We propose an integrated semantic web framework consisting of formal ontologies, web services, a reasoner and a rule engine that together recommend appropriate level of patient-care based on the defined semantic rules and guidelines. The classification of healthcare-associated infections within the HAIKU (Hospital Acquired Infections – Knowledge in Use) framework enables hospitals to consistently follow the standards along with their routine clinical practice and diagnosis coding to improve quality of care and patient safety. The HAI ontology (HAIO) groups over thousands of codes into a consistent hierarchy of concepts, along with relationships and axioms to capture knowledge on hospital-associated infections and complications with focus on the big four types, surgical site infections (SSIs), catheter-associated urinary tract infection (CAUTI); hospital-acquired pneumonia, and bloodstream infection. By employing statistical inferencing in our study we use a set of heuristics to define the rule axioms to improve the SSI case detection. We also demonstrate how the occurrence of an SSI is identified using semantic e-triggers. The e-triggers will be used to improve our risk assessment of post-operative surgical site infections (SSIs) for patients undergoing certain type of surgeries (e.g., coronary artery bypass graft surgery (CABG)).

**Keywords**  Ontologies · Knowledge modeling · Healthcare-associated infections · Surveillance · Semantic framework · Surgical site infections

**Introduction**

Healthcare-associated Infections (HAIs) affect millions of patients around the world, killing hundreds of thousands and imposing, directly or indirectly, a significant socio-economic burden on healthcare systems [1]. According to the Centers for Disease Control (CDC) [2], hospital-acquired infections in the U.S., where the point prevalence of HAIs among hospitalized patients is 4 %, result in an estimated 1.7 million infections, which lead to as many as 99,000 deaths and cost up to $45 billion annually [3, 4]. Similar or higher rates of HAI occur in other countries as well with an estimated 10.5 % of patients in Canadian hospitals having an HAI [5]. Clinical assessment and laboratory testing are generally used to detect and confirm an infection, identify its origin, and determine appropriate infection control methods to stop the infection from spreading within a healthcare institution. Failure to monitor, and detect HAI in timely manner can delay diagnosis, leading to complications (e.g., sepsis), and allowing an epidemic to spread.

To ensure the quality of care given to the patients in healthcare settings, it is crucial to have systems that monitor for cases of HAI [6]. Our knowledge-based surveillance infrastructure enables monitoring for HAIs and
generates an alert when a suspect, probable, or confirmed cases of HAI is detected. In this paper we focus on surgical site infections (SSIs), one of the most common healthcare associated infections, accounting for about 31% of all HAIs among hospitalized patients in 2010 in U.S. [7]. Diagnosis of an SSIs relies mainly on direct observation of physical signs and symptoms of infection in an incisional wound and a case cannot usually be confirmed solely by analyzing data given in laboratory reports. Given the diversity, complexity and heterogeneity of HAI data, availability of a reference vocabulary is a prerequisite of creating an integrated knowledge-based system. Despite several modifications and improvements to existing terminologies made by the Centers for Disease Control and Prevention (CDC) in the last decade, e.g., specifying the location of infections related to surgical operations and clarifying the criteria to identify the exact anatomic location of deep infections [8], inconsistencies, discrepancies, and confusion in the application of the criteria in different medical/clinical practices still exist, and there is a need for further improvement and clarification of the current nomenclature [9].

While the Centers for Disease Control and Prevention (CDC) has provided a certain criteria as a guideline [8] to prevent, control and reduce HAIs, in the HAI ontology project [10] we have brought together expertise in artificial intelligence, knowledge modeling, epidemiology, medicine, and infection control to explore how advances in semantic technology can improve the analysis and detection of HAIs. To develop a common understanding about the domain of infection control and to achieve data interoperability in the area of healthcare-associated infections, we present the HAI Ontology as part of the HAIKU (Hospital Acquired Infections – Knowledge in Use) project. The formal HAI ontology assists researchers and health professionals in analyzing medical records to identify and flag potential cases of HAIs among patients who could be at risk of acquiring an SSI.

In this paper we discuss the role and importance of the HAIKU semantic infrastructure to improve the detection of HAI using semantic web technologies. The paper is organized as follows: "Existing methods for detecting HAI" section presents an overview of existing tools and systems for managing nosocomial infections. The HAIKU ontology design and implementation along with the related semantic rules and axioms designed for intelligent alerting are presented in "The HAI ontology: an overview" and "The HAIKU framework for case detection and reporting" section, respectively. "Axiomatization using semantic and statistical analysis" section presents our axiomatization process informed by statistical analysis of existing datasets. The paper concludes in "Conclusion" section with a general discussion, a summary of findings, and anticipated future work.

### Existing methods for detecting HAI

Healthcare-associated Infections have been considered an important healthcare quality outcome since Florence Nightingale reduced mortality rates through the application of septic techniques in field hospitals during the Crimean war [11]. HAIs continue to be costly to individual patients and to the health system. Although there are several different types of HAIs, five of them account for nearly all cases. These HAI types are: pneumonia, surgical site infections, urinary tract infections, bloodstream infections, and gastrointestinal infections [3, 5]. The recognition that specific syndromes represent the majority of infections was an important advancement in efforts to reduce the incidence and impact of HAIs. While general approaches to reduce infections have been employed since the 1800s – including encouraging hand hygiene [12, 13] and environment cleaning [14, 15] – evidence-based preventive measures specifically designed for each of the five HAI syndromes now exist [16–20].

A cornerstone of HAI prevention and control is disease surveillance. The Centers for Disease Control and Prevention has specified explicit criteria and cohort definitions to support the surveillance of various HAI syndromes [6]. Their efforts in this domain began in the 1970s and led to the conduct of the SENIC Project [6], which evaluated the impact of infection surveillance on HAI incidence [21]. This study demonstrated that systematic tracking of HAIs coupled with physician-level feedback significantly reduces infection risk [21]. Other researchers [22, 23] have also described the use of electronic systems for the surveillance of hospital acquired infections, mainly through monitoring microbiology lab reports. As a result of the SENIC Project, hospital based infection control programs have become a standard practice; and surveillance is a primary function of these programs. The task of surveillance, however, is not trivial. It is instructive to consider surveillance for surgical site infections as an example. Each day, patients undergoing surgery must be identified, baseline information recorded, and a method of follow-up identified. Then, practitioners must follow patients for 30 days following the surgery to identify specific criteria indicative of infection [24]. This monitoring requires extensive review of medical records and possibly a telephone interview with the patient. This manual process is time consuming and is expensive, requiring highly skilled personnel. Due to the expense, hospitals may forego surveillance or focus only on a subset of patients. Neither of these alternatives is optimal and in spite of many years of experience and research, the detection and control of HAIs remains as a challenge.

However, many of the steps in the surveillance of HAI could, in theory, be automated. The cohort identification could be simplified by taking advantage of information contained in information systems used to manage operating rooms. Most of the criteria specifying increased risk of infection are contained
in other systems, such as laboratory, pharmacy, or administrative information systems. If this information could be combined in a consistent manner across disparate information systems, then it might be possible to reduce the costs of infection control programs or to increase the number of patients covered by them. The goal of our project is to create a logical framework using clear semantics to enable the consistent integration of different data sources necessary for HAI surveillance, with an initial focus on SSI.

The HAI ontology: An overview

Many factors [25], including environmental, organizational, procedural, and personal factors, contribute to the occurrence and severity of HAIs. The effectiveness of any detection method is highly dependent on the quality of the integrated information derived from different data sources in various settings including microbiological, clinical, and epidemiological data. We use ontologies along with other semantic technologies to align these different data sources with one another and with knowledge-bases, regulations, and processes. An ontology, or a formal explicit specification of shared conceptualization [26], provides a semantic framework for knowledge dissemination, exchange, and discovery via reasoning and inferencing. Ontologies capture the knowledge in a domain of interest through concepts, instances and relationships (taxonomic and associative). The taxonomic relationships organize concepts into sub/super (narrower/broader) concept tree structure, while associative relationships relate instances of defined concepts across taxonomies.

Methodology and data sources

The HAI Ontology has been implemented following an integrated and iterative V-model [27] methodology consisting of the following steps: i) scope definition; ii) data and knowledge acquisition; iii) conceptualization through defining the main concepts, their attributes and the relationships within the domain of interest; iv) integration; v) encoding using a formal ontology language; vi) documentation and vii) evaluation. In the conceptualization stage we have defined the ontological elements (concepts, relationships/attributes and logical axioms) based on expert interview, the results from a statistical analysis over our available datasets along with the CDC SSI case definition criteria [6, 24] and a case–control study performed over a systematic review of 156 studies on SSIs and their associated factors and attributes [28]. In this study, we used the Ottawa Hospital (TOH) research data warehouse, [28] which contains data from 1996 to the present time. Warehouse is a relational database that draws together data from multiple source systems including the most important operational information systems, such as laboratory (microbiology and clinical chemistry), pharmacy, operating room, clinical notes, encounters and diagnoses (classified using ICD codes) patient registration system, patient demographics and movement, and patient abstracts (e.g., discharge abstract database (DAD), and the national ambulatory care reporting system (NACRS)).

Moreover to develop and validate detection methods we have used clinical data along with the information extracted through exhaustive chart review to identify patients that have experienced surgical site infections. For surveillance of SSIs, the infection-control staff routinely collect demographic and operational data about selected patients undergoing one or more operative procedures during a specific observation time period.

At the integration phase several databases (e.g., those containing information on hospital morbidity and discharge abstracts), existing bio-ontologies (e.g., SNOMED-CT [29], ICD-9,1 HL7-RIM,2 FMA,3 CheBI,4 Infectious Disease Ontology (IDO)), and textual resources have been used to design and implement the integrated HAI Ontology. The databases and ontologies has been identified and selected based on our requirements and their compatibility with existing data warehouse schema architecture. The integration [10] has been done at two structural and semantic levels. The structural integration has been done by creating a homogeneous dataset in a standard ontology language. Semantic integration, which is more challenging, has been performed partly manually and partly (semi) automatically through using the SemanticScience Integrated Ontology (SIO) framework [30].

Ontological conceptualization

We use description logics and OWL 2.0 Web Ontology Language [31] to encode the ontological model. Figures 1 and 2 demonstrate, respectively, a segment of the HAI Ontology class hierarchy representing different types of hospital acquired infections and different types of processes and operations defined in the HAI Ontology, integrated within the SemanticScience Integrated Ontology (SIO) framework [30], created using Protégé5 ontology editor.

Figure 3 demonstrates a partial view of the HAI Ontology in Protégé with axiom definitions for Surgical Site Infections.

An extra-simple-time-ontology has been created to manage temporal aspects of HAIKU. We also use the SADI

1 International Classification of Diseases, Version 9
2 HL7 Reference Information Model: www.hl7.org/implement/standards/rim.cfm
3 The Foundational Model of Anatomy Ontology: sig.biostr.washington.edu/projects/fm/AboutFM.html
4 Chemical Entities of Biological Interest (CheBI): https://www.ebi.ac.uk/chebi/
5 http://protege.stanford.edu/
framework [32], which is a set of conventions for creating HTTP-based Semantic Web services. It consume RDF\(^6\) documents as input and produce RDF documents as output, which solves the syntactic interoperability problem as all services communicate in one language. We evaluate the HAI Ontology by assessing its competency to answer the intended queries. Also to examine consistency of the ontology we use logical reasoners. We have already formulated [10, 33] a broad range of semantic queries to improve HAI case identification and enumeration, evaluation of HAI risk factors or causative factors, identification or evaluation of diagnostic factors.

Below is an example [33] of captured knowledge in the HAI ontology represented in in N3 format [34], which is a variation of RDF with improved human-readability. It demonstrates an event such as “Diagnosis” “is performed for” a “patient” within an specific time “diagnosis time” and “identifies” an “incident”, which could be a “surgical site infection” as “consequence of” a “surgery”, in this case “coronary artery bypass graft” has been performed in a specific time (“event has time”) “surgery time”. Looking at the “blood culture result” for “blood culture test” that “has time” “test time”, “specifies finding”, which here is “positive blood culture finding” that “identifies microorganism”

---

\(^6\) The Resource Description Framework (RDF): www.w3.org/RDF/
"Serratia Proteamaculans" and its NCBI’s taxonomy number: 28151. Then, a “pharmacy service”, which is “a service for” the “patient” has been performed at pharmacy service time to “manage” a “drug product”, in this case “A-Hydrocort Inj” with a specific drug identification number (DIN) record.

```sparql
1 @prefix haio:<http://www.semanticweb.org/ontologies/2010/7/HAI.owl#> .
2 @prefix esto:<http://unbsj.biordf.net/ontologies/extra-simple-time-ontology.owl#> .
3 @prefix lern:<http://purl.oclc.org/SADI/LSRN/> .
4 @prefix sio:<http://semanticscience.org/resource/> .
5 @prefix xsd:<http://www.w3.org/2001/XMLSchema#> .
6 @prefix taxon:<http://purl.org/obo/owl/NCBITaxon#NCBITaxon_> .
7 :patient a haio:Patient .
8 :diagnosis a haio:Diagnosis; haio:is_performed_for :patient;
   haio:event_has_time :diagnosis_time; haio:identifies :incident .
9 :incident a haio:Surgical_site_infection; haio:is_consequence_of :surgery .
10 :surgery a haio:Coronary_artery_bypass_graft; haio:event_has_time :surgery_time .
11 :test a haio:Blood_culture;
   haio:has_result :blood_culture_result; haio:event_has_time :test_time .
12 :blood_culture_result a haio:Blood_culture_result;
   haio:specifies_finding :finding .
13 :finding a haio:Positive_blood_culture_finding;
14 # Serratia proteamaculans
15 haio:identifies_microorganism taxon:28151 .
16 :pharmacy_service a haio:Pharmacy_service; haio:is_service_for :patient ;
   haio:manages :drug_product; haio:event_has_time :pharmacy_service_time .
17 # 'is subject of'
19 # 'has attribute'
20 :din_record a lsrn:DIN_Record ; sio:SIO_000008 :din_attr .
21 # 'has value' "A-HYDROCORT INJ"
```

The HAIKU framework for case detection and reporting

As shown in Fig. 4, the HAIKU semantic web framework consists of formal ontologies, web services, a reasoner and a rule engine that together recommend appropriate level of actions based on the defined semantic rules and guidelines. The HAIKU semantic rules are modeled throughout a set of iterative processes that consists of domain and context specification, consensus knowledge acquisition from unstructured texts and literature, statistical/epidemiological analysis over existing structured data available from the TOH data warehouse, interviewing with domain experts and end-users, evaluation and conflicts resolution.

The semantic backbone, powered by the HAI ontology, assists us in reviewing medical records by identifying specific
Fig. 3 Part of the HAI Ontology in Protégé representing axiom definitions for SSIs

Fig. 4 The HAIKU framework for automatic case detection
terms and the association between them to generate patterns that indicate at-risk patients. By linking relevant pieces of data and information (e.g., signs and symptoms, type of medical procedures, length of hospitalization, drugs prescribed, names of infectious agents), an e-trigger can be fired when the semantics imply a possible risk of SSI, allowing preventative measures can be taken.

In the knowledge engineering phase, as demonstrated in Fig. 4, after implementing the integrated HAI ontology we used PSOA RuleML, to map the concepts in the ontology to instances of data in TOH datawarehouse. For example, the population of haio:is_performed for and haio:identifies is captured by the following rules:

```
1  And (  
2       ?diagnosis # haio:Diagnosis  
3       haio:is_performed_for(?diagnosis ?patient)  
4       haio:identifies(?diagnosis ?disease)  
5       ?disease # haio:Disease  
6       ?disease # ?diseaseClass )  
7       :-  
8  And (  
10      ?diagnosisRow # dwt:NhrDiagnosis(dwa:hdgWID->?diagnosisID  
11      dwa:hdgHraEncWID->?encounterID  
12      dwa:hdgCd->?diseaseCode)  
13      ?patient = External(modf:Patient_by_patWID(?patientID))  
14      ?diagnosis = External(modf:Diagnosis_by_hdgWID(?diagnosisID))  
15      ?diseaseClass = External(modf:disease_class_by_ICD10(?diseaseCode))  
16      ?incident = External(modf:Disease_by_diagnosis(?diagnosisID))  
```

**Axiomatization using semantic and statistical analysis**

We have defined a set of rule axioms to trigger specific actions under certain conditions. By using the ontology and a logical reasoner together the system can issue alerts to support infection control actions. We start our semantic analysis by converting the existing knowledge, based on the CDC guideline and our statistical analysis, into rule axioms. We specify the rule axioms through multiple criteria such as type and duration of surgery, patients’ age and specifications, comorbidities/existing conditions, etc.

In our model we analyzed TOH data related to 732 operative episodes among 729 patients (3 patients had 2

Table 1  Logistic regression coefficients in odds ratio indicating the predictive power of the trigger factors

<table>
<thead>
<tr>
<th>Trigger Factors</th>
<th>Odds Ratio</th>
<th>lower 95 % CI</th>
<th>Upper 95 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic antibiotics likely for SSI started on or after post-operative day 2</td>
<td>7.59</td>
<td>4.29</td>
<td>13.50</td>
</tr>
<tr>
<td>Duration of the antibiotics usage above a</td>
<td>1.06</td>
<td>1.03</td>
<td>1.09</td>
</tr>
<tr>
<td>Any CT (Computed Tomography) scan report with mention of specific SSI terms a</td>
<td>1.14</td>
<td>0.34</td>
<td>4.17</td>
</tr>
<tr>
<td>Readmission with presumed diagnosis of SSI a</td>
<td>15.44</td>
<td>6.78</td>
<td>38.18</td>
</tr>
<tr>
<td>Readmission to emergency room/reference center, or coded as urgent a</td>
<td>1.34</td>
<td>0.76</td>
<td>2.32</td>
</tr>
<tr>
<td>Any likely significant pathogens on wound culture a</td>
<td>1.38</td>
<td>0.69</td>
<td>2.70</td>
</tr>
</tbody>
</table>

a For specific conditions, refer to Table 2
b CI - Confidence Interval
episodes each) who received a coronary artery bypass graft (CABG) between July 1 2004 and June 30 2007. To define the indicators and trigger factors, we used existing knowledge in guidelines, biomedical literature, and insights obtained through the statistical analysis of data in TOH data warehouse. Based on the semantic and statistical analysis and also following the guideline presented in [28] we generate the following three rules for issuing alerts for suspect, probable and confirmed cases of SSIs.

**Rule 1. IssueAlertForSuspectCase R001**

**Class names:**
Hospitalized_Patient, surgical_procedure, Suspect_Case, Pharmacy_service, CT_Scan_Thorax, Abscess, CT-diagnosis, PostOperativeDay2, IV_Antibiotic

**Meaning:**
Hospitalized patients having surgical procedure AND ((having Antibiotics use started on or after postoperative day 2) OR (having CT Scan-Thorax report with mention of specific SSI terms (ex. Abscess)))

**Rule 2. IssueAlertForProbableCase R002**

**Class names:** Hospitalized_Patient, surgical_procedure, Probable_Case, blood_culture_Test, Finding, Positive_Blood_culture, Positive_gram_stain

**Meaning:**
Hospitalized patients having surgical procedure AND having blood culture test AND Finding = (Positive_Blood_culture OR Positive_gram_stain (from wounded site))

**Rule 3. IssueAlertForConfirmedCase R003**

**Class names:** Hospitalized_Patient, Disease, Disease_Code, Confirmed_Case,

**Meaning:**
Hospitalized patients having Disease AND Disease code = HAICD code

The semantic alerts are issued upon presentation of one or more indicators for probable, suspect or confirmed cases. The results from the statistical analysis are used to:

1. Power the knowledge acquisition phase by improving (or revising our) conceptualization of the domain
2. Assess the result obtained in response to queries from our knowledge-based system
The triggers most frequently associated with cardiac surgical site infection in the literature have been summarized [28]. We used multivariable logistic regression with data in the TOH data warehouse to generate conditional predicted probabilities of SSI, based on these trigger factors (Table 1) [28].

Area Under the Receiver Operating Characteristic (AUC-ROC) curve: 0.91 (10-fold cross-validated estimate was 0.90) 95 % CI: 0.88-0.94

As mentioned, the case identification rule has been defined in the HAIKU ontology as suspected (culture positive from wound or blood specimen) and probable (IV antibiotics or thoracic CT order). The graphs below display the distribution of predicted probabilities among, probable cases, suspected cases and the data including non-SSI instances (individuals). Figure 5a represents the density of predicted probability among probable cases (lab result indicative of infection). The detailed search terms for this analysis have been shown in Table 2.

Table 2  Search terms for trigger factors used in the model (adapted from [28])

<table>
<thead>
<tr>
<th>Trigger factor definitions</th>
<th>Search terms or codes used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microbiology Reports</strong></td>
<td></td>
</tr>
<tr>
<td>(Free text entry)</td>
<td></td>
</tr>
<tr>
<td>‘Any likely significant pathogens on wound culture’</td>
<td>‘obacter; ’staphylococcus aureus; ’escherichia; ’streptococcus; ’pseudomonas; ’morganella; ’serrata; ’stentrophomonas; ’klebsiella; ’proteus; ’providencia; ’coagulase negative’ (IF multiple gram negative)</td>
</tr>
<tr>
<td>‘Any likely significant pathogens on blood culture’</td>
<td>‘obacter; ’staphylococcus aureus; ’escherichia; ’streptococcus; ’pseudomonas; ’morganella; ’serrata; ’stentrophomonas; ’klebsiella; ’proteus; ’providencia; ’enterococcus; ’haemophilus; ’propionibacterium; ’coagulase negative’ (IF had non-human derived material implanted in index operative episode)</td>
</tr>
<tr>
<td><strong>Radiology Reports</strong></td>
<td></td>
</tr>
<tr>
<td>(Free text entry)</td>
<td></td>
</tr>
<tr>
<td>‘Any CT report with mention of specific SSI terms’</td>
<td>‘osteomyelitis; ’sternal infection; ’wound infection; ’mediastinitis; ’abscess; ’retrosternal fluid; ’endocarditis; ’phlegmon’</td>
</tr>
<tr>
<td><strong>Admitting Diagnoses</strong></td>
<td></td>
</tr>
<tr>
<td>(Free text entry)</td>
<td></td>
</tr>
<tr>
<td>‘Readmission with presumed diagnosis of SSI’</td>
<td>‘endocarditis; ’infect; ’cellulitis; ’incision; ’wound; ’osteomyelitis; ’sepsis; ’I + D; ’abscess; ’mediastinitis; ’debride; ’sternal’</td>
</tr>
<tr>
<td><strong>Pharmacy Records</strong></td>
<td></td>
</tr>
<tr>
<td>(Generic and Trade names, Drug Identification Numbers, Anatomic Therapeutic Chemical Classification codes, and American Hospital Formulary Service pharmacologic-therapeutic classifications for listed agents)</td>
<td></td>
</tr>
<tr>
<td>‘Systemic antibiotics likely for SSI started on or after post-operative day 2’</td>
<td>‘cefaclin; cephalixin; clindamycin; cloxacillin; ertapenem; imipenem; linezolid; meropenem; nafcillin; penicillin; piperacillin; rifampin; rifampicin; ticarcillin; vancomycin’</td>
</tr>
</tbody>
</table>
at Table 2. The peak around predicted probability of 0.1 is probably due to the gram stain test being non-specific for infection.

The distribution represented in Fig. 5b is also based on lab results, but selected from positive culture results only which is expected to be more specific to the presence of pathogen. However, overall accuracy of prediction in terms of AUC-ROC did not change significantly whether we use this criteria or the non-specific one above, potentially due to the small number of individuals that had culture alone (n=44) in our data resulting in the loss of precision in statistical quantity. Distribution of predicted probability among suspected cases (selected systemic antibiotics 2 days after index operation—see the list of antibiotics in Table 2) shown in Fig. 5c suggests that certain antibiotics orders are less predictive of SSI than other as seen in the large peak at low value of predictive probability. Highly right-skewed distribution of predicted probability density is observed among suspected case (thoracic CT result with suggestive terms of SSI) (Fig. 5d) implying the usefulness of the CT scan for the detection of SSI, although the number of the individuals receiving the scan is small (n=38) and thus its statistical significance is inconclusive as seen in its confidence interval crossing the null effect (i.e., 1.0).

The weak predictive power of culture diagnosis seen in our model does not preclude the importance of this test as it is already established indicator of the infection (Ref-CDC); rather, this could be caused by the low yield of culture diagnosis when specimens were drawn after the initiation of antibiotics treatment (ref – I will find it today if necessary) or other factors that is not measured in this study. In contrast, extremely strong effect of the presumed diagnosis of SSI at readmission seen in our model potentially reflects a high accuracy of readmission diagnosis in TOH. Thus, although existing knowledge such as the CDC guideline will play a central role in developing the detection rule, the predictive power of trigger factors may differ from one target population to another due to biological and non-biological variations reflecting local practice, such as the accuracy of laboratory tests, patterns of medication use, and clinical diagnosis. Although the accuracy and precision of statistical algorithm is limited to the quality of data (e.g., sample size and selection bias) and analytic methodology (e.g., model selection), it can be highly useful to quantify the importance of trigger factors for a target population of interest and thus assist to achieve best detection performance at specific context. Therefore, the importance of combining existing knowledge and statistical analysis will be highly critical once the evaluation of the detection systems is extended to other settings, especially when strong heterogeneity across healthcare or target population is expected.

E-triggers can be issued in response to queries retrieving confirmed, suspect and probable SSI cases using logical reasoners and rule engines. The logical reasoner controls the consistency and satisfiability of the results and reveals redundancies and hidden dependencies.

Conclusion

In behavioral economics, nudges are used for positive reinforcement and indirect suggestions to try to achieve non-forced compliance to improve the decision making of groups and individuals. Using the same analogy, we define a semantic infrastructure to issue semantic nudges to assist healthcare professionals and infection control practitioners in their decision-making process to effectively monitor healthcare-associated infections (HAIs), with a particular focus on surgical site infections (SSIs). Since the rate of HAIs is a major quality and performance indicator, healthcare settings are constantly under pressure to control, and minimize the incidence of HAIs.

SSIs impose a huge burden on patients, hospitals, and healthcare organizations. For the CABG procedure, post-surgery SSIs can substantially increase the length-of-stay in hospital [36]. Several known risk factors (e.g., obesity, old age, and diabetes, duration of surgery) have been associated with incidence of SSIs. Multiples lists of cautionary steps, guidelines and strategies have been provided by different infection control agencies to improve the SSIs surveillance and prevention, to reduce the incidents of SSI.

In this study, we presented the HAIKU semantic web platform. This platform captures and integrates knowledge from existing guidelines, standards, and databases and generates triggers to detect cases based on the HAI ontology (available for download from: http://surveillance.mcgill.ca/projects/haiku/HAI.owl), and the logical rules created through semantic and statistical analysis. The HAIKU framework enables health professionals and researchers to analyze retrospective data on health-care associated infections and generate predictive models using the existing knowledge (i.e., literature findings, guidelines, statistical models) to issue alerts on potential cases of HAIs. It also assists them in making informed decisions regarding different procedures and policies about healthcare associated infections. While in this paper we focused more on SSIs, the detection of other adverse events due to infection could possibly benefit from the HAIKU semantic framework as well, such as sepsis, C-difficile associated diarrhea, and urinary tract infection.

One strength of this study is the use of statistical methods along with semantic technology to define e-triggers, which provides a coherent, efficient and consistent view of the data extracted from various resources. Our research demonstrates that a combination of parameters indicating infections, e.g., antibiotic use and positive blood culture, along with the risk factors, e.g., having old age, duration of a surgical procedure
leads to more accurate triggers to issue alerts for early detection of HAIs.

The semantic rules can be used to recommend a certain course of actions given a specific situation occurring in a healthcare setting. Issuing timely alerts and warnings can increase the efficiency of the healthcare system, and improve the quality of care in healthcare settings. We plan to implement the HAIKU framework in multiple clinical sites so that we can evaluate the transferability of our approach. One limitation of our work is related to the temporal reasoning. Since we are using description logics to encode the HAI Ontology, temporal reasoning can easily lead to undecidability. So, we are currently working on alternative approaches to overcome this problem. In addition we are planning on enriching the ontological structure to extend our analysis for different types of HAIs, as well as other adverse events such as sepsis and bleeding.

Acknowledgments The Canadian Foundation for Innovation (CFI), the Canadian Institutes of Health Research (CIHR), the Natural Sciences & Engineering Council of Canada (NSERC), and Canadian Network for the Advancement of Research, Industry and Education (CANARIE) provide funding for this research.

References

24. Surgical Site Infection (SSI) Event, Procedure-Associated Module, Centers for Disease Control and Prevention, Office of Infectious Diseases, National Center for Emerging and Zoonotic Infectious


