

Using Generative Artificial Intelligence for Healthcare-Associated Infection Surveillance

Daniel J. Morgan,^{1,2} Katherine E. Goodman,^{1,3} Westyn Branch-Elliman,^{4,5} Erica S. Shenoy,^{6,7,8} Jorge L. Salinas,⁹ Makoto Jones,^{10,11} Sunny P. Singh,^{12,13} Gregory M. Schrank,^{1,14} Lisa Pineles,^{1,3} Shatha AlShanqeeti,¹⁵ Anthony D. Harris,^{1,2,3} and Eili Klein¹⁶

¹Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, Maryland, USA; ²Medical Care Center, VA Maryland Health Care System, Baltimore, Maryland, USA; ³Institute for Healthcare Computing, University of Maryland, North Bethesda, Maryland, USA; ⁴Center for the Study of Healthcare Innovation, Implementation, and Policy, Greater Los Angeles VA Healthcare System, Los Angeles, California, USA; ⁵Department of Medicine, UCLA David Geffen School of Medicine, Los Angeles, California, USA; ⁶Infection Control, Mass General Brigham, Boston, Massachusetts, USA; ⁷Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts, USA; ⁸Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; ⁹Department of Medicine, Stanford University, Palo Alto, California, USA; ¹⁰Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA; ¹¹Medical Care, Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah, USA; ¹²Applied Innovations and Medical Informatics (AIMI), Veterans Affairs Long Beach Healthcare System, Long Beach, California, USA; ¹³Office of Research and Development, Veterans Health Administration, Washington, DC, USA; ¹⁴Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA; ¹⁵Institute of Human Virology, University of Maryland, Baltimore, Maryland, USA; and ¹⁶Department of Emergency Medicine, Johns Hopkins Medicine, Baltimore, Maryland, USA

(See the Editorial Commentary by Alzyood on pages 480–3.)

Generative artificial intelligence (GenAI) is becoming widespread in society but has had limited impact on medicine. GenAI can rapidly digest large documents, making it particularly well suited for healthcare-associated infection (HAI) surveillance that is largely based on clinical notes. In this article, we review evidence on the application of GenAI for HAI surveillance and highlight the necessary steps for changes to practice. Small studies of simulated patients and retrospective studies of patient records have found GenAI can isolate individual criteria for HAI and make HAI determinations with accuracy similar to that of human experts. GenAI tools are being developed to assist in HAI detection. Use of GenAI for HAI review could reduce variability between facilities, improve efficiency of HAI review, and lead to more time for direct prevention work. For widespread adoption, GenAI HAI detection must be safe and effective in prospective clinical studies.

Keywords. artificial intelligence; infection; surveillance.

Generative artificial intelligence (GenAI) has rapidly gained prominence for its ability to create content and provide cogent analysis. It can summarize large amounts of text and other material in mere seconds. While GenAI has made inroads across many industries, such as computer coding and graphic design, its penetration into healthcare has been comparatively limited. Its current uses are predominantly administrative and aimed at reducing clinician burden—including serving as a scribe, responding to patient portal responses, and drafting prior authorization letters [1]. Most explorations of GenAI have analyzed clean and curated datasets where there are right and wrong answers to support activities, such as completing medical board examinations. More recently, there have been efforts to use GenAI to analyze complex clinical cases or provide rationale for diagnostic reasoning. Yet the implementation of GenAI

into real-time clinical decision making has been limited by computing infrastructure, real-time data accessibility, and regulatory requirements [2, 3].

The few explorations of GenAI in infectious disease have focused mostly on theoretical use in clinical vignettes or retrospective data. For example, large language models (LLMs) using earlier, nonreasoning models performed below the level of human clinicians in infectious disease consultations or in the management of patients with bloodstream infections [4, 5]. However, LLMs have unique strengths in language processing that could be a good fit for healthcare-associated infection (HAI) surveillance [6–9].

In this viewpoint, we review the research on GenAI for HAI surveillance, ranging from theoretical approaches to practical considerations for implementation. We analyze how GenAI may overcome current surveillance limitations and propose a pathway for responsible integration into clinical use. Key terms are presented in the Glossary.

Glossary

Application programming interface (API): A set of rules that allows programs to interface with each other. Often used to perform large, batch-processing tasks with large language models.

Received 18 August 2025; accepted 21 October 2025; published online 12 November 2025

Correspondence: D. J. Morgan, Department of Epidemiology and Public Health, University of Maryland School of Medicine, 10 S Pine St, MSTF 334, Baltimore, MD 21201 (dmorgan@som.umaryland.edu); E. Klein, Department of Emergency Medicine, Johns Hopkins Medicine, 5801 Smith Ave., Suite 3220, Baltimore, MD 21209 (eklein@jhu.edu).

Clinical Infectious Diseases® 2026;82(3):473–9

© The Author(s) 2025. Published by Oxford University Press on behalf of Infectious Diseases Society of America. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.
<https://doi.org/10.1093/cid/ciaf619>

Continued

Chatbot: A program for humans to interact with a large language model (LLM) through a conversational interface.

Generative artificial intelligence (GenAI): Artificial intelligence (AI) that uses generative models to produce text and other data.

GPT-4o: A commonly used large language model, often accessed through a chatbot.

Healthcare-associated infection (HAI): Complications of medical care, especially those reported as metrics of hospital performance, such as central line-associated bloodstream infection, catheter-associated urinary tract infection, or surgical site infection.

Large language model (LLM): A GenAI model that can accept free-text inputs and generate humanlike text and control processes.

LLaMA: A commonly used open-source LLM.

Machine learning: A general category of AI in which computers learn patterns or predictions from data without explicit programming; it is traditionally predictive, using structured data and outputting quantitative estimates (eg, probabilities).

CURRENT STATE OF HAI SURVEILLANCE AND REPORTING

HAIs, including central line-associated bloodstream infection (CLABSI), catheter-associated urinary tract infection (CAUTI), and surgical site infection (SSI), are complications of medical care, a significant portion of which are preventable [10]. Hospital metrics to track rates of HAIs were among the first to be publicly reported and linked to reimbursement [11]. The US Centers for Disease Control and Prevention's (CDC's) National Healthcare Safety Network (NHSN) provides definitions, a way to report surveillance and is used by >37 000 US facilities [12]. Similar definitions and reporting structures are used in Europe and Asia [13, 14].

While the visibility and importance of these metrics to a medical center's bottom line and reputation has driven investment in HAI tracking, HAI surveillance presents challenges. This surveillance is typically conducted by healthcare system-employed infection preventionists (IPs), who perform extensive review of the electronic health record (EHR) for HAI cases based on complex definitions. This process is resource intensive: the person-time required to perform surveillance activities ranges from 15–30 minutes for each CAUTI and up to 75 minutes for each CLABSI [6, 7, 15], which can occupy a third of all IP time [16]. HAI surveillance limits time for more active infection prevention activities, such as engagement with hospital wards to reduce HAIs. In 1 survey, 95% of infection control

experts reported that HAI surveillance distracted from other infection control activities [17]. Furthermore, despite increasingly detailed definitions from the CDC, reporting is subject to potential bias and human error, with pressure from payment structures and reputational impact to report lower HAI rates [15].

Lin et al [15] identified up to 2-fold interhospital variability in CLABSI rates when comparing an objective algorithmic approach with actual human based reporting. More recent studies have continued to document underreporting and variable and inconsistent interpretation of CLABSI definitions [18–21]. Issues with objectivity in reporting HAIs have led to an erosion of trust in these metrics by clinicians and administrators [22]. This has prompted the development of new metrics like hospital-onset bacteremia that are entirely based on EHR data and require no additional manual review. However, hospital-onset bacteremia is not currently a required reporting metric [23].

Traditional AI to Detect HAIs

Automated and partially automated surveillance can improve efficiency, but historically these techniques have been difficult to implement [24]. Before the emergence of GenAI, automated HAI surveillance centered on machine learning or other statistical algorithms that used structured data elements (eg, *International Classification of Diseases, Tenth Revision* codes) that could be readily extracted from the EHR [25]. A fundamental barrier to these automated HAI surveillance approaches is that important variables are only found often in unstructured clinician notes, and reliably extracting these criteria can be challenging [26, 27].

In addition, automated surveillance tools developed at one facility have not been readily transferable to others. One study, for example, identified variables in the EHR that could classify colorectal SSI but required local availability of model parameters and validation before broad implementation [28].

Natural Language Processing for HAIs

Before GenAI, the approach to accessing unstructured data from notes was natural language processing. While natural language processing offers some advantages, such as cheaper computing costs compared with LLMs, it also has important methodological and implementation barriers [29]—namely, it requires experts to build and test models, depends on institution-specific training datasets created through resource-intensive expert annotation, and demonstrates limited transferability between healthcare systems with different documentation practices [30, 31]. GenAI, which is trained on millions of available documents, has foundational knowledge of clinical information and is more generalizable and easier to implement with less expertise.

EXPLORATION OF GenAI TO DETECT HAIs

Preclinical studies performed on simulated cases have demonstrated the potential of chatbots and LLM-based tools to identify common HAIs [6]. Most exploration of GenAI for actual patients has been performed by pasting notes and data into chatbot windows. However, future tools will likely include more carefully developed passive surveillance that extracts data automatically from the EHR, requiring less human effort (Table 1).

Preclinical Proof-of-Concept Studies

Because of privacy concerns around commercial chatbots, initial studies were limited to synthetic or simulated patient records. However, these preliminary investigations have yielded encouraging results. Investigators explored the performance of GPT-4o to detect CAUTI using simulated cases, finding acceptable sensitivity and specificity. [7] Similarly, GPT-4o could identify CLABSI and CAUTI in CDC training cases. [8] These simulations, however, were relatively simple and did not reflect the reality of complex medical records.

Retrospective Evaluation of GenAI on Patient Records

As healthcare institutions have begun to implement secure chatbot environments, evaluation has expanded to real patient records. CLABSI has been the most studied HAI. Investigators found GPT-4o could identify 90% of CLABSI cases and exclude 75% of non-CLABSI bacteremia cases in 20 cases and 20 controls at a single hospital [6]. These reviews were efficient, requiring a median of 5 minutes versus the reported IP review time of 75 minutes [6].

A larger operational assessment within the US Department of Veterans Affairs (VA) examined CLABSI determination across 11 hospitals with standard prompting, data, and evaluation used by frontline IPs and physicians, and it found GPT-4o based AI-assisted review was as accurate as traditional review, faster, and less variable between sites [32]. Including detailed guidance from CDC-NHSN criteria in the prompt improved accuracy. Furthermore, with careful prompting, GenAI could highlight key data in unstructured notes, such as presence of a central line or secondary bloodstream infection, enabling more efficient manual review. However, GenAI still misclassified some CLABSIs or other HAIs after identifying data elements correctly (eg, identifying dates of central line placement and bacteremia but incorrectly concluding that a CLABSI was present when bacteremia occurred within a day of line placement).

In a single-center study, CAUTI surveillance with GenAI was performed on 21 CAUTI cases and 21 matched controls, reporting 95% sensitivity and 76% specificity (which improved to 90% with expert review of GenAI output) [33]. GenAI also identified 1 true CAUTI in the control group that was missed by facility review.

GenAI detection of SSI was explored using public health record abstractions and CDC-NHSN definitions. GPT-4o was 100% sensitive and 54% specific and could categorize the SSIs as superficial, deep, or organ space. These reviews required a median of 6 minutes per case, though it is important to emphasize that data was not from the medical record [9].

Implementing GenAI for HAI Surveillance

While these proof-of-concept studies show promise, transitioning to an operational system must follow similar foundational steps, whether this system is implemented as a secure chatbot to perform reviews by copying and pasting notes or as a tool with an application programming interface to automate or partially automate HAI identification. These steps include (1) the technical aspects of using GenAI; (2) decisions around the degree of human support versus automation; (3) hospital incorporation of GenAI for HAI detection; and (4) building GenAI into human workflow.

Technical Steps for GenAI Clinical Implementation

Technical steps for clinical implantation of GenAI include (1) capturing appropriate data; (2) using standardized prompts; and (3) exporting results in a useful format. First, collection of appropriate categories of data, such as clinical documentation and laboratory results, must be standardized and validated. Second, a generalizable model prompting strategy must be developed and validated to guide the GenAI system to capture the same elements and perform accurate and repeatable reviews with low variability. Ideally, a repository of validated prompts should be developed and maintained publicly by a regulatory body, such as the CDC or national organizations, such as the Society for Healthcare Epidemiology of America or the Association for Professionals in Infection Control. Third, GenAI results must be exported in a structured manner for incorporation into either current HAI reporting processes or new surveillance frameworks.

GenAI Support for Human Review Versus Automation

Even with technically sound GenAI systems that match or exceed human reviewer accuracy, healthcare organizations and regulators face important implementation decisions. Surveillance implementation options exist along a continuum (Figure 1): At one extreme is full automation, in which machines using GenAI and other methods would review cases independently and provide definitive determinations (eg, “CLABSI yes or no”), while at the other extreme GenAI is used *only* to extract a few key elements for CLABSI determination.

Hospital Incorporation of GenAI for HAI Detection

Current systems, such as Epic or Theradoc, currently provide identification of bacteremia and important hospital dates from structured data fields and are well accepted. For GenAI tools, most hospitals won't be designing their own specific systems but rather choosing from what is available from large EHR

Table 1. Role of Large Language Models in the Identification and Prevention of Different Types of Healthcare-Associated Infections

HAI	Data Needed for GenAI Review	How GenAI Could Be Used	Strength of Evidence for Use of GenAI	Study Limitations
CLABSI	Clinical notes, devices, microbiology, vital signs, ADT	To identify/rule out CLABSI	Multicenter cohort of frontline providers [32]	Retrospective, not implemented
CAUTI	Clinical notes, devices, microbiology, vital signs, ADT	To identify/rule out CAUTI	Preclinical findings [7, 8] and single cohort	Retrospective, 1 hospital, not implemented
SSI	Clinical notes, operational notes, microbiology, vital signs, ADT, radiology	To identify/rule out SSI	Small study of abstracted data [9]	Retrospective; did not use patient records
HOB	Clinical notes, devices, operational notes, microbiology, vital signs, ADT, radiology	To attribute cause of HOB and potential interventions ^a	None	Has not been studied

Abbreviation: ADT, admission, discharge, transfer; CAUTI, catheter-associated urinary tract infection; CLABSI, central line-associated bloodstream infection; GenAI, generative artificial intelligence; HAI, healthcare-associated infection; HOB, hospital-onset bacteremia; SSI, surgical site infection.

^aMetric defined only by blood culture ≥ 72 hours after admission.

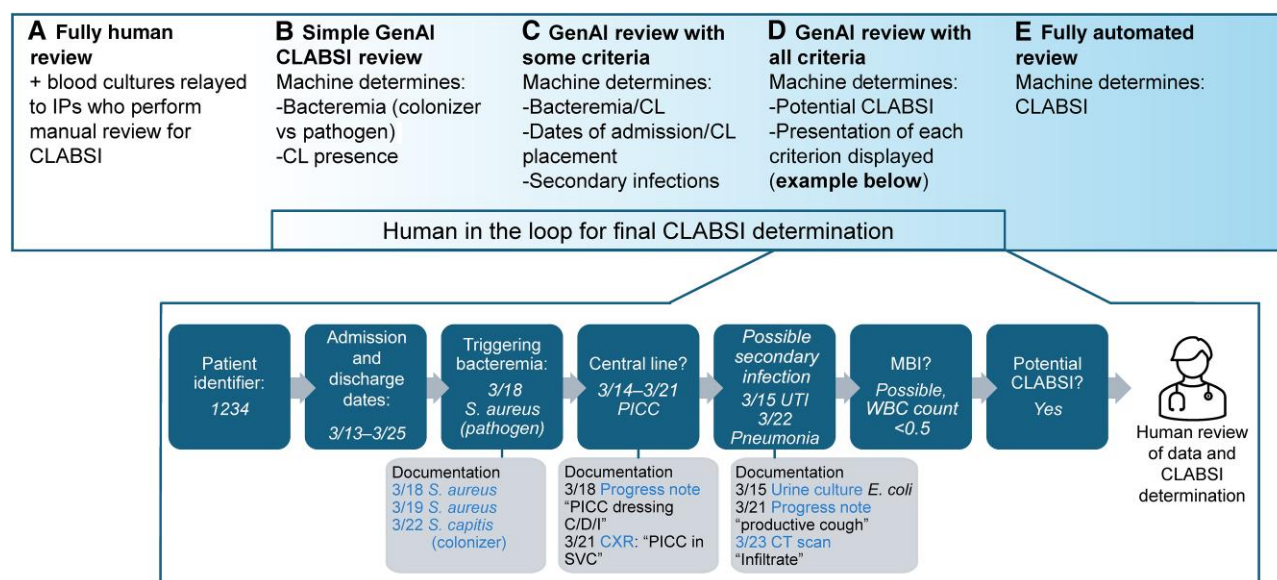


Figure 1. Range of implementation approaches for healthcare-associated infection central line (CL)–associated bloodstream infection (CLABSI) detection, from traditional human review through a range of more automated approaches and ending with a fully automated review without human oversight. An example is provided, demonstrating a potential generative artificial intelligence (GenAI) review for CLABSI, documenting specific criteria for CLABSI that could then be reviewed by infection prevention experts for final determination. Documentation in blue represents hyperlinks to elements in the electronic health record, and dates are given in month/date format. Abbreviations: C/D/I, clean, dry, and intact; CT, computed tomographic; *E. coli*, *Escherichia coli*; HOB, hospital-onset bacteremia; MBI, mucosal barrier infection; PICC, peripherally inserted central catheter; *S. aureus*, *Staphylococcus aureus*; *S. capitis*, *Staphylococcus capitis*; UTI, urinary tract infection; WBC, white blood cell (cells/microliter).

vendors. However, it is critical for developers, regulators, and hospitals to consider the range of different GenAI options and the degree of automation versus human oversight as products are created (Figure 1).

Building GenAI Into Human Workflow

Beyond decisions around technology and the EHR, the human hospital environment must be considered. Who currently performs HAI surveillance using what tools? How would a GenAI tool affect their workflow? This has not been addressed by most studies, with only the VA AI CLABSI study addressing how GenAI surveillance would be used by clinicians and the acceptability and usefulness of the tool [6–9, 32, 33].

Governance, Accountability, and Reliability

Organizations will require plans for ongoing evaluation of GenAI performance and sustainment with assistance from national organizations. Eventually this must be required for hospitals to perform or through a third-party with access to the EHR. Potentially, spot audits or review of reported cases could occur (or examination of the ratio of GenAI presumed cases vs those reported for signals to audit). Finally, protocols for ongoing evaluation of the results against reference standards should be established.

Potential Benefits of GenAI for HAI Detection

GenAI detection of HAIs could lead to more efficient review of records and less IP expert time committed to HAI surveillance

[6]. In addition, GenAI review may be less variable and more objective as it applies a single standard instead of the subjectivity of different reviewers at each hospital (although this must be determined in studies, given that LLMs have some stochasticity in their performance) [32]. Less variability in reviews could result in a fairer metric that could achieve the hoped-for benefits of public reporting: quality improvement to truly decrease infections. These potential benefits need to be assessed in real-world studies.

GenAI may also be helpful to identify potential causes of HAIs and interventions to reduce HAIs. For example, analysis of notes on patients with CAUTI could identify incorrect collection of samples from catheter bags from nursing notes, leading to education of nurses to improve sample collection.

Ultimately, definitions of HAIs from the CDC and other groups that define or report HAIs may change if reviews can be performed using GenAI in a standardized fashion across hospitals. Currently, HAI identification is complex, requiring trained reviewers who are generally applying exclusionary rules with multiple layers. For example, for CLABSI determination, there is an exclusion for secondary bloodstream infection without a culture at a secondary site (known by terms like *PNU2* or *IAB3*). These exclusions are hard to apply and even harder to audit. However, a change to definitions that could be applied in a standard way across hospitals by GenAI and displayed clearly for each case could lead to more transparency and consistency with clinical definitions of HAIs. For example, a simplified CLABSI definition could require the presence of a central line, bacteremia with a pathogen, and a GenAI's best explanation of a secondary infection.

CHALLENGES FOR GenAI HAI DETECTION

Accuracy

The first challenge will be how well GenAI identifies HAIs in a prospective fashion—how accurate is it compared with human experts across each individual HAI? Which method of review best identifies preventable HAIs that can be addressed by quality improvement? How can GenAI be implemented into clinical systems to maximize accuracy? While some studies have been published [6], larger implementation studies are needed.

AI Ethics and Bias

Successful GenAI HAI detection tools would introduce new ethical considerations. Access to GenAI-assisted HAI surveillance tools, as seen with EHR, billing, and other technologies, could further separate hospitals into the “haves” and “have nots” if these tools affected reimbursement and IP time for other high-impact investigations and activities. In addition, conscious and unconscious biases—related to racial or gender stereotypes, for example—can be embedded in the training data for LLMs, which has been demonstrated to influence their clinical

recommendations [34–36]. These potential biases should be evaluated during tool development and evaluation [34–36].

Finally, AI scribes are increasingly generating clinical notes and other documentation tools are coming, such as AI-generated discharge summaries [37]. As the unstructured inputs for GenAI HAI tools are increasingly written by LLMs, not humans, AI scribes could be trained to include content and phrasings to artificially reduce HAI rates. For example, scribes could preferentially avoid listing symptoms that are part of the CAUTI criteria.

Variation by Company or Model

Although prompts can be standardized, small changes and different base LLMs could modify outcomes. Even version updates to LLMs could change performance. This should be assessed routinely, and companies should be required to transparently display prompts and model updates and functions.

Costs

The cost of GenAI is likely to be a barrier. The models require resources each time they are run, and new tools are often costly for health systems. If spending shifts from infection control staff to specific technological tools, the net effect could be less flexibility and ability to respond to needs within an infection control program.

Policy Incentives

Goodhart's law—when a measure becomes a target, it ceases to be a good measure—may apply to GenAI HAI reviews. No process is perfect and there would be ways to “game” a new system based on GenAI review, such as preferentially documenting secondary infections.

GenAI FOR INFECTION CONTROL AND PATIENT SAFETY

Expansion of Infection Surveillance

Many HAI surveillance programs are limited to a few types of HAIs, with the choice driven by regulatory requirements (eg, CLABSI, CAUTI, and select HAIs). However, many other infections may be important but impractical to monitor given that they are rare—for example, cardiac device-associated infections, which are rare but catastrophic. With automated review of standardized definitions, the number of procedures for which HAI surveillance could be efficiently conducted could expand dramatically. Other potential targets include other surgical infections, nonventilator hospital-associated pneumonia, complications of endoscopy, and complications of antibiotic prophylaxis after various procedures [26, 38, 39].

Outbreak Detection and Investigation

Identifying clusters or outbreaks of infections with similar characteristics or exposures often depends on humans

reviewing records. GenAI may be used to monitor for specific exposures or atypical infections if patient notes are available. For example, if identical organisms are identified, common exposures may be rapidly explored with GenAI analysis of notes. Similarly, GenAI is capable of identifying outpatient risk factors for avian influenza [40].

Other Patient Safety Metrics Detectable by GenAI

GenAI has been explored for other areas of quality improvement,—for example, the sepsis metric or reasons for antibiotic use [41]. GenAI could also improve screening for risk factors that influence the chance of infections, such as long-term care exposure [42]. Beyond HAIs, other traditional hospital patient safety metrics, such as falls and pressure injuries, might be identified by GenAI review as they are primarily documented in notes and are reported not immediately but with a delay, which would allow all data to accrue.

CONCLUSIONS

GenAI is promising for HAI detection and other patient safety metrics. A handful of studies have found that current technology can assist infection prevention teams in case review. Plans for developing these are moving forward in the VA and EHR companies. If effectively implemented, these tools could improve efficiency of metric collection, potentially allowing expansion of surveillance to other infections, and provide more time for infection prevention and control teams to devote to high-yield activities. Priorities for the field are (1) large prospective studies of GenAI tools; (2) evaluation of human use of tools and cognitive biases that occur; and (3) reconsideration of optimal HAI definitions for reviews performed by GenAI.

Notes

Financial support. None.

Potential conflicts of interest. The authors: No reported conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

- Morgan DJ, Rodman A, Goodman KE. How physicians can prepare for generative AI. *JAMA Intern Med*. doi:10.1001/jamainternmed.2025.4914.
- Rodman A, Buckley TA, Manrai AK, Morgan DJ. Artificial intelligence vs clinician performance in estimating probabilities of diagnoses before and after testing. *JAMA Netw Open* 2023; 6:e2347075.
- Goodman KE, Morgan DJ, Hoffmann DE. Clinical algorithms, antidiscrimination laws, and medical device regulation. *JAMA* 2023; 329:285–6.
- Kanjilal S. Flying into the future with large language models. *Clin Infect Dis* 2024; 78:867–9.
- Schwartz IS, Link KE, Daneshjou R, Cortés-Penfield N. Black box warning: large language models and the future of infectious diseases consultation. *Clin Infect Dis* 2024; 78:860–6.
- 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; 46:305–8.
- Perret J, Schmid A. Application of OpenAI GPT-4 for the retrospective detection of catheter-associated urinary tract infections in a fictitious and curated patient data set. *Infect Control Hosp Epidemiol* 2024; 45:96–9.
- Wiemken TL, Carrico RM. Assisting the infection preventionist: use of artificial intelligence for health care-associated infection surveillance. *Am J Infect Control* 2024; 52:625–9.
- 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.
- Schreiber PW, Sax H, Wolfensberger A, Clack L, Kuster SP, Swissnoso. The preventable proportion of healthcare-associated infections 2005–2016: systematic review and meta-analysis. *Infect Control Hosp Epidemiol* 2018; 39:1277–95.
- Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006; 355:2725–32.
- Centers for Disease Control and Prevention. National Healthcare Safety Network (NHSN). 2025. Available at: <https://www.cdc.gov/nhsn/index.html>. Accessed 12 August 2025.
- Hansen S, Sohr D, Geffers C, et al. Concordance between European and US case definitions of healthcare-associated infections. *Antimicrob Resist Infect Control* 2012; 1:28.
- Asia Pacific Society of Infection Control. 2022. APSIC guide for prevention of catheter associated urinary tract infections (CAUTIs). Available at: <https://apsic-apac.org/wp-content/uploads/2022/06/APSIC-CAUTI-guidelines-2022.pdf>. Accessed 6 October 2025.
- Lin MY, Hota B, Khan YM, et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. *JAMA* 2010; 304:2035–41.
- 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. *Infect Dis Health* 2016; 21:36–40.
- Schrank GM, Branch-Elliman W, Leekha S, et al. Perceptions of health care-associated infection metrics by infection control experts. *JAMA Netw Open* 2023; 6:e238952.
- Mayer J, Greene T, Howell J, et al. Agreement in classifying bloodstream infections among multiple reviewers conducting surveillance. *Clin Infect Dis* 2012; 55:364–70.
- Larsen EN, Gavin N, Marsh N, Rickard CM, Runnegar N, Webster J. A systematic review of central-line-associated bloodstream infection (CLABSI) diagnostic reliability and error. *Infect Control Hosp Epidemiol* 2019; 40:1100–6.
- Kim JJ, Kang H, Khan AN, et al. A comparison of National Healthcare Safety Network surveillance definitions to physician review for central-line-associated bloodstream infections. *Infect Control Hosp Epidemiol* 2023; 44:2062–4.
- Worth LJ, Brett J, Bull AL, McBryde ES, Russo PL, Richards MJ. Impact of revising the National Nosocomial Infection Surveillance System definition for catheter-related bloodstream infection in ICU: reproducibility of the National Healthcare Safety Network case definition in an Australian cohort of infection control professionals. *Am J Infect Control* 2009; 37:643–8.
- MacEwan SR, Gaughan AA, Beal EW, Hebert C, DeLancey JO, McAlearney AS. Concerns and frustrations about the public reporting of device-related healthcare-associated infections: perspectives of hospital leaders and staff. *Am J Infect Control* 2023; 51:633–7.
- Howard-Anderson J, Morgan DJ. Moving beyond central line-associated bloodstream infections. *Ann Intern Med* 2024; 177:822–3.
- van Mourik MSM, Perencevich EN, Gastmeier P, Bonten MJM. Designing surveillance of healthcare-associated infections in the era of automation and reporting mandates. *Clin Infect Dis* 2018; 66:970–6.
- Streefkerk HRA, Verkooijen RP, Bramer WM, Verbrugh HA. Electronically assisted surveillance systems of healthcare-associated infections: a systematic review. *Euro Surveill* 2020; 25:1900321.
- Branch-Elliman W, Sundermann AJ, Wiens J, Shenoy ES. Leveraging electronic data to expand infection detection beyond traditional settings and definitions (part II/III). *Antimicrob Steward Healthc Epidemiol* 2023; 3:e27.
- Shenoy ES, Branch-Elliman W. Automating surveillance for healthcare-associated infections: rationale and current realities (part I/III). *Antimicrob Steward Healthc Epidemiol* 2023; 3:e25.
- Mulder T, Kluytmans-van den Bergh MFQ, van Mourik MSM, et al. A diagnostic algorithm for the surveillance of deep surgical site infections after colorectal surgery. *Infect Control Hosp Epidemiol* 2019; 40:574–8.
- Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural language processing for real-time catheter-associated urinary tract infection surveillance: results of a pilot implementation trial. *Infect Control Hosp Epidemiol* 2015; 36:1004–10.

30. Woeltje KF, Lin MY, Klompas M, Wright MO, Zuccotti G, Trick WE. Data requirements for electronic surveillance of healthcare-associated infections. *Infect Control Hosp Epidemiol* **2014**; 35:1083–91.
31. Tvardik N, Kergourlay I, Bittar A, Segond F, Darmoni S, Metzger MH. Accuracy of using natural language processing methods for identifying healthcare-associated infections. *Int J Med Inform* **2018**; 117:96–102.
32. Morgan D. Using an LLM for frontline providers to identify healthcare-associated infections across 11 hospitals. Presented at: Stanford Artificial Intelligence and Medical Imaging (AIMI) Symposium; 3 **June 2025**; Palo Alto, CA.
33. Alshanqeeti S, Coffey KC, McDermott K, et al. Comparing generative artificial intelligence vs. experts for detection of catheter-associated urinary tract infection. *Clin Infect Dis* **2026**; 82:148–50.
34. Bouguettaya A, Stuart EM, Aboujaoude E. Racial bias in AI-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models. *NPJ Digit Med* **2025**; 8:332.
35. Ayoub NF, Balakrishnan K, Ayoub MS, Barrett TF, David AP, Gray ST. Inherent bias in large language models: a random sampling analysis. *Mayo Clin Proc Digit Health* **2024**; 2:186–91.
36. Omar M, Soffer S, Agbareia R, et al. Sociodemographic biases in medical decision making by large language models. *Nat Med* **2025**; 31:1873–81.
37. Morgan DJ, Branch-Elliman W, Goodman KE. Time to study implementation of AI-generated discharge summaries. *JAMA Intern Med* **2025**; 185:825–6.
38. Branch-Elliman W, Sundermann AJ, Wiens J, Shenoy ES. The future of automated infection detection: innovation to transform practice (part III/III). *Antimicrob Steward Healthc Epidemiol* **2023**; 3:e26.
39. Branch-Elliman W, Lamkin R, Shin M, et al. Promoting de-implementation of inappropriate antimicrobial use in cardiac device procedures by expanding audit and feedback: protocol for hybrid III type effectiveness/implementation quasi-experimental study. *Implement Sci* **2022**; 17:12.
40. Goodman KE, Shams SM, Magder LS, Baghdadi JD, Morgan DJ, Harris AD. Generative AI-based surveillance for avian influenza across a statewide healthcare system. *Clin Infect Dis* **2025**; 81:900–3.
41. Yuan K, Yoon CH, Gu Q, et al. Leveraging transformers and large language models with antimicrobial prescribing data to predict sources of infection for electronic health record studies. *medRxiv* 24305966 [Preprint]. April 19, 2024. Available from: <https://doi.org/10.1101/2024.04.17.24305966>.
42. Goodman KE, Robinson ML, Shams SM, et al. Identification of long-term care facility residence from admission notes using large language models. *JAMA Netw Open* **2025**; 8:e2512032.