgeries, 16 (2.9%) were SSIs. We reviewed the 16 SSI cases and 16 selected controls without SSI matched by year. The patients were predominantly male (91%, 29/32; 14/16 in SSI group and 15/16 in the non-SSI group). Mean age was 66.2 years (SD 10.9) in the SSI group and 63.8 (SD 7.3) in the non-SSI group. Among the 32 patients, 53% (17/32) were white and 47% (15/32) were black, and 3% (1/32) reported Hispanic ethnicity. Of the surgeries, 40.6% (13/32) were cervical spine fusions, 46.9% (15/32) were lumbar fusions, and 12.5% (4/32) were thoracic and thoracolumbar fusions. Average length of stay for the index surgery admission was 22.6 days (SD 19.6) in SSI group and 6.1 days (SD 3.8) in non-SSI group (Table 1).

All 16 SSI were correctly identified by GenAI and one of 16 (6.3%) non-SSI control was incorrectly identified as SSI (false positive). In this instance, GenAI incorrectly reasoned that a non-SSI was an SSI due to evidence of a paraspinal fluid collection on MRI. However, the fluid collection was a CSF leak from a dural tear without radiological features to suggest infection with absence of clinical and microbiological criteria needed to establish SSI. The sensitivity of GenAI in detection of SSI was 100% (95% CI 79.4% to 100%) and specificity was 93.7% (95% CI 69.8% to 99.8%). At the fixed SSI prevalence rate of 50% in this study, the negative predictive value was 100% (95% CI 78.20% to 100.00%) and positive predictive value was 94.1% (95% CI 70.6% to 99.1%). Although GenAI accurately identified all SSI cases, it was discordant for depth of infection. Of 16 SSI cases, 1 was superficial and 15 were deep incisional SSI as deemed by IP experts. GenAI identified 11 deep infections and 5 organ space infections. SSI classification by GenAI did not match initial IP nurse classification in 37.5% (6/16) cases. When these cases were retrospectively reviewed by IP physician and nurses, GenAI classification was correct in 4/6 of discordant cases.

Table 1. Patient demographics

	All	SSI	Non-SSI
Gender	32	16	16
Male	29 (91%)	14 (88%)	15 (94%)
Female	3 (9%)	2 (12%)	1 (6%)
Race			
White	17 (53%)	7 (44%)	10 (63%)
Black	15 (47%)	9 (56%)	6 (37%)
Ethnicity			
Hispanic	1 (3%)	1(6%)	0
Not Hispanic	31 (97%)	15 (94%)	16 (100%)
Age at SSI date (years), mean (min/max)	65 (42, 89)	66.2 (42, 89)	63.8 (49, 76)
Length of hospital stay (days), mean (min/max)	14 (2, 80)	22.6 (3, 80)	6.1 (2, 15)

Median total time for GenAI review in SSI cases was 9 minutes (IQR 7–21) and 9 minutes (IQR 7–13) in non-SSI controls. Time to copy and paste patient record information into GenAI chatbot and receive SSI determination was a median of 4 minutes (IQR 3–7). Median time needed to review discordant cases was 22 minutes (IQR 9–34) compared to 9 minutes (IQR 7–13) in concordant cases. When IPs reviewed GenAI output, they asked GenAI additional questions in 8 cases, of which 7 were SSI cases. Only 1 case required further chart review during this process. Median time spent on manual review of GenAI output in cases that required additional questions or expert review was 6 minutes (IQR 4–12) and 6 minutes (IQR 4–11) when no additional questions or chart review were performed. Perceptions of the process were provided by IP nurses after each review. Subjective assessment of effort required for an SSI review with GenAI assistance was "moderate" in 50% (8/16) and "little effort" in 50% (8/16) of cases. Non-SSI cases required less effort with 18.7% (3/16) requiring "moderate effort," 50% (8/16) requiring "little effort," and 31.3% (5/16) requiring "no effort." No GenAI assisted reviews were reported to require "a lot of effort."

Discussion

In a single center case control study reviewing all SSIs over 4 years, we found GenAI identified all SSIs. GenAI was highly sensitive and specific and had a negative predictive value of 100%. GenAI differed with IP nurses on determining depth of infection, but on further review was more accurate than IP nurses. Reviews took a median of 9 minutes.

GenAI was able to identify all patients with SSI. This was likely due to most patients having many of the criteria needed to diagnose an SSI and all patients with SSI requiring a surgical intervention (while most patients without an SSI did not have a second surgery). Our findings are similar to a study by Flores-Balado et. al. where a natural language processing (NLP) tool created for specific

https://doi.org/10.1017/ice.2026.10429 Published online by Cambridge University Press

hospitals performed well in detecting SSI following hip replacement surgery.[11] In our study, all patients with SSI required a surgical intervention which likely increased identification as well as the likelihood of having a positive culture indicating infection.

In the one case where GenAI incorrectly detected an SSI, it reasoned that the presence of a fluid collection was indicative of an infection. Clinically this was considered a cerebrospinal fluid leak with lack of clinical and microbiological criteria for SSI and improvement without antibiotics, emphasizing the importance of expert review of GenAI output to confirm SSI. It is important to note, however, that the prompts used (Appendix) did not define the phrase "other evidence of infection" which may have contributed to this misclassification. We found GenAI was able to effectively extract and present key elements for SSI determination from clinical data that aids in expert review. For example, was purulent drainage identified? Did the patient have symptoms of SSI? etc. (see Appendix). This is similar to earlier studies using simulated cases[12,13] and a more recent studies done on catheter associated urinary tract infections[8] and central line associated infections[9] indicating that GenAI may be best used to identify specific data elements for human review rather than a fully automated review with determination of SSI or not.

GenAI classification of depth of SSI did not match the classification of IP nurse in a subset of cases. Notably, GenAI classification was correct in most cases on repeat review. This is likely because classification of deep vs organ space SSI does not correlate to severity of infection and reporting the level of infection is not required for reimbursement, potentially contributing to less thorough original reviews by IPs. Although differentiating between deep and organ space SSIs does not influence reimbursement, it is important for identifying mechanisms of infection and enabling targeted prevention strategies.

SSI is a publicly reported HAI that is tied to reimbursement in the US and elsewhere. Overall HAI surveillance and reporting is resource-intensive and constitutes up to 61% of IP expert time.[12,14] There is no standardized estimate of the time infection preventionists spend per case in the literature. Anecdotally, this time was estimated to range between 20 and 60 minutes by IP nurses. HAI identification has been reported as distracting from other IP activities by IP experts.[15] This could potentially be reduced with the use of GenAI as the time spent to review one case using GenAI ranged between 4 and 66 minutes with a median review time of 9 minutes. The time required for review could be further improved with automation of feeding patient data into GenAI (as cutting and pasting was the greatest time requirement). In this study, only 28.1% of cases required additional questions or chart review by IP nurses and of these, only one was not an SSI. Our findings support using GenAI as a screening tool or "first pass review" for SSI to augment the performance of expert reviewers. In this center, a total of 555 surgeries were reviewed by IP experts and only 2.9% were SSI. Given the high negative predictive value, if applied to all 555 cases with the same prevalence, GenAI could potentially reduce manual chart review by 97.1%. This is similar to a more resource intense NLP intervention by Flores-Balado et. al which found an 88.95% reduction in manual review of clinical records.[11] Furthermore, using GenAI for SSI surveillance may improve equity in resource-limited settings with few IPs.

Our study had limitations including using a single GenAI model (OpenAI-Chat GPT-4o) due to lack of other accessible HIPAA and privacy-compliant GenAI systems and unblinded review of IP nurses to determine necessary elements for SSI or not SSI from output. This study was conducted in the US VA system which has an older, more male population but where most patients receive their postoperative care within the same system.[16] While this minimizes loss of follow up, it potentially limits applicability to other systems. Notably this is a single-center study with a small sample size. However, this proof-of-concept study provides early evidence that GenAI can augment IP expert workflow potentially decreasing the time spent on SSI surveillance, which is one of the more difficult HAI to monitor.

Conclusion

Generative AI accurately identified spinal SSI compared to traditional IP nurses' review. Automated GenAI assisted human review could improve SSI metric identification and reporting.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/ice.2026.10429.

Financial support. No funding was received for this work.

Competing interests. KCC reports consulting fees from CARB-X and institutional research funding from Institute for Healthcare Improvement unrelated to this work. ADH reports consulting fees from UpToDate, Innoviva, and Pfizer unrelated to this work. JDB reports research grant support from AHRQ K08 and CDC SHEPheRD contract unrelated to this work. DJM reports research grant support from CDC, NIH, AHRQ and VA HSR unrelated to this work and reimbursement for planning or speaking at conferences from SHEA, IDSA, WHO, and ACP. Authors SA, AG, MKR, KEG, LP, WBE report no competing interests.

References

1. Centers for Disease Control and Prevention. National Healthcare Safety Network (NHSN) patient safety component manual: chapter 9 – surgical site infection (SSI) event. 2026. https://www.cdc.gov/nhsn/pdfs/pscmanual/9pscssicurrent.pdf Accessed March 13, 2026
2. Lupo BD, Jameson WP, Quinones CJ, *et al.* Risk factors and outcomes of surgical site infections of the spine: a retrospective multi-center analysis. *J Clin Med* 2025;14:3520. doi:10.3390/jcm14103520.
3. Edmiston CE, Leaper DJ, Chitnis AS, Holy CE, Chen BPH. Risk and economic burden of surgical site infection following spinal fusion in adults. *Infect Control Hosp Epidemiol* 2023;44:88–95. doi:10.1017/ice.2022.32.
4. Freire-Archer M, Sarraj M, Koziarz A, *et al.* Incidence and recurrence of deep spine surgical site infections: a systematic review and meta-analysis. *Spine* 2023;48:E269. doi:10.1097/BRS.0000000000004713.
5. Stone PW, Dick A, Pogorzelska M, Horan TC, Furuya EY, Larson E. Staffing and structure of infection prevention and control programs. *Am J Infect Control* 2009;37:351–357. doi:10.1016/j.ajic.2008.11.001.
6. Shenoy ES, Branch-Elliman W. Automating surveillance for healthcare-associated infections: rationale and current realities (Part I/III). *ASHE* 2023;3:e25. doi:10.1017/ash.2022.312.
7. Badia JM, Casanova-Portoles D, Membrilla E, Rubiés C, Pujol M, Sancho J. Evaluation of ChatGPT-4 for the detection of surgical site infections from electronic health records after colorectal surgery: a pilot diagnostic accuracy study. *J Infect Public Health* 2025;18:102627. doi:10.1016/j.jiph.2024.102627.
8. Alshanqeeti S, Coffey K, Mcdermott K, *et al.* Comparing generative artificial intelligence vs. experts for detection of catheter-associated urinary tract infection (CAUTI). *Clin Infect Dis* 2025;9:ciaf486. doi:10.1093/cid/ciaf486.
9. Morgan DJ, AlShanqeeti S, Coffey KC, *et al.* Using generative artificial intelligence to identify central line-associated bloodstream infections. *clin infect dis* 2025;ciaf636. doi:10.1093/cid/ciaf636.
10. Windisch P, Dennstädt F, Koechli C, *et al.* The impact of temperature on extracting information from clinical trial publications using large language models. *Cureus* 2024;16:e75748. doi:10.7759/cureus.75748.
11. Flores-Balado Á, Castresana Méndez C, Herrero González A, *et al.* Using artificial intelligence to reduce orthopedic surgical site infection surveillance

workload: algorithm design, validation, and implementation in 4 Spanish hospitals. *Am J Infect Control* 2023;51:1225–1229. doi:10.1016/j.ajic.2023.04.165.

12. Wiemken TL, Carrico RM. Assisting the infection preventionist: use of artificial intelligence for health care-associated infection surveillance. *Am J Infect Control* 2024;52:625–629. doi:10.1016/j.ajic.2024.02.007.
13. Rodriguez-Nava G, Egoryan G, Goodman KE, Morgan DJ, Salinas JL. Performance of a large language model for identifying central line-associated bloodstream infections (CLABSI) using real clinical notes. *Infect Control Hosp Epidemiol.* 2024;46:1–4. doi:10.1017/ice.2024.164.
14. Mitchell BG, Hall L, Halton K, MacBeth D, Gardner A. Time spent by infection control professionals undertaking healthcare associated infection surveillance: a multi-centred cross sectional study. *Infection, Disease & Health* 2016;21:36–40. doi:10.1016/j.idh.2016.03.003.
15. Schrank GM, Branch-Elliman W, Leekha S, *et al.* Perceptions of health care–associated infection metrics by infection control experts. *JAMA Network Open* 2023;6:e238952. doi:10.1001/jamanetworkopen.2023.8952.
16. Pindyck T, Gupta K, Strymish J, *et al.* Validation of an electronic tool for flagging surgical site infections based on clinical practice patterns for triaging surveillance: operational successes and barriers. *Am J Infect Control* 2018;46:186–190. doi:10.1016/j.ajic.2017.08.026.

https://doi.org/10.1017/ice.2026.10429 Published online by Cambridge University Press