SURGICAL INFECTIONS
Volume 21, Number 8, 2020

DOI: 10.1089/sur.2019.238

Developing a Surgical Site Infection Surveillance System Based on Hospital Unstructured Clinical Notes and Text Mining

Marta Luisa Ciofi Degli Atti,[1] Fabrizio Pecoraro,[2] Simone Piga,[1]
Daniela Luzi,[2] and Massimiliano Raponi[3]

Abstract

Background: Electronic surveillance using clinical and administrative data from multiple sources has been reported as a tool for surveillance of surgical site infections (SSIs), but experiences are limited. In this study, we aimed to assess the accuracy of a text-searching algorithm to detect SSIs in children based on the application of regular expressions of unstructured clinical notes collected through different information systems.
Methods: We developed an information system data warehouse that integrates data provided by electronic health and administrative records for patients who underwent surgical procedures in index weeks when active SSIs surveillances was conducted. To capture whether the patient developed an SSI, we developed a customized application to analyze clinical notes and code descriptions applying a pattern-matching algorithm based on regular expressions. We described the SSI cases detected by the active surveillance and the text-searching algorithm. To assess the accuracy in identifying the SSIs through the two methods, we adopted a reference standard that calculated the total number of SSIs as those detected by active surveillance plus those derived by the text-searching algorithm that was missed by active surveillance.
Results: Compared with the total number of SSIs used as a reference standard, both methods had a specificity of 100%, a positive predictive value of 100%, and a negative predictive value >99.5%. Sensitivity was 70% for the text-mining algorithm and 60% for the active surveillance. Accuracy was >99% with both methods. The kappa value was 0.46.
Conclusions: Compared with conventional surveillance of SSIs, a text-searching algorithm is a valid tool for case finding that has the potential to reduce drastically the workload of conventional surveillance, which involved direct contact with all families.

Keywords: electronic surveillance; surgical site infections; text mining

Surgical site infections (SSIs) are one of the most frequent healthcare-associated infections (HAI), accounting for around 20% of all HAIs [1]. These infections are associated with longer post-operative in-hospital stays, additional surgical procedures, and a higher mortality rate, representing a negative impact on patient outcomes and healthcare costs [2–4]. The incidence of SSIs differs according to age, type of surgery, use of implants, and surgical incision classification [5–7]. In children, SSI rates within 30 days after surgery range from 2.3% to 6% [6,7].

Surveillance of SSIs is a crucial component of strategies to reduce their incidence, as it enables evaluation of clinical outcomes, monitoring of hospital and surgical team performance, and evaluation of the impact of quality improvement programs on risk reduction. A number of methods have been developed for the surveillance of SSIs [8], including in-patient surveillance, post-discharge follow-up, and automated analysis of administrative or clinical data [9–12].

Active surveillance of SSIs and post-discharge follow-up of patients by infection control teams generally are considered accurate and reliable methods to estimate the incidence of SSIs but are time-consuming and expensive [13]. Identification of SSIs from administrative data, such as diagnosis and procedure codes, requires fewer resources but may be inaccurate, because this depends on the quality and completeness of the electronic administrative data [14]. Electronic surveillance using clinical

[1]Clinical Epidemiology Unit and [3]Medical Direction, Bambino Gesù Children's Hospital, Rome, Italy.
[2]National Research Council, Institute for Research on Population and Social Policies, Rome, Italy.

and administrative data from multiple sources (i.e., surgical, laboratory, radiology, and medication records, physicians' and nursing notes, and diagnosis at discharge) has been reported as a tool for surveillance of SSIs, but experiences are limited [15–17]. The aim of the present study was to assess the accuracy of a text-searching algorithm to detect SSIs in children based on the application of regular expressions of unstructured clinical notes collected through different information systems.

Patients and Methods

We conducted this study at the Ospedale Pediatrico Bambino Gesù (OPBG), an academic tertiary-care children's hospital located in Rome, Italy, with 607 in-patient beds and approximately 24,000 annual admissions. The study was approved by the Hospital Ethics Committee; there was no intervention on the participants.

At the time the study was performed, the Hospital Infection Control Committee (ICC) conducted active surveillance for SSIs on children aged 0 to 17 years undergoing surgical procedures during one index week per quarter [9]. During the index weeks, the list of patients undergoing surgical procedures was obtained from the operating list orders. These children were followed up by trained nurses to identify SSIs up to 30 days from surgery; to collect information on SSIs, data were drawn from clinical records if patients were still hospitalized or through a telephone interview with parents if they had been discharged home. The SSIs were classified according to the case definitions of the Centers for Disease Control and Prevention, Atlanta, GA USA [18] (Table 1). Infections present at the time of surgery were excluded.

For the current study, we developed an information system data warehouse that integrates data provided by electronic health and administrative records for patients who underwent surgical procedures in the SSI active surveillance index weeks from the first quarter of 2016 to the fourth quarter of 2018 (12 index weeks). The electronic sources of data included surgical intervention records, hospital discharge notes, emergency admission records, out-patient visits, and laboratory findings (Table 2). The data warehouse was developed in collaboration with the hospital information technology team and the data management office.

As shown in Fig. 1, this set of records constituted the data sources layer of the information system architecture. To select the data collected in these systems and store them in the integrated database, an extract, transform, and load tool (ETL tool in Fig. 1) was implemented, adopting the IBM Cognos Business Intelligence suite. In particular, this tool accomplished this task through the following two steps: (1) it extracted the list of surgical interventions performed during the SSI active surveillance index weeks; and (2) it detected from the source systems all the events related to the patient that occurred within 30 days from the surgical intervention and loaded them into the integrated database.

To capture whether the patient suffered an SSI, we developed a customized application (analysis tool in Fig. 1) to analyze clinical notes and code descriptions applying a pattern-matching algorithm based on regular expressions (Fig. 2). The application was implemented adopting the Qt framework (https://www.qt.io), a free and open-source widget toolkit that supports the creation of graphic user interface applications. It also provides an embedded regular expression engine that enabled us to implement the algorithm presented in this paper easily to capture an SSI within the explored narrative texts. In particular, the algorithm was based on a set of search items that suggest the presence of an infection within 30 days after surgery. The first group of terms aims to capture whether an infection had been detected during the child's examination (Group 1.1 in Fig. 2), whereas the second one verifies whether this infection involved the surgical site (Group 1.2 in Fig. 2). The adopted algorithm first verifies if,

Table 1. Surgical Site Infection (SSI) Case Definitions (Adapted from Reference 19)

Superficial SSI	Infection involves only skin and subcutaneous tissue and patient has at least one of the following: a. Purulent drainage from the superficial incision. b. Organisms isolated from an aseptically obtained culture of fluid or tissue from the superficial incision. c. At least one of the following signs or symptoms of infection: Pain or tenderness, localized swelling, redness, or heat, and superficial incision that is deliberately opened by surgeon and is culture positive or not cultured. d. Diagnosis of superficial incisional SSI by the surgeon or attending physician.
Deep SSI	Infection involves deep soft tissues (e.g., fascial and muscle layers) and patient has at least one of the following: a. Purulent drainage from the deep incision but not from the organ/space component of the surgical site. b. A deep incision spontaneously dehisces or is deliberately opened by a surgeon and is culture-positive or not cultured when the patient has at least one of the following signs or symptoms: Fever or localized pain or tenderness. c. An abscess or other evidence of infection involving the deep incision is found during direct examination, during re-operation, or by histopathologic or radiologic examination. d. Diagnosis of a deep incisional SSI by a surgeon or attending physician.
Organ/space SSI	Infection involves any part of the body, excluding the skin incision, fascia, or muscle layers, that is opened or manipulated during the operative procedure and patient has at least one of the following: a. Purulent drainage from a drain placed through a stab wound into the organ/space b. Organisms isolated from an aseptically obtained culture of fluid or tissue in the organ/space. c. An abscess or other evidence of infection involving the organ/space that is found during direct examination, during re-operation, or by histopathologic or radiologic examination. d. Diagnosis of an organ/space SSI by a surgeon or attending physician.

Table 2. Source Databases of Information System Data Warehouse and Variables Extracted and Adopted within Text-Mining Algorithm

Source database	*Variables*
Surgical intervention records	• Date of the intervention • ICD-9 CM codes of procedure and disease • Text description of procedure
Hospital discharge notes	• Date of hospital admission and hospital discharge • ICD-9 CM codes of diagnosis at discharge • Text description of patient's physical examination during hospitalization • Text description of patient's clinical summary at discharge
Emergency admission records	• Date of emergency department (ED) admission • ICD-9 CM codes of the diagnosis at ED discharge • Text description of patient's clinical history and physical examination at ED admission • Text description of patient's clinical summary at ED discharge
Outpatient visits	• Date of outpatient visit • ICD-9 CM codes of diagnosis • Text description of patient's clinical history, physical examination, and diagnosis
Laboratory records	• Date of site culture • Result of culture

ICD = International Classification of Diseases.

within the same sentence, there is a term of one group followed by a term of the other group. If a sentence matches this regular expression, the algorithm verifies if this positive result is a false positive that could be generated by the presence of a negative term (Group 2 in Fig. 2). An isolate from a site culture also was associated with a possible SSI. The terms were selected according to the literature (15–17) by the ICC team, which consists of the hospital Medical Director as chairman, infectious disease physicians, infection control nurses, physicians who are experts in infection control, microbiologists, pharmacists, chief surgeons, and physicians and nurses of clinical departments.

All possible SSI cases detected by the algorithm were investigated by an ICC physician expert in infection control, who reviewed the clinical records of the patients. Possible SSI cases were defined as confirmed if they met the criteria reported in Table 1.

In this paper, the feasibility of the information system and the text-mining algorithm have been tested considering the surgical procedures carried out during 12 SSI active surveillance index weeks from the first quarter of 2016 to the fourth quarter of 2018.

We described the SSI cases detected by the active surveillance and the text-searching algorithm. To assess the accuracy in identifying the SSIs through the two methods, we adopted a reference standard that calculated the total number of SSIs as those detected by active surveillance plus those derived by the text-searching algorithm that were missed by active surveillance. We then calculated sensitivity, specificity, the positive predictive value, the negative predictive value, accuracy, and the F score of the two surveillance methods. We also calculated Cohen's kappa coefficient to capture the level of agreement of the two methods.

Results

A total of 3,434 surgical procedures were performed during the study period and included in the analysis. Through active surveillance, 2,944 (85.7%) procedures were traced, whereas the remaining patients were lost to follow-up. Among patients who were followed actively, there were 18 SSIs, 15 (83.3%) of which appeared after discharge.

The text-searching algorithm explored the clinical notes of the 3,434 patients undergoing surgical procedures; of these, 1,100 (32.0%) had at least one hospital follow-up visit within

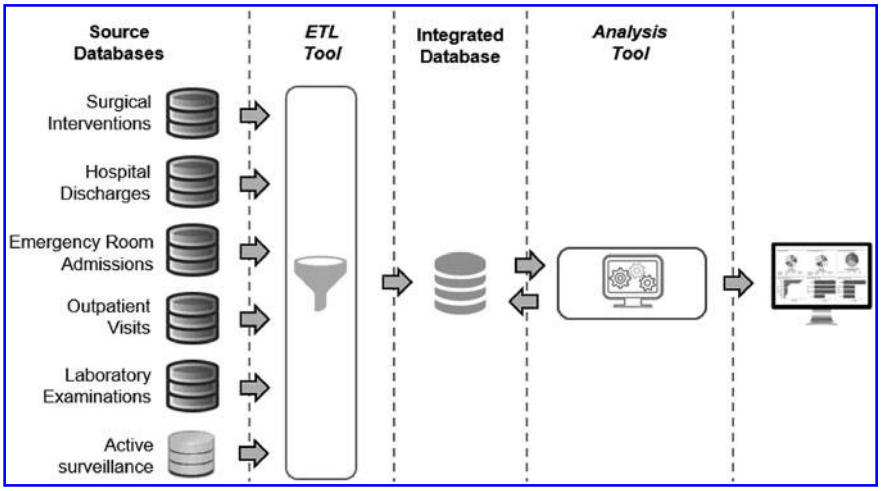

FIG. 1. Description of information system high-level architecture underlining its components: Source databases where data were collected; the ETL tool that transforms these data and stores them in the integrated database (data warehouse); the customized application developed to analyze the integrated data, which provides an interface to the user with results of text-mining algorithm.

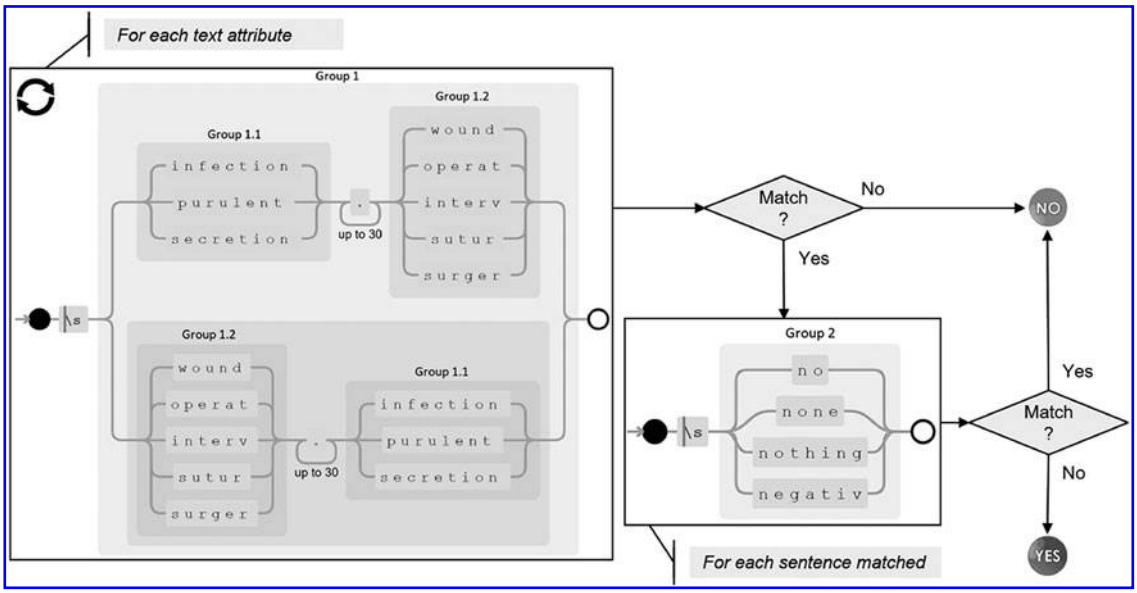

FIG. 2. Description of algorithm based on standard expressions to analyze clinical notes and code descriptions. Group 1 includes list of terms to capture whether an infection had been detected during child's examination (Group 1.1) and if this infection involved the surgical site (Group 1.2). Group 2 includes list of terms to verify if results of Group 1 could be generated by presence of a negative term.

30 days after surgery, and 118 (3.5%) were still hospitalized at 30 days. The algorithm identified 48 possible SSIs, 9 (18.7%) of which were in the active surveillance group, and an additional 12 (25.0%) were confirmed after chart review. The total of the confirmed cases was 21 (43.7%), of which 47.6% occurred after hospital discharge.

The two methods together identified a total of 30 SSIs. Among the 12 SSI cases that were missed by active surveillance, 9 (75%) occurred in children who were lost to follow up, and 3 (25%) were in children whose families were interviewed by telephone and did not report signs or symptoms of SSI. All of the nine SSI cases that were missed by the text-searching algorithm involved children who were discharged home and did not return to the hospital for a follow-up visit within 30 days of surgery.

Table 3 describes the performance of active surveillance and the text-searching algorithm in detecting SSIs. Compared with the total number of SSIs used as a reference standard, both methods had a specificity of 100%, a positive predictive value of 100%, and a negative predictive value >99.5%. Sensitivity was 70% for the text-mining algorithm and 60% for the active surveillance. Accuracy was >99% with both methods. The kappa value was 0.46.

Discussion

In this study, we compared a conventional surveillance of SSIs based on active follow-up of children within 30 days of surgery with electronic surveillance based on a text-searching algorithm. This electronic surveillance had a sensitivity of 70% and a positive predictive value of 100%, documenting its reliability. Conventional surveillance methods of SSI based on patient follow-up and manual chart review are particularly time consuming; for this reason, they may be limited to selected types of surgery or to subsamples of patients undergoing surgical procedures [7]. Electronic surveillance tools based on administrative and clinical data can improve the detection of SSI cases, and previous studies have reported sensitivities ranging from 60.0% to 97.8% [20]. Site culture results, hospital re-admissions, and out-patient visits were considered by other authors as data required for electronic surveillance of SSI [21]. When we conducted this study, we did not have in-place complete electronic medical records; therefore, the text-searching algorithm explored only the electronic components of the medical records that were in place, which included hospital discharge notes, emergency visits, hospital admissions, surgical summaries, out-patient visits, and incision culture results. Mining of these electronic sources of information identified possible SSIs that were confirmed if cases were reported by active surveillance, or, if not, were confirmed through patient chart review by a physician expert in infection surveillance and control. Overall, the text-mining algorithm identified as possible cases 1.4% of the 3,434 patients who underwent surgical procedures; therefore, only a small proportion of possible cases were verified by the infection control team. Compared with conventional surveillance of SSIs, this semi-automated surveillance clearly has the potential of drastically reducing the workload of conventional surveillance, which implied active contact with all families. As expected, most of the SSIs appeared after hospital discharge [9]. Cases missed by the text-searching algorithm were patients who did not return to the hospital within 30 days after surgery, confirming that the algorithm was valid in detecting possible SSI cases seen by the hospital staff. In our study, cases missed by the algorithm in patients who did not return to the hospital within 30 days after surgery were identified through active surveillance with telephone calls to families and the administration of a structured questionnaire on SSI onset. In settings where follow-up visits are not scheduled routinely after surgical procedures, or where the rate of attendance at these visits is sub-optimal, a combination of surveillance methods should be used to monitor SSI occurrence.

TABLE 3. SURVEILLANCE PERFORMANCE OF ACTIVE SURVEILLANCE AND TEXT-MINING ALGORITHM

	SSI (n)	*No SSI (n)*	*Sensitivity (%; 95% CI)*	*Specificity (%; 95% CI)*	*Positive predictive value (%; 95% CI)*	*Negative predictive value (%; 95% CI)*	*Accuracy (%)*	*F Score (%)*	*Kappa value*
Text mining	21	3.413	70.0 (68.3–71.7)	100 (100–100)	100 (100–100)	99.7 (99.4–99.9)	99.7 (99.4–99.9)	82.4 (81.1–83.7)	0.46
Active surveillance	18	2.926	60.0 (58.3–61.6)	100 (100–100)	100 (100–100)	99.6 (99.3–99.8)	99.6 (99.3–99.8)	75.0 (73.4–76.6)	
Total cases	30	3.404	–	–	–	–	–	–	–

CI = confidence interval.

A further advantage of this method is the completeness of the patients included in the denominator. In fact, the active follow-up surveillance based on telephone interviews of parents failed to trace 14% of the families. Cases not detected by this conventional surveillance method included SSIs that occurred in children whose parents did not respond to telephone interviews or were not able to report SSIs signs or symptoms. Because of the complexity in defining SSIs and the need to have complete information on the occurrence of these HAI infections, semi-automated electronic surveillance can aid in case finding.

This study has some limitations to be considered. It represents the experience of a single institution, and it was not possible to compare our results with those of other hospitals. To reduce inter-observer variability, chart review of possible cases was performed by only one physician expert in infection control; however, the implementation of the tool as a routine surveillance method will involve a larger number of trained infection control practitioners. We limited our analysis to a follow-up period of 30 days after surgery, which is the surveillance time frame recommended for procedures without implants [2]. Further analysis should be conducted to investigate the reliability of electronic surveillance based on a text-searching algorithm over longer follow-up periods. Validation of the tool also should be conducted and could consist of checking of the actual data with an independent data source, such as an infection control practitioner or infectious diseases physician/microbiologist who has not been involved in the initial data collection and determination, who independently examines a random sample of cases to determine their classification and subsequently cross-check against existing data.

Surveillance of SSIs is important to target quality improvement interventions, monitor temporal trends, identify alert signals, and perform external benchmarking [5–7]. The main objective of this study was to assess the validity of an electronic tool for SSI surveillance to be adopted at the hospital level. In fact, previous studies have shown that the validity of automated or semi-automated surveillance of SSI must be investigated carefully, because cases may be misclassified by administrative or clinical data [14–16]. The results we obtained enabled us to implement the text-searching algorithm as a tool for case finding applied to all surgical procedures. The availability of timely and complete data on SSI incidence enables evaluation of the quality of surgical care in all pediatric specialties. Moreover, the analysis of narrative texts, which commonly are collected in different hospital information systems, has the advantage that the proposed method can be replicated easily in different settings. Text-searching algorithms may be developed for the surveillance of other quality measures, including HAI, such as central-line-associated blood stream infections. A further development of this algorithm will be to consider risk factor stratification of SSI rates regarding the ''factors associated with SSI development, such as ''category of operative procedure,'' ''duration of operative procedure,'' and ''incision class.'' The collection of data on a larger number of cases may enable design and validation of predictive scores of SSIs in children. This is a promising field to enhance the collection of outcome indicators and their use to improve the quality of care of children undergoing surgery.

Funding Information

No financial support was received.

Author Disclosure Agreement

No competing financial interests exist.

References

1. European Centre for Disease Prevention and Control. Point prevalence survey of healthcare associated infections and antimicrobial use in European acute care hospitals. Stockholm: ECDC; 2013.
2. European Centre for Disease Prevention and Control. Surveillance of surgical site infections and prevention indicators in European hospitals – HAISSI protocol. *Available at:* http://ecdc.europa.eu/sites/portal/files/documents/HAI-Net-SSI-protocol-v2.2.pdf
3. Badia, JM, Casey AL, Petrosillo N, et al. Impact of surgical site infection on healthcare costs and patient outcomes: A systematic review in six European countries. J Hospital Infect 2017;96: 1–15.
4. Shepard J, Ward W, Milstone A, et al. Financial impact of surgical site infections on hospitals: The hospital management perspective. JAMA Surg 2013;148:907–914.
5. Horwitz JR, Chwals WJ, Doski JJ, et al. Pediatric wound infections: A prospective multicenter study. Ann Surg 1998; 227:553–558.
6. Uludag O, Rieu P, Niessen M, et al. Incidence of surgical site infections in pediatric patients: A 3-month prospective study in an academic pediatric surgical unit. Pediatr Surg Int 2000;16:417–420.
7. Bucher BT, Guth RM, Elward AM, et al. Risk factors and outcomes of surgical site infection in children. J Am Coll Surg 2011;212:1033–1038.
8. Smyth ET, Emmerson AM. Surgical site infection surveillance. J Hosp Infect 2000;45:173–184.
9. Keller S, Grass F, Tschan F, et al. Comparison of surveillance of surgical site infections by a national surveillance program and by institutional audit. Surg Infect 2019; Published ahead of print. doi:10.1089/sur.2018.211.
10. Ciofi degli Atti ML, Serino L, Piga S, et al. Incidence of surgical site infections in children: Active surveillance in an Italian academic children's hospital. Ann Ig 2017;29: 46–53.
11. Brossette SE, Hacek DM, Gavin PJ, et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: The future of infection control surveillance? Am J Clin Pathol 2006;125:34–39.
12. Ju MH, Ko CY, Hall BL, et al. A comparison of 2 surgical site infection monitoring systems. JAMA Surg 2015;150: 51–57.
13. Protocol for Surgical Site Infection Surveillance with a Focus on Settings with Limited Resources. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO *Available at:* https://www.who.int/infection-prevention/tools/surgical/SSI-surveillance-protocol.pdf
14. van Mourik MS, van Duijn PJ, Moons KG, et al. Accuracy of administrative data for surveillance of healthcare-associated infections: A systematic review. BMJ Open 2015; 27;5: e008424.
15. Michelson JD, Pariseau JS, Paganelli WC. Assessing surgical site infection risk factors using electronic medical records and text mining. Am J Infect Control 2014;42:333–336.
16. Cato KD, Liu J, Cohen B, et al. Electronic surveillance of surgical site infections. Surg Infect 2017;18:498–502.
17. Kulaylat AN. Measuring surgical site infections in children: Comparing clinical, electronic, and administrative data. J Am Coll Surg 2016;222:823–830.
18. Sherman ER, Heydon KH, St John KH, et al. Administrative data fail to accurately identify cases of healthcare-associated infection. Infect Control Hosp Epidemiol 2006; 27:332–337.
19. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control 2008;36:309–332.
20. Freeman R, Moore LS, García Álvarez L, et al. Advances in electronic surveillance for healthcare-associated infections in the 21st Century: A systematic review. J Hosp Infect 2013;84:106–119.
21. Woeltje KF, Lin MY, Klompas M, et al. Data requirements for electronic surveillance of healthcare-associated infections. Infect Control Hosp Epidemiol 2014;35:1083–1091.

Address correspondence to:
Dr. Marta Luisa Ciofi degli Atti
Clinical Epidemiology Unit
Bambino Gesù Children's Hospital
Rome
Italy

E-mail: marta.ciofidegliatti@opbg.net

This article has been cited by:

1. Kushan De Silva, Noel Mathews, Helena Teede, Andrew Forbes, Daniel Jönsson, Ryan T. Demmer, Joanne Enticott. 2021. Clinical notes as prognostic markers of mortality associated with diabetes mellitus following critical care: A retrospective cohort analysis using machine learning and unstructured big data. *Computers in Biology and Medicine* **132**, 104305. [Crossref]
2. A. Egli, J. Schrenzel, G. Greub. 2020. Digital microbiology. *Clinical Microbiology and Infection* **26**:10, 1324-1331. [Crossref]