HEALTHCARE INFECTIONS

Epidemiological and Mathematical Modelling of Healthcare-Associated Infections: From Transmission Insight to Implementable Prediction

Arthi

Faculty of Nursing, Sri Ramachandra Institute of Higher Education and Research (SRIHER), Chennai, Tamil Nadu, India

Published November 2026

Abstract

Background: Healthcare-associated infections (HAIs) remain an important cause of preventable patient harm, prolonged hospitalisation, increased healthcare expenditure and antimicrobial resistance. Surveillance describes the frequency and distribution of infections, while epidemiological and mathematical models can help explore transmission pathways, estimate risk and assess prevention strategies. However, the translation of model outputs into routine infection-prevention practice remains challenging.

Objective: This narrative review examines epidemiological surveillance, mathematical transmission models, statistical prediction and machine-learning approaches used in the study of HAIs. It focuses on the strengths and limitations of these approaches, the quality and transportability of the evidence, and the requirements for translating model outputs into infection-prevention decisions.

Methods: A targeted, non-systematic literature search was undertaken for this narrative review. PubMed-indexed literature and publicly accessible publications from the World Health Organization, the Indian Council of Medical Research-related literature, the Centers for Disease Control and Prevention and peer-reviewed journals were consulted. Search concepts included healthcare-associated infection, nosocomial transmission, mathematical modelling, agent-based modelling, prediction models, machine learning, surveillance and antimicrobial resistance. Priority was given to systematic reviews, methodological guidance and studies illustrating transmission modelling or implementation. No formal dual screening, quantitative synthesis or comprehensive risk-of-bias assessment was undertaken.

Discussion: Compartmental and agent-based models provide different ways to represent transmission between patients, healthcare workers, devices and healthcare environments. Statistical and machine-learning approaches can support risk estimation and surveillance, but model performance in development datasets does not guarantee transportability. Recent systematic reviews identify substantial activity in artificial-intelligence-based HAI detection and prediction, while the wider literature highlights the importance of external evaluation, calibration and transparent reporting. Models that exclude community and long-term-care settings may also misrepresent the sources and pathways of hospital-associated pathogens.

Conclusion: The value of HAI modelling depends not only on methodological sophistication but also on the quality of surveillance data, transparent assumptions, independent evaluation and evidence that model-informed decisions improve infection-prevention practice. Future work should prioritise context-sensitive models, representative validation and prospective evaluation of clinical and public-health impact.

Keywords: healthcare-associated infections; mathematical modelling; transmission dynamics; epidemiological surveillance; prediction models; machine learning; antimicrobial resistance; infection prevention and control

1 Introduction

Healthcare-associated infections are infections acquired during the process of receiving healthcare that were not present or incubating at the time of admission. They occur across acute-care hospitals, long-term-care facilities and other healthcare settings. Their consequences extend beyond the affected patient: they increase the demand for diagnostic and therapeutic resources, contribute to antimicrobial use and resistance, and may amplify pathogen transmission within and beyond health-care facilities. The World Health Organization's 2024 global report on infection prevention and control identifies HAIs and antimicrobial resistance as continuing threats to patient safety and health-system performance, while emphasising the importance of effective infection-prevention programmes and water, sanitation and hygiene services [1].

Routine surveillance remains essential, but surveillance counts alone rarely explain the mechanisms responsible for infection or predict how transmission may change after an intervention. Epidemiological and mathematical modelling can complement surveillance by representing the movement of pathogens through populations, estimating the potential effect of interventions and identifying patterns that may not be apparent from aggregate rates. Statistical prediction models address a related but distinct question: which patients, wards or facilities are more likely to experience an infection or an outbreak?

These methods are increasingly relevant in the context of antimicrobial resistance (AMR). In India, Gandra and colleagues have highlighted the substantial burden of resistant organisms associated with HAIs and the need to strengthen hospital epidemiology and infection-prevention research capacity [2]. Their discussion of the ICMR HAI surveillance network, which involves 26 tertiary-care hospitals, illustrates both the value of coordinated surveillance and the need for stronger context-specific research. The reported carbapenem resistance among organisms causing central line-associated bloodstream infections ranged from 53% for Pseudomonas species to 77% for Acinetobacter species [2]. Such findings underline the importance of understanding not only how frequently infections occur, but also how resistant organisms are introduced, transmitted and sustained in healthcare settings.

This review does not claim to identify a previously unrecognised methodological problem. Instead, it brings together evidence across transmission modelling, statistical prediction and machine learning around a practical question: what is required for a model to become useful to an infection-prevention programme rather than remaining a research output? The Indian surveillance context is used as an illustrative setting in which data availability, healthcare-system diversity and local AMR patterns make model transportability especially important.

2 Search approach and scope

This article is a narrative review and is not presented as a systematic review or meta-analysis. A targeted, non-systematic search was undertaken during preparation of the manuscript. PubMed-indexed literature and publicly accessible publications from WHO, ICMR-related sources, CDC and peer-reviewed journal websites were consulted. Search concepts combined terms relating to health-care-associated infection, hospital-acquired infection, nosocomial transmission, mathematical modelling, compartmental models, agent-based models, epidemiological surveillance, prediction models, artificial intelligence, machine learning and antimicrobial resistance.

Priority was given to systematic reviews, methodological guidance, studies that described transmission models in healthcare settings, and publications that discussed validation or implementation. Reference lists of relevant review articles were also used to identify additional sources. The search was not designed to retrieve every study in the field. No independent duplicate screening, formal study-level risk-of-bias assessment or quantitative synthesis was performed. Accordingly, the findings are presented as a critical narrative synthesis and should not be interpreted as an exhaustive evidence map. The SANRA scale provides a useful framework for considering the quality of narrative reviews, including the importance of a clear rationale, aims, search description, referencing and scientific reasoning [3].

3 Surveillance as the evidentiary base—and its limitations

Surveillance provides the empirical foundation for HAI epidemiology. Common measures include incidence, prevalence, incidence density and device-associated infection rates expressed in relation to device-days. These measures support comparisons over time and can help identify wards or patient groups requiring further investigation. Surveillance systems also support antimicrobial-resistance monitoring and can provide baseline information for evaluating infection-prevention initiatives.

However, surveillance indicators require careful interpretation. A change in an observed infection rate may reflect a true change in transmission, but it may also be influenced by changes in patient case-mix, device utilisation, length of stay, microbiological testing, case definitions or completeness of reporting. A reduction in infection density, for example, may partly reflect a reduction in device exposure rather than a change in the probability of infection per exposure day. Similarly, improved case ascertainment may initially increase reported rates even when surveillance quality has improved.

These issues matter directly to modelling. A model is only as meaningful as the outcome and exposure data used to construct it. If case definitions change over time or differ between hospitals, apparent differences may reflect measurement rather than epidemiology. Aggregated surveillance data can reveal broad trends, but they may not contain the patient-level information needed to reconstruct contact networks, movement between wards, device exposure or colonisation status. Gandra and colleagues note that India lacks comprehensive country-wide HAI rates and emphasise the need to strengthen hospital epidemiology and surveillance capacity [2]. This limits the extent to which models developed in other healthcare systems can be transferred without local evaluation.

4 Transmission modelling: Compartmental and agent-based approaches

4.1 Compartmental models

Compartmental models divide a population into groups according to infection or colonisation status. The susceptible–infectious–recovered (SIR) framework is a foundational example, while susceptible–exposed–infectious–recovered (SEIR) models add a latent or exposed state. In healthcare settings, models may be extended to distinguish colonised and infected patients, patients with and without devices, healthcare workers, environmental reservoirs or different wards. Such structures can be used to explore how contact rates, admission prevalence, length of stay, screening, isolation and decolonisation may influence transmission.

The main advantage of compartmental models is their relative transparency. Their assumptions and transition parameters can often be expressed mathematically and explored through sensitivity analyses. They can also help compare intervention scenarios when direct experimental evaluation is difficult. Nevertheless, the simplicity that makes these models interpretable can also be a limitation. Hospitals are not homogeneous, well-mixed populations. Patients differ in susceptibility and exposure; healthcare workers move between patients and wards; devices create specific routes of exposure; and environmental contamination may persist after an infected patient has left.

Adding compartments can improve realism, but it also increases the number of parameters that must be estimated. When data are insufficient, parameters may be based on assumptions or borrowed from other settings. This can produce models that fit observed data while leaving uncertainty about the mechanisms represented. Consequently, model outputs should be interpreted in relation to the assumptions and data used, rather than as precise forecasts of future infection rates.

4.2 Agent-based models

Agent-based models represent individual entities—such as patients, healthcare workers, rooms, devices or facilities—and define how they interact over time. This approach can capture heterogeneity in contact patterns, patient movement, staffing arrangements and exposure opportunities. It may therefore be useful for exploring questions that are difficult to represent in a single population-average transmission rate.

Henriot and colleagues developed an agent-based susceptible–exposed–infected model to simulate the transmission dynamics of bloodborne pathogens within hospitals. Their model was illustrated using hepatitis C virus transmission in an Egyptian hospital and a further example involving hepatitis B virus transmission in Ethiopia [4]. The study demonstrates how individual-level interactions and invasive procedures can be represented in a hospital transmission model.

Agent-based models do not eliminate uncertainty. Their results depend on the accuracy of the contact structure, behavioural rules, parameter values and calibration data. A model developed for one hospital may not represent another hospital with different staffing, patient flow, device availability or infection-prevention practices. These models can also require substantial computational and data resources. Their strongest contribution is often to compare plausible scenarios under clearly stated assumptions, rather than to provide a universally applicable numerical forecast.

4.3 The hospital–community interface

Hospitals should not always be modelled as closed systems. Patients may enter hospital already colonised or infected, move between acute-care and long-term-care facilities, or acquire organisms in the community before admission. Healthcare workers, visitors and inter-facility transfers may also connect otherwise separate settings. A model that attributes every hospital-onset event to transmission occurring within the hospital may therefore misrepresent the contribution of importation and wider transmission networks.

A systematic scoping review by Lin and colleagues examined epidemiological models of the community contribution to hospital-associated pathogens. Of 29 eligible models, the review found that different studies represented long-term-care facilities and non-healthcare community settings to varying degrees, highlighting gaps in how the wider transmission system is incorporated [5]. This evidence supports the need for models that make the boundaries of the system explicit and consider community and long-term-care settings where relevant.

5 Statistical models: Risk estimation and intervention evaluation

Statistical models are commonly used to identify factors associated with HAI occurrence and to estimate the probability or rate of infection. Logistic regression may be used for binary outcomes, such as whether an infection occurred during a specified period. Poisson regression can be used for count outcomes when an appropriate exposure denominator is available, while negative binomial regression may be more suitable when count data are overdispersed. Multilevel or mixed-effects models can account for clustering of patients within wards or hospitals.

These methods can produce interpretable estimates and can be useful for examining associations between patient characteristics, device exposure, antimicrobial use and infection outcomes. However, association should not be confused with causation. A variable associated with infection in an observational model does not necessarily represent a modifiable cause. For example, length of stay may be associated with infection because patients who remain hospitalised longer accumulate more exposure time, because severe illness increases both length of stay and infection risk, or because infection itself prolongs admission.

Time-series approaches, including interrupted time-series analysis, can be used to examine changes in infection rates after the introduction of an intervention or policy. Their usefulness depends on the quality and length of the pre-intervention and post-intervention observations and on appropriate consideration of secular trends, seasonality, concurrent interventions and changes in case ascertainment. Where these factors are not addressed, apparent intervention effects may be misleading.

The appropriate model therefore depends on the question. Transmission models ask how infection may spread and how interventions may alter transmission. Statistical models estimate associations or risks in observed data. Evaluation designs ask whether an intervention changed outcomes. These approaches can complement one another, but they should not be treated as interchangeable.

6 Machine learning and prediction: Promise and validation requirements

Machine-learning approaches—including random forests, gradient boosting, support vector machines and neural networks—can accommodate large numbers of predictors and complex non-linear relationships. In HAI research, these methods have been explored for early detection, risk prediction, surveillance support and identification of patients who may benefit from additional infection-prevention measures.

A systematic review by Barbati and colleagues identified 121 studies evaluating artificial intelligence for HAI detection and prediction. The authors reported increasing research activity and found that AI approaches showed promising performance, while also identifying limitations in precision and the quality of the evidence base [6]. These findings indicate that the field is active, but the number of published models should not be confused with the number of models ready for routine clinical use.

A prediction model may perform well in the dataset used to develop it and perform less well when applied to a different hospital, patient population or period. This is particularly important in HAI research because infection prevalence, microbiological testing, antimicrobial resistance patterns, patient mix and clinical workflows can vary substantially between institutions. External evaluation is therefore essential before a model is used to guide decisions in a new setting.

Discrimination describes how well a model separates individuals who experience an outcome from those who do not. Calibration describes the agreement between predicted and observed risks. A

model may rank patients reasonably well but systematically overestimate or underestimate their absolute risk. Such errors can affect decisions about isolation, screening, surveillance intensity and allocation of infection-prevention resources. Evaluation should therefore consider more than a single area-under-the-curve statistic.

Interpretability is another consideration. Techniques such as feature-importance measures and SHAP values can help describe which variables contribute to model predictions. However, explaining how a model generated a prediction is not the same as identifying a causal mechanism. A variable may be predictive because it acts as a proxy for severity, healthcare exposure or another unmeasured factor. Interpretation should therefore remain cautious, particularly when model outputs are used to justify interventions.

7 Reporting, model evaluation and risk of bias

Transparent reporting is necessary for readers to assess how a prediction model was developed, evaluated and intended to be used. The TRIPOD+AI statement provides updated reporting guidance for prediction models developed using regression or machine-learning methods and for studies evaluating their performance [7]. It supports clear reporting of participant selection, outcome definitions, predictors, missing-data handling, model development, evaluation and intended use.

Reporting guidance should not be mistaken for a guarantee of methodological quality. A manuscript may report its methods clearly while still having limitations in participant selection, predictor measurement, outcome definition or analysis. Model evaluation should include internal validation to estimate optimism and external evaluation in independent data when possible. Calibration, discrimination and clinical utility should be considered in relation to the decision the model is intended to support.

The distinction between reporting and implementation is important. A well-reported model is easier to appraise and reproduce, but it is not automatically effective in practice. The central question is whether using the model changes decisions in a way that improves patient or health-system outcomes. This requires prospective evaluation in the clinical workflow, including attention to staff response, alert burden, data availability, resource requirements and unintended consequences.

8 Public-health applications and the implementation gap

HAI modelling has several potential applications. It can support analysis of surveillance trends, investigation of outbreaks, comparison of screening and isolation strategies, assessment of environmental cleaning, monitoring of antimicrobial resistance and planning of infection-prevention resources. Each application requires a clearly defined decision context. A model designed to estimate patient-level risk may not be suitable for estimating the impact of a hospital-wide intervention, while a transmission model may not provide sufficiently precise individual risk estimates for bedside decisions.

Economic evaluation can connect epidemiological modelling with decisions about resource allocation. For example, Brain and colleagues evaluated enhanced cleaning and disinfection of shared medical equipment using data from a stepped-wedge cluster randomised trial involving 5,002 inpatients. Their economic evaluation reported that the intervention was associated with reduced HAIs and cost savings compared with usual care [8]. This illustrates how infection-prevention interventions can be assessed not only in terms of infection outcomes but also in relation to resource use.

Economic conclusions, however, remain dependent on the intervention, setting, costing assumptions and study design.

The implementation gap arises when models are developed and published but are not evaluated in routine practice. Barriers may include limited electronic health-record infrastructure, incomplete microbiology data, lack of trained analytical personnel, poor interoperability, alert fatigue, and uncertainty about who is responsible for acting on model outputs. In resource-constrained settings, the complexity of a model may itself become a barrier if it requires data that are not routinely collected.

For India, local relevance is particularly important. Gandra and colleagues argue that infection-prevention strategies successful in high-income settings may not be directly transferable because of differences in epidemiology, infrastructure, care delivery, resources and microbiological patterns [2]. Modelling approaches should therefore be designed and evaluated with local data and local decision-makers, rather than imported without adaptation.

9 Research priorities

Several priorities emerge from the literature reviewed.

- **Strengthen surveillance data:** improve consistency of case definitions, microbiological reporting, device-exposure measurement and patient-level data capture across healthcare facilities.
- **Represent transmission beyond the hospital:** include community, long-term-care and inter-facility movement when these pathways are relevant to the pathogen and research question.
- **Improve model transparency:** state assumptions, data sources, missing-data procedures, parameter uncertainty and intended decision use.
- **Prioritise independent evaluation:** assess model performance across hospitals, time periods and patient populations, with particular attention to calibration and clinical utility.
- **Evaluate real-world impact:** determine whether model-informed decisions improve infection-prevention processes or patient outcomes, rather than relying only on retrospective predictive performance.
- **Build interdisciplinary capacity:** combine epidemiology, microbiology, clinical expertise, biostatistics, data science, health economics and implementation research.

These priorities are consistent with the need for stronger hospital epidemiology and infection-prevention research capacity in India, including the development of locally relevant research programmes and better use of existing surveillance infrastructure [2].

10 Conclusion

Epidemiological surveillance, mathematical transmission models, statistical approaches and machine-learning methods provide complementary tools for understanding and preventing health-care-associated infections. Compartmental models offer transparent representations of population-level transmission, while agent-based models can explore individual interactions and heterogeneous exposure. Statistical and machine-learning approaches can support risk estimation and detection, but their usefulness depends on careful evaluation and transportability. None of these methods can compensate for incomplete surveillance data or poorly defined outcomes.

The next stage of HAI modelling should be judged not simply by methodological complexity or predictive accuracy in a development dataset, but by the credibility of its assumptions, performance in independent settings and demonstrated contribution to infection-prevention decisions. Transparent reporting, context-sensitive validation and prospective implementation research are essential if modelling is to move from analytical promise to measurable improvement in patient safety.

Declarations

Funding: No external funding was received for the preparation of this narrative review.

Conflicts of interest: The author declares no conflict of interest related to this manuscript.

Ethics approval: Not applicable. This narrative review did not involve human participants, identifiable patient data or animal subjects.

Author contribution: The author conceptualised the review, undertook the targeted literature search, synthesised the literature and prepared the manuscript.

Data availability: Not applicable. No new dataset was generated or analysed for this narrative review.

References

[1] World Health Organization. *Global report on infection prevention and control 2024*. Geneva: World Health Organization; 2024. Available from: https://www.who.int/publications/i/item/9789240103986

[2] Gandra S, Vasudevan AK, Warren DK, Singh SK. Strengthening hospital epidemiology & infection prevention research capacity in India to curb antimicrobial resistance. *Indian J Med Res*. 2024;159(1):7-9. doi:10.4103/ijmr.ijmr_1919_23.

[3] Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev*. 2019;4:5. doi:10.1186/s41073-019-0064-8.

[4] Henriot P, El-Kassas M, Anwar W, Girgis SA, El Gaafary M, Jean K, et al. An agent-based model to simulate the transmission dynamics of bloodborne pathogens within hospitals. *PLoS Comput Biol*. 2025;21(2):e1012850. doi:10.1371/journal.pcbi.1012850.

[5] Lin G, Poleon S, Hamilton A, Salvekar N, Jara M, Haghpanah F, et al. The contribution of community transmission to the burden of hospital-associated pathogens: a systematic scoping review of epidemiological models. *One Health*. 2025;20:100951. doi:10.1016/j.onehlt.2024.100951.

[6] Barbati C, Viviani L, Vecchio R, Arzilli G, De Angelis L, Baglivo F, et al. Artificial intelligence use and performance in detecting and predicting healthcare-associated infections: a systematic review. *Artif Intell Med*. 2026;172:103321. doi:10.1016/j.artmed.2025.103321.

[7] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378.

[8] Brain D, Sivapragasam N, Browne K, White NM, Russo PL, Cheng AC, et al. Economic evaluation of enhanced cleaning and disinfection of shared medical equipment. *JAMA Netw Open*. 2025;8(4):e258565. doi:10.1001/jamanetworkopen.2025.8565.