



Universiteit  
Leiden  
The Netherlands

## **Leveraging AI-based prediction in perioperative and critical care: from model development to clinical implementation**

Meijden, S.L. van der

### **Citation**

Meijden, S. L. van der. (2025, May 6). *Leveraging AI-based prediction in perioperative and critical care: from model development to clinical implementation*. Retrieved from <https://hdl.handle.net/1887/4245255>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4245255>

**Note:** To cite this publication please use the final published version (if applicable).



# Chapter 4

## **Automated Identification of Postoperative Infections to Allow Prediction and Surveillance Based on Electronic Health Record Data: Scoping Review**

**S.L. van der Meijden**, A.M. van Boekel, H. van Goor, R.G.H.H. Nelissen,  
J.W. Schoones, E.W. Steyerberg, B.F. Geerts, M.G.J. de Boer, M.S. Arbous

*JMIR Medical Informatics, 10 Sep 2024*



## ABSTRACT

### Background

Postoperative infections remain a crucial challenge in health care, resulting in high morbidity, mortality, and costs. Accurate identification and labeling of patients with postoperative bacterial infections is crucial for developing prediction models, validating biomarkers, and implementing surveillance systems in clinical practice.

### Objective

This scoping review aimed to explore methods for identifying patients with postoperative infections using electronic health record (EHR) data to go beyond the reference standard of manual chart review.

### Methods

We performed a systematic search strategy across PubMed, Embase, Web of Science (Core Collection), the Cochrane Library, and Emcare (Ovid), targeting studies addressing the prediction and fully automated surveillance (ie, without manual check) of diverse bacterial infections in the postoperative setting. For prediction modeling studies, we assessed the labeling methods used, categorizing them as either manual or automated. We evaluated the different types of EHR data needed for the surveillance and labeling of postoperative infections, as well as the performance of fully automated surveillance systems compared with manual chart review.

### Results

We identified 75 different methods and definitions used to identify patients with postoperative infections in studies published between 2003 and 2023. Manual labeling was the predominant method in prediction modeling research, 65% (49/75) of the identified methods use structured data, and 45% (34/75) use free text and clinical notes as one of their data sources. Fully automated surveillance systems should be used with caution because the reported positive predictive values are between 0.31 and 0.76.

### Conclusions

There is currently no evidence to support fully automated labeling and identification of patients with infections based solely on structured EHR data. Future research should focus on defining uniform definitions, as well as prioritizing the development of more scalable, automated methods for infection detection using structured EHR data.

## 4.1. Introduction

Postoperative bacterial infections, including deep or superficial surgical site infections (SSIs), urinary tract infections (UTIs), and pneumonia, are the most frequent complications after surgery. Postoperative infections can be categorized into subtypes, usually based on location or severity according to the Clavien-Dindo classification [1]. The overall incidence of postoperative infections within 30 days of surgery varies between 6.5% and 25% [2-4]. Considering the 313 million patients undergoing surgery globally each year, these postoperative infections have an enormous impact on population health and overall health care costs [5]. Effective postoperative infection prevention and management require early detection of high-risk patients through prediction and data-driven surveillance. It is imperative for developing and validating prediction and surveillance systems to be able to accurately identify patients who have postoperative infections. Machine learning modeling practices use the term “labeling” for the identification of patients with the outcome of interest. Labeling and surveillance are both challenges due to underreporting in (hospital) complication registries, ranging from 38% to 77% when compared with a manual chart review [6,7]. Consequently, the current reference standard for identifying patients with postoperative infections relies on labor-intensive manual chart review, with an estimated 1.5 full-time equivalents per 10,000 admissions [8,9]. Furthermore, manual surveillance and labeling are prone to interobserver variability [10,11] and human errors [12], highlighting the need for more robust methods to address this devastating postoperative problem.

To achieve a more objective, cost-effective, and resource-efficient identification of patients with postoperative infections, it is imperative to leverage the electronic health records (EHRs) to automatically detect patients with infections without human checking on high-risk patients based on readily available EHR data. Different types of data are present within the EHR, including structured, tabular, and free-text records in which diagnoses and clinical symptoms are reported. A previously performed systematic review identified semiautomated and fully automated surveillance methods for hospital-acquired infections (HAIs) [13]. As more than 90% of the included systems required manual checking of infectious cases, it was concluded that fully automated surveillance of HAIs cannot be routinely used yet in health care settings.

To go beyond manual labeling and manual surveillance and to explore the current methods and criteria used in prediction modeling studies, the aim of this study was to perform a scoping review on available labeling methods for postoperative infections and fully automated surveillance systems (ie, not requiring manual checking). We aimed to (1) evaluate the current methods and criteria used to label patients with postoperative infections in prediction modeling and biomarker validation studies, (2) explore available automated surveillance methods and their performance (sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) in comparison with reference standard manual chart review, and (3) determine the necessary data types and sources needed to perform automated detection of postoperative infections.

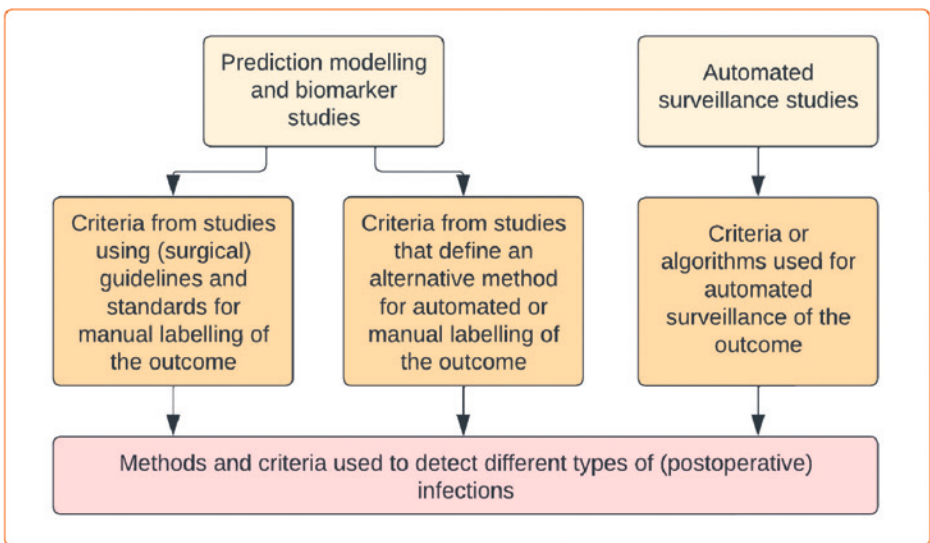
## 4.2. Methods

### 4.2.1. Overview

This scoping review combined 2 literature searches to evaluate current methods used by prediction modeling and biomarker validation studies to label patients with postoperative infections and the use of automated surveillance systems to identify patients with postoperative infections based on EHR data. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist was used. The protocol was registered on Open Science Framework [14].

### 4.2.2. Search Strategy

First, prediction modeling validation studies using machine learning methods, statistical models, and biomarkers to predict postoperative infections were identified. Second, a separate search was performed to identify studies on automated surveillance for postoperative and other hospital-acquired infections (Figure 1). Surveillance studies focusing on surgical populations often only investigate SSIs. As we aimed to study all bacterial infections that may occur after surgery, surveillance studies in a hospital-wide setting were also included. Both searches were performed in PubMed, Embase (OVID), Web of Science (Core Collection), the Cochrane Library, and Emcare (OVID). Studies were included from inception (ie, 1966) to August 1, 2023. The search queries were generated with help from an information specialist (JWS) from the Leiden University Medical Center. The details of the search queries are provided in Appendix A of Multimedia Appendix 1.



**Figure 1.** Data sources from the literature for identifying infections in prediction modeling or biomarker studies and automated surveillance studies.

#### 4.2.3. Selection Criteria

The selection of studies was performed in Covidence, a program used to manage systematic literature searches. The inclusion and exclusion criteria used are presented in Table S1 in Multimedia Appendix 1. All titles and abstracts were screened by 2 independent reviewers (AMVB and BFG for prediction models; SLVDM and BFG for surveillance studies). The full texts of all potentially relevant studies were retrieved and assessed by 2 reviewers (SLVDM and BFG) for eligibility. Any disagreement on the inclusion or exclusion of studies was resolved through reassessment and discussion with a third reviewer (MSA). The data from the different reports were collected by 1 researcher (AMVB or SLVDM), and inconsistencies were checked for by a second researcher (BFG).

#### 4.2.4. Data Extraction and Definitions

The following data were extracted for the prediction modeling studies: name of the prediction tool, type of prediction tool (machine learning, biomarker, and statistical model), surgical subpopulations, type of postoperative infection predicted, and criteria and guidelines used to manually or automatically label patients with infections. Manual labeling involves individuals conducting EHR chart reviews and applying specific criteria, often derived from surgical guidelines, to determine the presence or absence of infections in patient records. The criteria for diagnosing patients with an infection, for example, from a reference guideline from the literature, were identified and extracted.

For automated surveillance studies, the population, study design, years of data collection, type of infection surveyed, type of algorithm used, definition used to automatically detect infections, reference standard used to compare the automated method with, type of validation performed, and performance metrics reported compared with the reference standard were collected. The main metrics used to assess performance were the method's sensitivity, specificity, PPV, and NPV. Other metrics extracted are presented in Table S17 in Multimedia Appendix 1, including the area under the receiver operating characteristic curve, accuracy, F1-score,  $\kappa$  score, Pearson correlation coefficient, and agreement percentage. Only performance metrics were assessed for surveillance studies, as for prediction modeling studies, and no accuracy of the labeling method compared with a reference standard was determined.

#### 4.2.5. Data Synthesis

For each method to identify and label patients with infections, the data type categories needed from the EHR were assessed to identify infections based on the definition used. These could be structured EHR data (type A), including tabular information stored, such as complication registries, medication information, and vital signs; free-text clinical notes (type B), including all clinical information stored in free-text, such as discharge letters and daily reports; microbiology results (type C), which is seen as a separate category, as it differs per hospital how well-structured this information is stored [15]; and an additional interpretation layer (whether the results are positive) is needed to use this information; or imaging results (type D), or a combination of these categories. The definitions were further differentiated

based on the data types and criteria needed to adhere to the definitions in Appendix E in Multimedia Appendix 1.

Some prediction models and surveillance systems are focused on predicting or detecting all severity types of bacterial infections, while others focus only on infections requiring pharmacological or surgical treatment. For example, some definitions include the prescription of antibiotics as one of the criteria, while others base their criteria on clinical symptoms only. As the severity of the infections surveyed or predicted influences the intended use case and number of infections identified, we classified the definitions according to the Clavien-Dindo scale [1]. Finally, the performance of the automated infection surveillance systems compared with that of the reference standard manual review was visualized per subtype of infection. The results were grouped according to the type of infection, such as HAI (type not further specified), SSI, pneumonia, anastomotic leakage and abdominal infections, bloodstream infections (including central venous catheter-related infections and sepsis), and UTIs. Infections that did not belong to one of these groups were categorized as “other.”

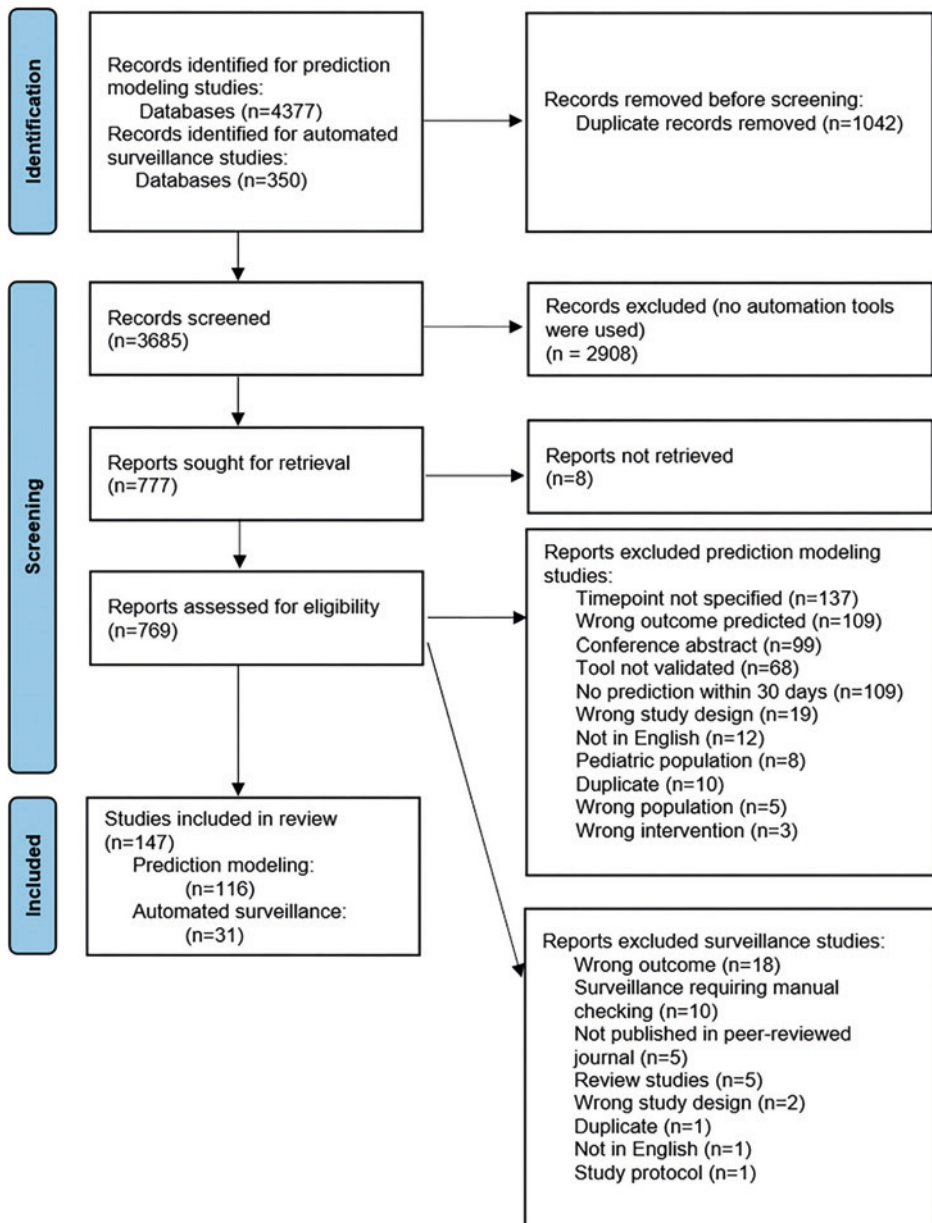
## **4.3. Results**

### **4.3.1. Overview**

We included a total of 147 studies published between 2003 and 2023 (Figure 2). Of these, 116 studies focused on the prediction of postoperative infections; either the development and validation of prediction models or a predictive biomarker were performed, or validation was performed of preexisting risk scores. These included the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator (ACS NSQIP; 33/116 studies), the National Nosocomial Infection Surveillance System (NNIS; 4/116 studies), and the Surgical Risk Preoperative Assessment System (SURPAS; 5/116 studies).

Out of 116 studies, 4 did not report the methodology used to determine which patients had the outcome of interest (ie, postoperative infection). In 83% (97/116) of prediction modeling studies, manual labeling based on diagnostic guidelines was performed, or a publicly available, manually labeled database was used, such as the participant use data file from the ACS NSQIP program (Table 1). A total of 13% (15/116) of studies used an alternative, non–guideline-based method to label patients with infections, 11 of whom used manual labeling, 3 of whom did not explicitly mention manual or automatic labeling, and 1 of whom used automatic labeling. In total, 93% (108/116) of the prediction modeling studies used manual labeling to determine the outcome of interest or a manually labeled, publicly available data set to perform their research.





**Figure 2.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.

**Table 1:** Definitions of patients with bacterial infections. Type A = structured EHR data, including tabular information stored such as complication registries, medication information, and vital signs; Type B = free-text clinical notes, including all clinical information stored in free-text such as discharge letters and daily reports; Type C = microbiology results; Type D = imaging results.

Type of infection	Origin of definition	Reference	A) Structured	B) Free-text	C) Microbiology	D) Imaging	Min. Clavien-Dindo
<b>HAI</b>							
	Diagnostic guidelines	WHO [19]	x	x	x	x	1
	Diagnostic guidelines	(E)CDC [20]	x	x			1
	Automated surveillance	Ehrentraut [21]		x			1
	Automated surveillance	Sakji [22]		x			1
	Automated surveillance	Tvardik [23]		x			1
<b>Pneumoniae</b>							
	Diagnostic guidelines	(E)CDC/ASC NSQIP [24]	x	x	x	x	1
	Prediction modeling	Kinlin [25]	x	x	x	x	1
	Automated surveillance	Blacky [26]	x		x		1
	Automated surveillance	Bouzbid [17]	x		x		2
	Automated surveillance	Cato [16]	x		x		1
	Automated surveillance	FitzHenry [27]		x			1
	Automated surveillance	Tvardik [23]		x			1
	Automated surveillance	Colborn [28]	x				1
	Automated surveillance	Stern [29]	x			x	1
<b>SSI</b>							
	Diagnostic guidelines	CDC/ASC NSQIP [30]	x	x	x	x	1-3a
	Diagnostic guidelines	WHO [31]	x	x	x	x	1-3a
	Prediction modeling	Daneman [32]	x				1
	Prediction modeling	Weller [18]	x				2
	Prediction modeling	Crispin [33]	x				3a
	Prediction modeling	Martin [34]	x	x			2
	Automated surveillance	Campillo-Gimenez [35]	x	x			1
	Automated surveillance	Cato [16]	x		x		1
	Automated surveillance	FitzHenry [27]		x			1
	Automated surveillance	Leclerc [36]	x		x		1
	Automated surveillance	Leth [37]	x		x		2
	Automated surveillance	Suzuki [38]	x	x	x		2
	Automated surveillance	Tvardik [23]		x			1

[continued on next page]

Table 1. [continued]

Type of infection	Origin of definition	Reference	A) Structured	B) Free-text	C) Microbiology	D) Imaging	Min. Clavien-Dindo
	Automated surveillance	Thirukumaran [39]		x			1
	Automated surveillance	Colborn [28]	x				1
Abdominal and AL							
	Diagnostic guidelines	Study Group of Rectal Cancer AL definition [40]	x	x		x	1
	Prediction modeling	Stidham [41]	x				3a
	Prediction modeling	Miyakita [42]	x				3b
	Prediction modeling	Mckenna [43]	x				2
	Prediction modeling	Nudel [44]	x	x			3a
	Prediction modeling	Kawai [45]		x		x	2
	Prediction modeling	Lin 2022 [46]		x		x	2
	Prediction modeling	Shi [47]				x	1
	Prediction modeling	Van Kooten [48]		x		x	1
UTI							
	Diagnostic guidelines	(E)CDC/ASC NSQIP [49]	x	x	x		2
	Prediction modeling	Cheng, Liu 2022 [50]		x	x		1
	Automated surveillance	Bouam 2003 [51]			x		1
	Automated surveillance	Bouzbid 2011 [17]	x		x		2
	Automated surveillance	Branch-Elliman [52]	x	x			1
	Automated surveillance	Cato [16]	x		x		1
	Automated surveillance	Choudhuri [53]	x		x		1
	Automated surveillance	FitzHenry [27]		x			1
	Automated surveillance	Leth [37]	x		x		2
	Automated surveillance	Redder [54]	x		x		2
	Automated surveillance	Tvardik [23]		x			1
	Automated surveillance	vanderWerff [55]	x	x	x		2
	Automated surveillance	Venable [56]	x	x			N/A
	Automated surveillance	Wald [57]	x		x		1
	Automated surveillance	Colborn [28]	x				1
Bloodstream infections							
	Diagnostic guidelines	ACS NSQIP [58]	x		x		1
	Diagnostic guidelines	ACCP/SCCM (sepsis-3 criteria) [59]	x				2
	Automated surveillance	Blacky [26]	x		x		1

[continued on next page]

Table 1. [continued]

Type of infection	Origin of definition	Reference	A) Structured	B) Free-text	C) Microbiology	D) Imaging	Min. Clavien-Dindo
	Automated surveillance	Bouam [51]			x		1
	Automated surveillance	Bouzbid [17]	x		x		2
	Automated surveillance	Cato [16]	x		x		1
	Automated surveillance	FitzHenry [27]		x			1
	Automated surveillance	Leal [60]			x		1
	Automated surveillance	Leal [61]			x		1
	Automated surveillance	Lin [62]			x		1
	Automated surveillance	Redder [54]	x		x		2
	Automated surveillance	Tvardik [23]		x			1
	Automated surveillance	Valik [63]	x		x		2
	Automated surveillance	Venable [56]	x	x			N/A
	Automated surveillance	Woeltje [64]	x		x		1
	Automated surveillance	Colborn [28]	x				1
C. Difficile							
	Automated surveillance	Dubberke [65]			x		1
	Automated surveillance	vanderWerff [66]			x		1
External ventricular and lumbar drain-related meningitis							
	Automated surveillance	vanMourik [67]	x		x		1
MRSA							
	Automated surveillance	Peterson [68]			x		1
PJI							
	Automated surveillance	Fu [69]		x			1
Neurological							
	Prediction modeling	Cheng, Bai [70]	x	x	x		1

Abbreviations: AL = Anastomotic leakage, ACS NSQIP = American College of Surgeons National Surgical Quality Improvement Program, CDC = Centers for Disease Control and Prevention, HAI = hospital-acquired Infections, PREZIES = PREventie van ZIEkenhuisinfecties door surveillance (prevention of hospital infections through surveillance, a Dutch initiative), SSISS = Surgical Site Infection Surveillance Service, NHSN = National Healthcare Safety Network (CDC), IDSA = Infectious Diseases Society of America, ACCP/SCCM = American College of Chest Physicians/Society of Critical Care Medicine, STS = Society of Thoracic Surgeons, VASQIP = Veterans Affairs Surgical Quality Improvement Program, WHO = World Health Organization, EAU = European Association of Urology, ECDC = European Centre for Disease Prevention and Control, PHE = Public Health England, NICE = The National Institute for Health and Care Excellence.

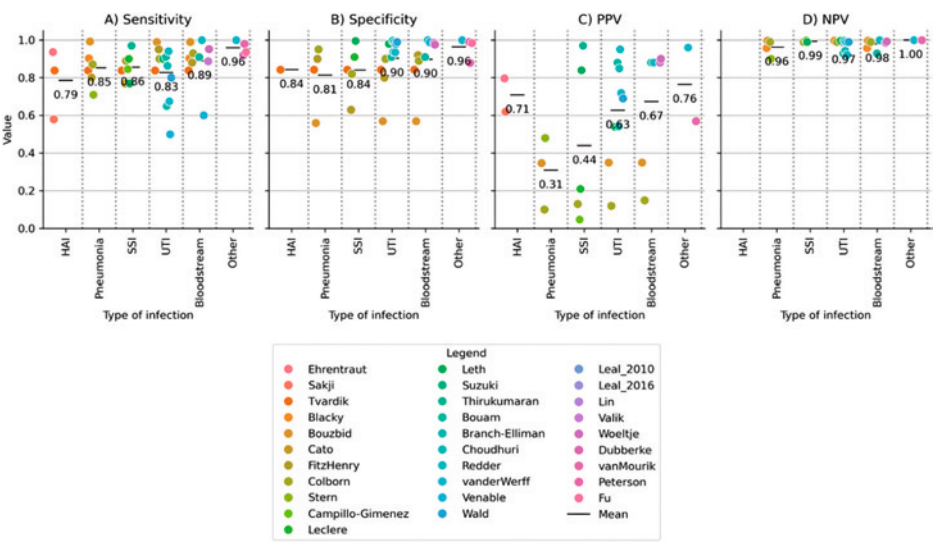
#### 4.3.2. Automated Surveillance

We included 31 automated surveillance studies for bacterial infections. Surveillance was performed and reported per patient, admission, procedure, patient days, or culture. Different types of surveillance systems were studied, and some studies have reported on more than 1 method. Most often (21/31, 68%), a set of criteria or rules was defined to automatically detect infections based on EHR data, followed by natural language processing (NLP) algorithms for free-text from the EHR (7/31, 23%) and other classification algorithms such as logistic regression (3/31, 10%). Except for one study [25], all the studies validated their automated surveillance algorithms against a reference standard (manual chart review, often according to one of the established diagnostic guidelines). Comparing the automated surveillance algorithm to manual chart review according to the established guidelines resulted in a range of sensitivity (0.79-0.96), specificity (0.81-0.96), PPV (0.31-0.76), and NPV (0.96-1.00) estimates for the different types of infection (Figure 3). The performance of all the combinations of postoperative infection data needed to run the automated surveillance algorithm varied (Figure 4). Reported performance per surveillance algorithm is provided in Table S17 in Multimedia Appendix 1.

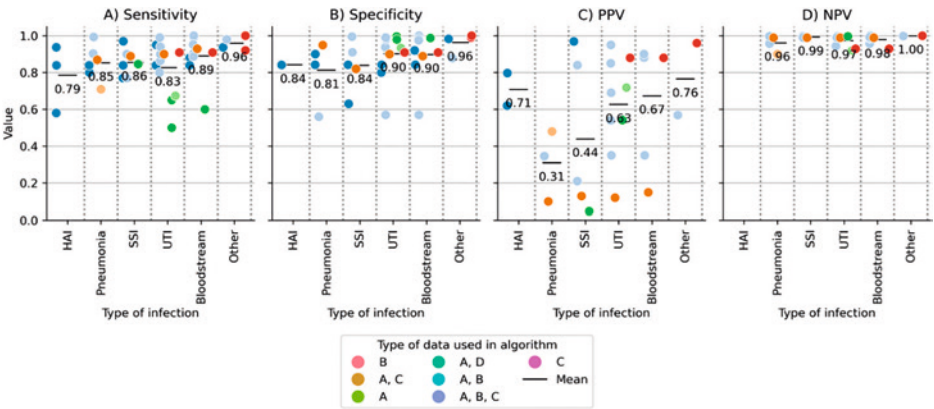
#### 4.3.3. Electronic Health Record Data for Automated Identification and Surveillance

In the 147 included studies, 75 different methods and definitions were used to identify different types of bacterial infections. A total of 56% (42/75) used 2 or more datatypes to label, diagnose, or surveil infections, and 45% (34/75) required free-text and clinical notes as at least one of their data sources. In Table 1, the different types of data from the EHR needed to automatically detect patients with an infection are specified for each diagnostic guideline or infection definition used in the different prediction modeling studies or automated surveillance methods. Figure 5 shows the total number of methods used to identify patients with bacterial infections and the different data categories used. Most frequently (20/75, 27%), a combination of microbiology results and structured EHR data was used, followed by free-text (13/75, 17%) and structured EHR data (11/75, 15%). In total, 45% (34/75) of the identified methods used free text and clinical notes as one of their data sources.

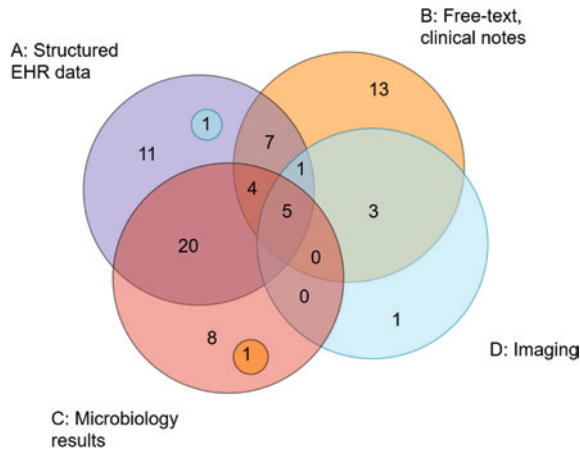




**Figure 3.** Performance of automated surveillance of postoperative infections compared with manual reference standard chart review. Panel A is the sensitivity, B is the specificity, C is the PPV, and D is the NPV. HAI: hospital-acquired infection; NPV: negative predictive value; PPV: positive predictive value; SSI: surgical site infection; UTI: urinary tract infection.



**Figure 4.** Performance per data type category used in automated surveillance algorithms. A=Structured electronic health record data only (eg, registrations and medication), B=Free-text clinical notes, C=microbiology results. Panel A is the sensitivity, B is the specificity, C is the PPV and D is the NPV. HAI: hospital-acquired infection; NPV: negative predictive value; PPV: positive predictive value; SSI: surgical site infection; UTI: urinary tract infection.



**Figure 5.** Venn diagram of the types of data used to identify bacterial infections in the included studies and guidelines. The data were divided into structured electronic health record data, free-text and clinical notes, microbiology results, and imaging. In total, 75 unique definitions were identified for different types of bacterial infections. EHR: electronic health record.

For hospital-acquired infections (no specification of subtype), free-text information was needed for all definitions and methods, limiting the ability to detect patients with an infection based on structured EHR data. For pneumonia, some automated surveillance studies have identified patients without the need for free-text information [24,25], but they did include culture results in their definition. For SSIs and UTIs, a wide range of criteria were used compared with other types of infections. Abdominal surgery-related anastomotic leakage and abdominal infections were identified based on antibiotic treatment or surgical reinterventions supplemented with free-text data or imaging results. Bacterial culture data, in combination with structured EHR parameters, are used in most methods for detecting bloodstream infections. For *Clostridium difficile* infections, cerebral extraventricular and lumbar drain-related meningitis, methicillin-resistant *Staphylococcus aureus*, and prosthetic joint infection, the authors used a maximum of 2 criteria from different categories to define infection. Prediction modeling studies that did not use manual chart review for labeling patients in the data set relied on the registration of infections or the performance of surgical interventions, sometimes in combination with antibiotic administration [32].

When assessing infection severity according to the different Clavien-Dindo definitions, most (64%, 48/75) were based on identifying infections according to a Clavien-Dindo score of 1 or more. This indicates that, based on the registration of infection or clinical criteria only, infections were surveyed and predicted. In 23% of definitions (17/75), the prescription of antibiotic therapy or surgical intervention was included as the criterion, resulting in a Clavien-Dindo score of 2 or higher.

## 4.4. Discussion

This scoping review assessed the methods and criteria used for identifying postoperative bacterial infections in prediction modeling and fully automated surveillance studies. We identified a total of 75 different methods and definitions from 147 included studies to identify patients with different types of bacterial infections. We found that 45% (34/75) used unstructured free-text and clinical notes as at least one of their data sources. Furthermore, out of 116 postoperative infection prediction studies, 108 (93%) used manual labeling based on self-defined criteria or diagnostic guidelines or used publicly available manually labeled databases. In addition, among the 31 automated surveillance studies, various methods, such as NLP, classification algorithms, and predefined criteria or rules on structured data, were used to automatically detect infections. Compared with manual chart review, automated surveillance systems have reported sensitivities for different types of infections ranging from 0.79 to 0.96, specificities from 0.81 to 0.96, PPVs from 0.31 to 0.76, and NPVs from 0.96 to 1.00. Finally, we found that different criteria were used among both prediction and surveillance studies to identify patients with infections, indicating that there is no uniform definition being used. Given the current use of different types of criteria and data used in prediction and surveillance studies, we were not able to identify or formulate a uniform and reliable method to automatically label patients with infections based on structured EHR data.

Prediction and surveillance of postoperative infections are crucial for early detection and assessment of the impact of preventative interventions but are currently hindered because the labeling of these cases is performed by resource-intensive manual chart review. In contrast to a previous study on semiautomated surveillance where high-risk patients were manually checked [13], we included only fully automated surveillance systems that were built to avoid requiring any human intervention. However, human intervention might still be required to incorporate the systems as well as to clean and preprocess the EHR data. Furthermore, we broadened the scope by assessing current labeling methods for prediction modeling studies, which, with some exceptions, were based on manual labeling according to established guidelines. In line with our findings, the predominant use of manual labeling was reported in a meta-analysis on the predictive performance of machine learning algorithms for SSI prediction [71]. Although manual labeling based on chart review is still the predominant method and is considered the reference standard, it must be noted that manual labeling may be flawed due to human errors and interobserver variability [9,10,12]. Furthermore, validating models only on national registries and databases limits the generalizability of developed prediction models and surveillance systems to other settings [72].

We extensively researched different definitions and methods from prediction modeling studies, guidelines, and surveillance studies to identify patients with bacterial infections that may occur after surgery and summarized different types of data needed to adhere to the different definitions. This study has several limitations. First, heterogeneity between studies (eg, differences in study design) prevented a meta-analysis, making it difficult to draw generalizable conclusions on optimal labeling methods. However, combining different types of studies

allowed us to generate insight into the current methods of labeling and identifying patients with infections. Second, the distinction between structured and unstructured data may differ according to hospital data set and region (eg, microbiology results can be registered as free-text or tabular data). Despite these limitations, we could identify a lack of uniform definitions for labeling of postoperative infections exists, and that manual labeling is currently the predominant method. Third, pre-existing infections could have impacted the performance of surveillance algorithms and prediction models as well as label reliability [73]. This could explain the relatively lower PPVs and warrants further research before reliable implementation of automated surveillance systems.

Different types of data were used among the definitions and methods, including structured tabular data, microbiological data, free-text data, and imaging results. The importance of reliable, high-quality outcome data is essential for the reliable use of artificial intelligence and surveillance systems [74]. Using structured EHR data is preferable, as free text is often subject to misinterpretation and contains personal patient-specific data that conflict with privacy legislation and thus have restrictions on data use [75]. By extracting free-text information, NLP shows promise in uncovering postoperative infections from free-text data. However, challenges remain with respect to generalizability [76], transparency, reliability, and potential biases, including concerns about accuracy or unintended errors [77,78]. Furthermore, NLP methods can be computationally expensive, depend on the quality of the input data, and are influenced by nuances in language, dialects, and medical jargon. Considering that NLP methods can vary significantly in complexity, ranging from simple string searches to advanced neural networks, future research should investigate whether increased complexity leads to improved surveillance accuracy. The use of microbiology results in definitions is prevalent, despite their occasional unreliability due to the possibility of false negatives or positives, causing under- or overreporting of infections [13], and heterogeneous storage practices. This reliance on microbiology results could lead to errors or inconsistencies in infection identification.

Accurately identifying patients with infections based on an automated analysis of EHR data remains a challenge, and validation is difficult owing to the limitations of manual chart review, which until now has remained the reference standard for postoperative infections and other relevant patient outcomes. Manual labeling based on manual EHR chart review is unfeasible when scaling artificial intelligence-based or statistical prediction models to more than one hospital, with 100,000 patient records each. In some of the included studies, alternative approaches were identified that relied on treatments and other structured data sources [27,31-33,41-43]. For future prediction model development and surveillance, alternative approaches to identifying patients with infection should be explored, such as focusing on pharmacological and interventional treatments performed by clinicians, as these approaches are often stored in a structured format in the EHR system [27]. Emphasis should be placed on the consensus on the definition and whether it is worse to miss infections that do not require treatment compared with those that do. Compared with sensitivity, specificity, and NPV, automated surveillance systems have a lower PPV where heterogeneity is observed between the different

types of infections. The PPV to detect pneumonia and SSIs is lower compared with other types of infections. This could be due to variations in clinical presentation, differences in diagnostic criteria, or the inherent complexity and variability of these particular infections. A lower PPV in general could be due to the use of low classification cutoffs to not miss any cases, but it could also indicate that the reference standard manual labeling may have resulted in erroneous labels and that the systems found infections where the human annotator did not [79]. In addition to detecting individual patients with infections, automated surveillance systems hold promise for assessing hospital incidence rates, predicting rates of complications, and evaluating the effectiveness of quality improvement initiatives, where the emphasis may shift from high PPVs to broader statistical insights.

In conclusion, there is currently no evidence to support fully automated labeling and identification of patients with infections based solely on structured EHR data. This is due to the diverse definitions of postoperative infection and the need for unstructured data types, such as free text and clinical notes, which were required as data sources in nearly half of the instances to assess an infection. Furthermore, manual labeling was still the predominant method in prediction modeling studies. Fully automatic surveillance methods may result in overreporting due to a relatively low PPV and heavy reliance on free-text data. Future research must focus on defining uniform or globally accepted definitions of postoperative infection that use criteria that can be extracted from the EHR, as well as prioritizing the development of more scalable automated methods for infection detection using EHR data.

## Multimedia Appendices

Multimedia Appendices are available on: [medinform.jmir.org/2024/1/e57195](https://medinform.jmir.org/2024/1/e57195)





## REFERENCES

1. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg*. 2009;250(2):187-96.
2. International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. *Br J Anaesth*. 2016;117(5):601-9.
3. Gillespie BM, Harbeck E, Rattray M, Liang R, Walker R, Latimer S, et al. Worldwide incidence of surgical site infections in general surgical patients: a systematic review and meta-analysis of 488,594 patients. *Int J Surg*. 2021;95:106136.
4. Wan YI, Patel A, Achary C, Hewson R, Phull M, Pearse RM, et al. International Surgical Outcomes Study (ISOS) Group. Postoperative infection and mortality following elective surgery in the international surgical outcomes study (ISOS). *Br J Surg*. 2021;108(2):220-7.
5. Weiser TG, Haynes AB, Molina G, Lipsitz SR, Esquivel MM, Uribe-Leitz T, et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet*. 2015;385 Suppl 2.
6. Ubbink DT, Visser A, Gouma DJ, Goslings JC. Registration of surgical adverse outcomes: a reliability study in a university hospital. *BMJ Open*. 2012;2(3).
7. Veen EJ, Janssen-Heijnen MLG, Bosma E, de Jongh MAC, Roukema JA. The accuracy of complications documented in a prospective complication registry. *J Surg Res*. 2012;173(1):54-9.
8. Du M, Xing Y, Suo J, Liu B, Jia N, Huo R, et al. Real-time automatic hospital-wide surveillance of nosocomial infections and outbreaks in a large Chinese tertiary hospital. *BMC Med Inform Decis Mak*. 2014;14:9.
9. Brossette SE, Hacek DM, Gavin PJ, Kamdar MA, Gadbois KD, Fisher AG, et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance? *Am J Clin Pathol*. 2006;125(1):34-9.
10. Klompas M. Interobserver variability in ventilator-associated pneumonia surveillance. *Am J Infect Control*. 2010;38(3):237-93.
11. Tokars JL, Richards C, Andrus M, Klevens M, Curtis A, Horan T, et al. The changing face of surveillance for health care-associated infections. *Clin Infect Dis*. 2004;39(9):1347-52.
12. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2018;25(10):1419-28.
13. Streefkerk HRA, Verkooijen RP, Bramer WM, Verbrugh HA. Electronically assisted surveillance systems of healthcare-associated infections: a systematic review. *Euro Surveill*. 2020;25(2):1900321.
14. van der Meijden SL. Scoping Review Protocol - Automated Detection of Postoperative Infections to Allow Prediction and Surveillance Based on EHR Data - a Scoping Review. 2022. URL: <https://osf.io/4cuge/> [accessed 2023-11-29].
15. Rhoads DD, Sintchenko V, Rauch CA, Pantanowitz L. Clinical microbiology informatics. *Clin Microbiol Rev*. 2014;27(4):1025-47.
16. Prevention of hospital-acquired infections. World Health Organization. 2002. URL: [https://iris.who.int/bitstream/handle/10665/67350/WHO\\_CDS\\_CSR\\_EPH\\_2002.12.pdf](https://iris.who.int/bitstream/handle/10665/67350/WHO_CDS_CSR_EPH_2002.12.pdf) [accessed 2024-08-14].
17. CDC/NHSN surveillance definitions for specific types of infections. National Healthcare Safety Network. 2024. URL: [https://www.cdc.gov/nhsn/pdfs/pscmanual/17pscnosinfdef\\_current.pdf](https://www.cdc.gov/nhsn/pdfs/pscmanual/17pscnosinfdef_current.pdf) [accessed 2024-01-01].
18. Ehrentraut C, Ekholm M, Tanushi H, Tiedemann J, Dalianis H. Detecting hospital-acquired infections: a document classification approach using support vector machines and gradient tree boosting. *Health Informatics J*. 2018;24(1):24-42.
19. Sakji S, Gicquel Q, Pereira S, Kergourlay I, Proux D, Darmoni S, et al. Evaluation of a French medical multi-terminology indexer for the manual annotation of natural language medical reports of healthcare-associated infections. *Stud Health Technol Inform*. 2010;160(Pt 1):252-6.
20. 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.
21. Pneumonia (Ventilator-associated [VAP] and non-ventilator-associated pneumonia [PNEU]) event. National Healthcare Safety Network. 2024. URL: <https://www.cdc.gov/nhsn/pdfs/pscmanual/6pscvapcurrent.pdf> [accessed 2024-01-01].
22. Kinlin LM, Kirchner C, Zhang H, Daley J, Fisman DN. Derivation and validation of a clinical prediction rule for nosocomial pneumonia after coronary artery bypass graft surgery. *Clin Infect Dis*. 2010;50(4):493-501.

23. Blacky A, Mandl H, Adlassnig KP, Koller W. Fully automated surveillance of healthcare-associated infections with MONI-ICU: a breakthrough in clinical infection surveillance. *Appl Clin Inform.* 2011;2(3):365-72.
24. Bouzbid S, Gicquel Q, Gerbier S, Chomarat M, Pradat E, Fabry J, et al. Automated detection of nosocomial infections: evaluation of different strategies in an intensive care unit 2000-2006. *J Hosp Infect.* 2011;79(1):38-43.
25. Cato KD, Liu J, Cohen B, Larson E. Electronic surveillance of surgical site infections. *Surg Infect (Larchmt).* 2017;18(4):498-502.
26. FitzHenry F, Murff HJ, Matheny ME, Gentry N, Fielstein EM, Brown SH, et al. Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Med Care.* 2013;51(6):509-16.
27. Colborn KL, Zhuang Y, Dyas AR, Henderson WG, Madsen HJ, Bronsert MR, et al. Development and validation of models for detection of post-operative infections using structured electronic health records data and machine learning. *Surgery.* 2023;173(2):464-71.
28. Stern SE, Christensen MA, Nevers MR, Ying J, McKenna C, Munro S, et al. Electronic surveillance criteria for non-ventilator-associated hospital-acquired pneumonia: assessment of reliability and validity. *Infect Control Hosp Epidemiol.* 2023;1-7.
29. Surgical site infection event (SSI). National Healthcare Safety Network. 2024. URL: <https://www.cdc.gov/nhsn/pdfs/pscmanual/9pscscsscurrent.pdf> [accessed 2024-01-01].
30. Global guidelines for the prevention of surgical site infection. World Health Organization. 2018. URL: <https://www.who.int/publications/i/item/9789241550475> [accessed 2023-12-01].
31. Daneman N, Simor AE, Redelmeier DA. Validation of a modified version of the national nosocomial infections surveillance system risk index for health services research. *Infect Control Hosp Epidemiol.* 2009;30(6):563-569.
32. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. *Stat Methods Med Res.* 2018;27(11):3271-3285.
33. Crispin A, Klinger C, Rieger A, Strahwald B, Lehmann K, Buhr HJ, et al. The DGAV risk calculator: development and validation of statistical models for a web-based instrument predicting complications of colorectal cancer surgery. *Int J Colorectal Dis.* 2017;32(10):1385-1397.
34. Martin S, Turner E, Nguyen A, Thornton B, Nazerali RS. An evaluation of the utility of the Breast Reconstruction Risk Assessment score risk model in prepectoral tissue expander breast reconstruction. *Ann Plast Surg.* 2020;84(5S Suppl 4).
35. Campillo-Gimenez B, Garcelon N, Jarno P, Chaplain JM, Cuggia M. Full-text automated detection of surgical site infections secondary to neurosurgery in Rennes, France. *Stud Health Technol Inform.* 2013;192:572-575.
36. Leclère B, Lasserre C, Bourigault C, Juvin ME, Chaillet MP, Mauduit N, et al. Matching bacteriological and medico-administrative databases is efficient for a computer-enhanced surveillance of surgical site infections: retrospective analysis of 4,400 surgical procedures in a French university hospital. *Infect Control Hosp Epidemiol.* 2014;35(11):1330-1335.
37. Leth RA, Nørgaard M, Uldbjerg N, Thomsen RW, Møller JK. Surveillance of selected post-caesarean infections based on electronic registries: validation study including post-discharge infections. *J Hosp Infect.* 2010;75(3):200-204.
38. Suzuki H, Clore GS, Perencevich EN, Hockett-Sherlock SM, Goto M, Nair R, et al. Development of a fully automated surgical site infection detection algorithm for use in cardiac and orthopedic surgery research. *Infect Control Hosp Epidemiol.* 2021;42(10):1215-1220.
39. Thirukumaran CP, Zaman A, Rubery PT, Calabria C, Li Y, Ricciardi BF, et al. Natural language processing for the identification of surgical site infections in orthopaedics. *J Bone Joint Surg Am.* 2019;101(24):2167-2174.
40. Rahbari NN, Weitz J, Hohenberger W, Heald RJ, Moran B, Ulrich A, et al. Definition and grading of anastomotic leakage following anterior resection of the rectum: a proposal by the international study group of rectal cancer. *Surgery.* 2010;147(3):339-351.
41. Stidham RW, Waljee AK, Day NM, Bergmans CL, Zahn KM, Higgins PDR, et al. Body fat composition assessment using analytic morphomics predicts infectious complications after bowel resection in Crohn's disease. *Inflamm Bowel Dis.* 2015;21(6):1306-1313.
42. Miyakita H, Sadahiro S, Saito G, Okada K, Tanaka A, Suzuki T. Risk scores as useful predictors of perioperative complications in patients with rectal cancer who received radical surgery. *Int J Clin Oncol.* 2017;22(2):324-331.
43. McKenna NP, Bews KA, Cima RR, Crowson CS, Habermann EB. Development of a risk score to predict anastomotic leak after left-sided colectomy: which patients warrant diversion? *J Gastrointest Surg.* 2020;24(1):132-143.

44. Nudel J, Bishara AM, de Geus SWL, Patil P, Srinivasan J, Hess DT, et al. Development and validation of machine learning models to predict gastrointestinal leak and venous thromboembolism after weight loss surgery: an analysis of the MBSA-QIP database. *Surg Endosc*. 2021;35(1):182-191.
45. Kawai K, Hirakawa S, Tachimori H, Oshikiri T, Miyata H, Kakeji Y, et al. Updating the predictive models for mortality and morbidity after low anterior resection based on the National Clinical Database. *Dig Surg*. 2023;40(3-4):130-142.
46. Lin V, Tsouchnika A, Allakhverdiiev E, Rosen AW, Gögenur M, Clausen JSR, et al. Training prediction models for individual risk assessment of postoperative complications after surgery for colorectal cancer. *Tech Coloproctol*. 2022;26(8):665-675.
47. Shi J, Wu Z, Wu X, Shan F, Zhang Y, Ying X, et al. Early diagnosis of anastomotic leakage after colorectal cancer surgery using an inflammatory factors-based score system. *BJS Open*. 2022;6(3).
48. van Kooten RT, Bahadoer RR, Ter Buurkes de Vries B, Wouters MWJM, Tollenaar RAEM, Hartgrink HH, et al. Conventional regression analysis and machine learning in prediction of anastomotic leakage and pulmonary complications after esophagogastric cancer surgery. *J Surg Oncol*. 2022;126(3):490-501.
49. Urinary Tract Infection (Catheter-Associated Urinary Tract Infection [CAUTI] and Non-Catheter-Associated Urinary Tract Infection [UTI]) Events. National Healthcare Safety Network. 2024. URL: <https://www.cdc.gov/nhsn/pdfs/pscmanual/7p-scauticurrent.pdf> [accessed 2024-01-01]
50. Cheng X, Liu Y, Wang W, Yan J, Lei X, Wu H, et al. Preoperative risk factor analysis and dynamic online nomogram development for early infections following primary hip arthroplasty in geriatric patients with hip fracture. *Clin Interv Aging*. 2022;17:1873-1883.
51. Bouam S, Girou E, Brun-Buisson C, Karadimas H, Lepage E. An intranet-based automated system for the surveillance of nosocomial infections: prospective validation compared with physicians' self-reports. *Infect Control Hosp Epidemiol*. 2003;24(1):51-55.
52. 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(9):1004-1010.
53. Choudhuri JA, Pergamit RF, Chan JD, Schreuder AB, McNamara E, Lynch JB, et al. An electronic catheter-associated urinary tract infection surveillance tool. *Infect Control Hosp Epidemiol*. 2011;32(8):757-762.
54. Redder JD, Leth RA, Møller JK. Incidence rates of hospital-acquired urinary tract and bloodstream infections generated by automated compilation of electronically available healthcare data. *J Hosp Infect*. 2015;91(3):231-236.
55. van der Werff SD, Thiman E, Tanushi H, Valik JK, Henriksson A, Ul Alam MU, et al. The accuracy of fully automated algorithms for surveillance of healthcare-associated urinary tract infections in hospitalized patients. *J Hosp Infect*. 2021;110:139-147.
56. Venable A, Dissanaike S. Is automated electronic surveillance for healthcare-associated infections accurate in the burn unit? *J Burn Care Res*. 2013;34(6):591-597.
57. Wald HL, Bandle B, Richard A, Min S. Accuracy of electronic surveillance of catheter-associated urinary tract infection at an academic medical center. *Infect Control Hosp Epidemiol*. 2014;35(6):685-691.
58. Moore LJ, Moore FA, Todd SR, Jones SL, Turner KL, Bass BL. Sepsis in general surgery: the 2005-2007 National Surgical Quality Improvement Program perspective. *Arch Surg*. 2010;145(7):695-700.
59. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315(8):801-810.
60. Leal J, Gregson DB, Ross T, Flemons WW, Church DL, Laupland KB. Development of a novel electronic surveillance system for monitoring of bloodstream infections. *Infect Control Hosp Epidemiol*. 2010;31(7):740-747.
61. Leal JR, Gregson DB, Church DL, Henderson EA, Ross T, Laupland KB. The validation of a novel surveillance system for monitoring bloodstream infections in the Calgary Zone. *Can J Infect Dis Med Microbiol*. 2016;2016:2935870.
62. Lin MY, Woeltje KF, Khan YM, Hota B, Doherty JA, Borlawsky TB, et al. Centers for Disease Control Prevention Epicenter Program. Multicenter evaluation of computer automated versus traditional surveillance of hospital-acquired bloodstream infections. *Infect Control Hosp Epidemiol*. 2014;35(12):1483-1490.
63. Valik JK, Ward L, Tanushi H, Müllersdorf K, Ternhag A, Aufwerber E, et al. Validation of automated sepsis surveillance based on the Sepsis-3 clinical criteria against physician record review in a general hospital population: observational study using electronic health records data. *BMJ Qual Saf*. 2020;29(9):735-745.

64. Woeltje KF, McMullen KM, Butler AM, Goris AJ, Doherty JA. Electronic surveillance for healthcare-associated central line-associated bloodstream infections outside the intensive care unit. *Infect Control Hosp Epidemiol*. 2011;32(11):1086-1090.
65. Dubberke ER, Nyazee HA, Yokoe DS, Mayer J, Stevenson KB, Mangino JE, et al. Implementing automated surveillance for tracking *Clostridium difficile* infection at multiple healthcare facilities. *Infect Control Hosp Epidemiol*. 2012;33(3):305-308.
66. van der Werff SD, Fritzing M, Tanushi H, Henriks-son A, Dalianis H, Ternhag A, et al. The accuracy of fully automated algorithms for surveillance of healthcare-onset infections in hospitalized patients. *Antimicrob Steward Healthc Epidemiol*. 2022;2(1)
67. van Mourik MSM, Groenwold RHH, Berkelbach van der Sprenkel JW, van Solinge WW, Troelstra A, Bonten MJM. Automated detection of external ventricular and lumbar drain-related meningitis using laboratory and microbiology results and medication data. *PLoS One*. 2011;6(8)
68. Peterson KE, Hacek DM, Robicsek A, Thomson Jr RB, Peterson LR. Electronic surveillance for infectious disease trend analysis following a quality improvement intervention. *Infect Control Hosp Epidemiol*. 2012;33(8):790-795.
69. Fu S, Wyles CC, Osmon DR, Carvour ML, Sagheb E, Ramazanian T, et al. Automated detection of periprosthetic joint infections and data elements using natural language processing. *J Arthroplasty*. 2021;36(2):688-692.
70. Cheng L, Bai W, Song P, Zhou L, Li Z, Gao L, et al. Development and validation of a nomograph model for post-operative central nervous system infection after craniocerebral surgery. *Diagnostics (Basel)*. 2023;13(13):2207.
71. Wu G, Khair S, Yang F, Cheliger C, Southern D, Zhang Z, et al. Performance of machine learning algorithms for surgical site infection case detection and prediction: a systematic review and meta-analysis. *Ann Med Surg (Lond)*. 2022;84:104956.
72. de Hond AAH, Shah VB, Kant IMJ, Van Calster B, Steyerberg EW, Hernandez-Boussard T. Perspectives on validation of clinical predictive algorithms. *NPJ Digit Med*. 2023;6(1):86.
73. van Rooden SM, Tacconelli E, Pujol M, Gomila A, Kluytmans JAJW, Romme J, et al. A framework to develop semiautomated surveillance of surgical site infections: an international multicenter study. *Infect Control Hosp Epidemiol*. 2020;41(2):194-201.
74. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med*. 2022;5(1):2.
75. Humbert-Droz M, Mukherjee P, Gevaert O. Strategies to address the lack of labeled data for supervised machine learning training with electronic health records: case study for the extraction of symptoms from clinical notes. *JMIR Med Inform*. 2022;10(3)
76. Khambete MP, Su W, Garcia JC, Badgeley MA. Quantification of BERT diagnosis generalizability across medical specialties using semantic dataset distance. *AMIA Jt Summits Transl Sci Proc*. 2021;2021:345-354.
77. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. *Nat Med*. 2023;29(10):2396-2398.
78. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. 2021;27(12):2176-2182.
79. Vassar M, Holzmann M. The retrospective chart review: important methodological considerations. *J Educ Eval Health Prof*. 2013;10:12.

