Journal of Hospital Infection 110 (2021) 139–147

Available online at www.sciencedirect.com

Journal of Hospital Infection

journal homepage: www.elsevier.com/locate/jhin

The accuracy of fully automated algorithms for surveillance of healthcare-associated urinary tract infections in hospitalized patients

S.D. van der Werff [a,*], E. Thiman [a,b], H. Tanushi [a,c], J.K. Valik [a,b], A. Henriksson [d], M. Ul Alam [d], H. Dalianis [d], A. Ternhag [a,b], P. Nauclér [a,b]

[a] *Department of Medicine Solna, Division of Infectious Disease, Karolinska Institutet, Stockholm, Sweden*
[b] *Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden*
[c] *Department of Data Processing & Analysis, Karolinska University Hospital, Stockholm, Sweden*
[d] *Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden*

A R T I C L E I N F O

Article history:
Received 17 December 2020
Accepted 27 January 2021
Available online 3 February 2021

Keywords:
Automated surveillance
Algorithms
Healthcare-associated infection
Natural language processing
Urinary tract infections

S U M M A R Y

Background: Surveillance for healthcare-associated infections such as healthcare-associated urinary tract infections (HA-UTI) is important for directing resources and evaluating interventions. However, traditional surveillance methods are resource-intensive and subject to bias.

Aim: To develop and validate a fully automated surveillance algorithm for HA-UTI using electronic health record (EHR) data.

Methods: Five algorithms were developed using EHR data from 2979 admissions at Karolinska University Hospital from 2010 to 2011: (1) positive urine culture (UCx); (2) positive UCx + UTI codes (International Statistical Classification of Diseases and Related Health Problems, 10th revision); (3) positive UCx + UTI-specific antibiotics; (4) positive UCx + fever and/or UTI symptoms; (5) algorithm 4 with negation for fever without UTI symptoms. Natural language processing (NLP) was used for processing free-text medical notes. The algorithms were validated in 1258 potential UTI episodes from January to March 2012 and results extrapolated to all UTI episodes within this period ($N = 16{,}712$). The reference standard for HA-UTIs was manual record review according to the European Centre for Disease Prevention and Control (and US Centers for Disease Control and Prevention) definitions by trained healthcare personnel.

Findings: Of the 1258 UTI episodes, 163 fulfilled the ECDC HA-UTI definition and the algorithms classified 391, 150, 189, 194, and 153 UTI episodes, respectively, as HA-UTI. Algorithms 1, 2, and 3 had insufficient performances. Algorithm 4 achieved better performance and algorithm 5 performed best for surveillance purposes with sensitivity 0.667 (95% confidence interval: 0.594–0.733), specificity 0.997 (0.996–0.998), positive predictive value 0.719 (0.624–0.807) and negative predictive value 0.997 (0.996–0.997).

* Corresponding author. Address: Department of Medicine Solna (MedS), Division of Infectious Disease, Karolinska Institutet, 171 77, Stockholm, *Sweden*. Tel.: + 46 76 3434 432.

E-mail address: suzanne.ruhe.van.der.werff@ki.se (S.D. van der Werff).

https://doi.org/10.1016/j.jhin.2021.01.023

Conclusion: A fully automated surveillance algorithm based on NLP to find UTI symptoms in free-text had acceptable performance to detect HA-UTI compared to manual record review. Algorithms based on administrative and microbiology data only were not sufficient.

Introduction

Healthcare-associated infections (HCAIs) annually affect millions of patients, are a major burden for the healthcare system, and are associated with prolonged hospital stay, increased morbidity, mortality, and costs [1—3]. Healthcare-associated urinary tract infections (HA-UTIs) account for nearly 20% of all HCAIs, affecting nearly 870,000 patients yearly in Europe [3].

A significant proportion of HCAIs can be prevented [1]. Therefore, to allocate necessary resources and evaluate the effect of interventions, continuous surveillance with feedback to healthcare personnel and stakeholders is important [4,5]. Much HCAI surveillance is currently based on time-consuming and resource-intensive manual review of patient records, which is also prone to subjective interpretation and surveillance bias [6—8].

With the use of electronic health records (EHRs), there is increasing access to detailed electronic health data. This digitalization allows automated surveillance systems to replace manual approaches and to generate standardized and continuous surveillance data [9]. However, surveillance algorithms need to be thoroughly validated before being implemented in a clinical setting.

In this study, the aim was to develop a fully automated rule-based surveillance algorithm using EHR data for the detection of HA-UTI in hospitalized patients, and validate it against manual record review according to the HA-UTI definitions of, primarily, the European Centre for Disease Prevention and Control (ECDC) and, secondly, the US Centers for Disease Control and Prevention (CDC). To demonstrate a possible use-case, the best-performing algorithm was used to determine HA-UTI incidence during a three-year period in all hospitalized patients.

Methods

Study design and data source

This retrospective observational study was performed at the Karolinska University Hospital (KUH) in Stockholm, Sweden. We used prospectively entered routine healthcare data from the electronic health record (EHR) stored as a duplicate of the operating EHR system in a research databank called Health Bank — Swedish Health Record Research Bank, comprising all patients receiving care at KUH between 2006 and 2013 [10]. Data included demographics, hospital administrative data, International Classification of Diseases (ICD)-10 codes, microbiological results, clinical chemistry results, physiological parameters, medication, and medical notes. The study population consisted of hospitalized patients aged $\geq$18 years between 2010 and 2013, the most recent period available (Figure 1). Patients admitted to obstetric wards were excluded because of lack of complete data. The study was approved by the Regional Ethical Review Board in Stockholm under permission nos. 2016/2309-32 and 2012/1838-31/3.

The algorithms were functionally developed using a *development dataset* of all 2979 admissions with a positive urine culture (UCx) from July 2010 to March 2011. Additionally, to perform natural language processing (NLP), this dataset was complemented with annotated free-text about UTI symptoms from 200 admissions with positive UCx from 2012 (selected so that they did not overlap with the validation dataset). To assess algorithm performance, admissions from a *validation period cohort* from January to March 2012 ($N = 15{,}986$) were divided into three separate groups: (1) admissions with at least one positive UCx ($N = 679$), (2) admissions with only negative UCx ($N = 1913$), and (3) admissions without UCx performed ($N = 13{,}394$) (Figure 1).

From this validation period cohort, a *validation dataset* of admissions ($N = 933$), in which UTI presence or not was classified based on medical record review as the reference standard, was selected based on the three admission groups. From admission group 1, the admissions with positive UCx, 533 admissions were selected and annotated, stratified based on day of culture, for the validation dataset. To include patients most likely to have HA-UTI, all admissions were included when positive UCx were only present on day 3 or later of admission ($N = 375$), or when positive UCx were present on day 1 or 2 and on day 3 or later of admission ($N = 14$). To include patients most likely to have community-acquired (CA)-UTI, all admissions during one month (January 2012) were included in which positive UCx were only present on day 1 or 2 of admission ($N = 144$). From the admission groups 2 and 3, admissions with only negative UCx or without UCx performed, 200 randomly sampled admissions per group were selected and annotated for the validation dataset. All performed UCx during admission were regarded as potential UTI episodes, and an admission with no UCx counted as one potential UTI episode.

The medical record review was performed by five trained professionals and UTI classification was based on the ECDC and CDC definitions (see below). As a run-in period, six patients were reviewed together, and further reviewing was performed independently with an overlap of 70 patients. There was substantial agreement between reviewers, with Cohen's κ of 0.82—1.00 for UTI classification. Complicated cases were classified using a consensus decision. The reviewers were blinded to the results of the algorithms.

Case definitions of HA-UTI

The rule-based algorithms were developed to detect microbiologically confirmed symptomatic HA-UTIs according to the definitions of ECDC (UTI-A definition) and CDC [11,12]. Symptomatic UTIs that were not microbiologically confirmed (ECDC UTI-B definition) were not assessed. Furthermore, HA-UTI were categorized as urinary catheter (CAD; catheter à

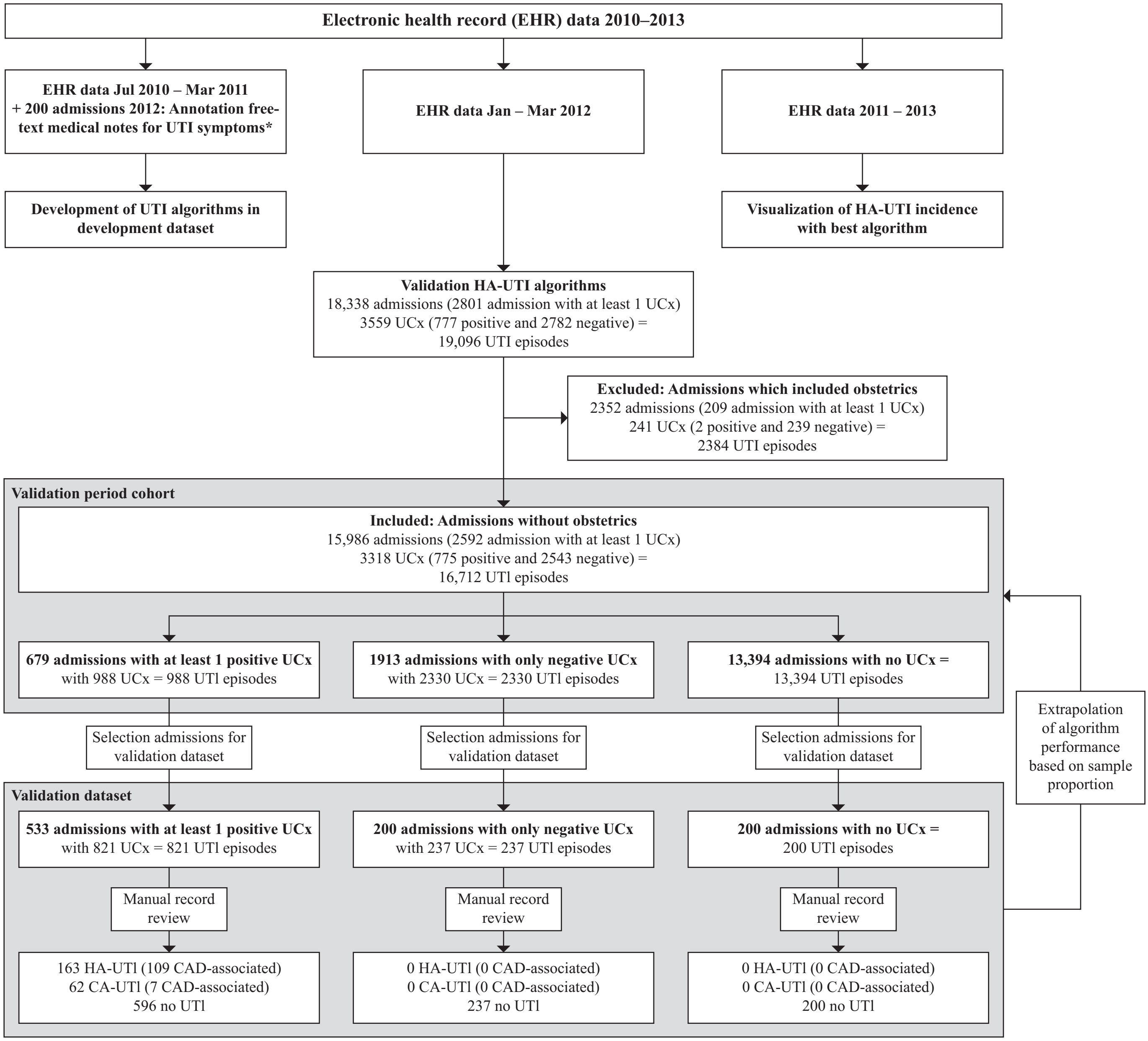

Figure 1. Flow chart of study. *Selected so that they did not overlap with the validation dataset. CA, community-acquired; CAD, urinary catheter (catheter à demeure); HA, healthcare-associated; UCx, urine culture; UTI, urinary tract infection. UTI episode: all performed UCx were regarded as potential UTI episodes during an admission, and admissions with no UCx counted as one potential UTI episode. Positive UCx: urine culture with ≤ 2 pathogens and with at least one pathogen having $>10^5$ colony-forming units (cfu) per millilitre of urine. Negative UCx: urine culture with >2 pathogens, only mixed flora or ≤ 2 pathogens with $\leq 10^5$ cfu per millilitre of urine.

demeure)-associated or not, as described by ECDC and CDC. When multiple episodes during an admission fulfilled the UTI criteria, a window of 14 days, based on the culture date, after the first positive episode was applied where no new HA-UTI episodes could be registered. If the criteria were met again after those 14 days, a new episode of HA-UTI could be registered, similar to the Repeat Infection Timeframe (RIT) criteria of the CDC [12].

Algorithms

Five possible rule-based algorithms for the detection of UTI were developed in the development dataset (Figure 2):

1. Positive UCx;
2. Positive UCx combined with UTI related ICD-10 codes (Appendix Table A.1) recorded during admission;

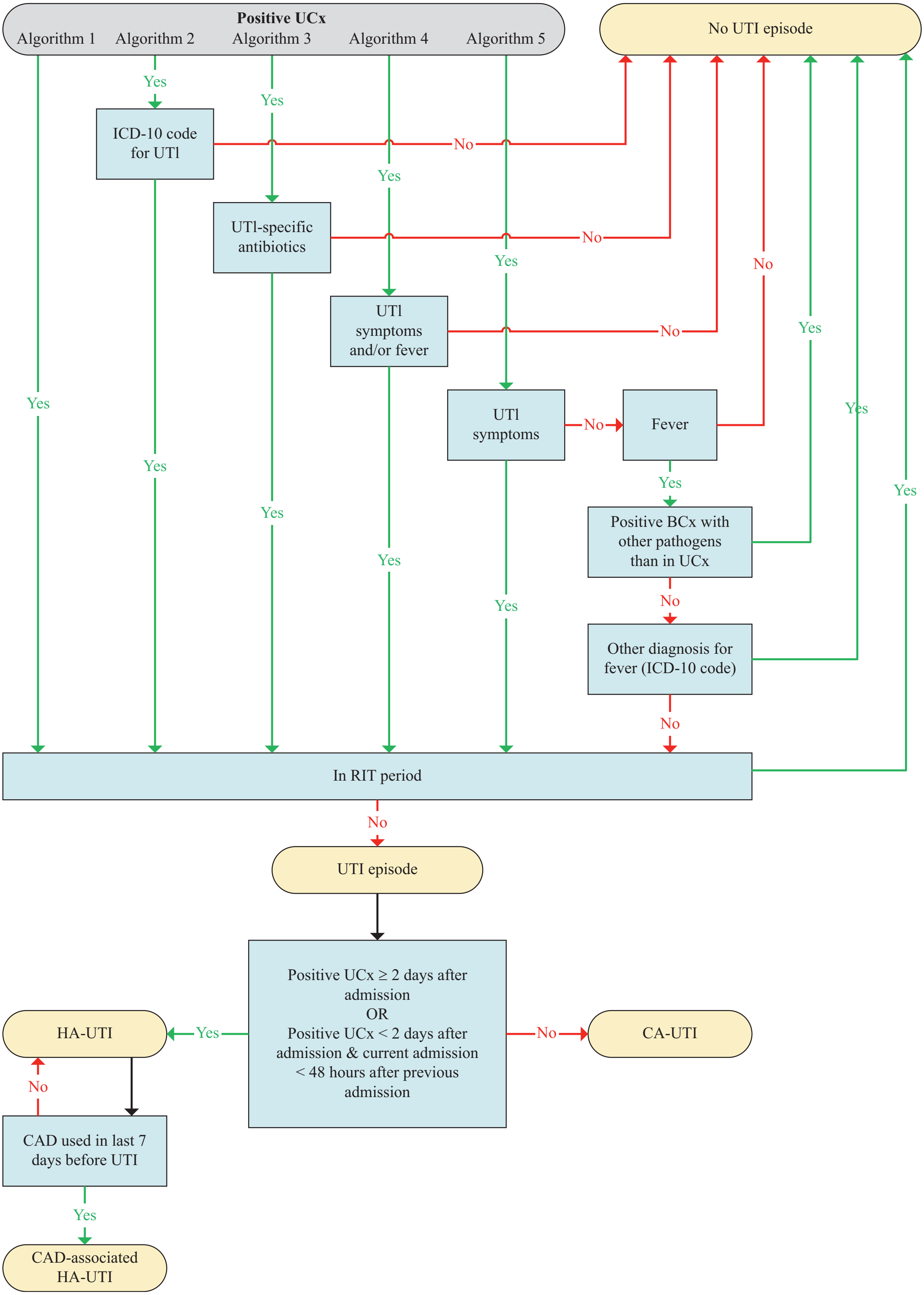

Figure 2. Flow diagram of how the five different rule-based algorithms classify healthcare-associated urinary tract infection. BCx, blood culture; CA, community-acquired; CAD, urinary catheter (catheter à demeure); HA, healthcare-associated; ICD-10, International

3. Positive UCx combined with UTI-specific antibiotics (Appendix Table A.1);
4. Positive UCx combined with fever and/or UTI symptoms (in accordance with ECDC and CDC definitions, respectively);
5. Algorithm 4 with negating rule for cases with fever and without UTI symptoms by non-correspondent positive blood cultures (BCx) or relevant ICD-10 codes (Appendix Table A.1).

Positive UCx was defined as UCx with not more than two pathogens (any bacteria or fungi except mixed flora) and with at least one pathogen having $>10^5$ colony-forming units (cfu) per millilitre of urine. UTI-specific antibiotics were considered when started one day before to seven days after the positive UCx. Fever, UTI symptoms, and non-correspondent positive BCx were taken into account when present within two days before or after the positive UCx. UTIs were defined as healthcare-associated (HA) when the UCx was taken two days or more after admission (on day 3 of admission where admission is day 1) or within two days after admission (on day 1 or 2 of admission) if a previous admission within 48 h was present. HA-UTIs were defined as CAD-associated when a CAD term was present within seven days before the UCx. The earlier described RIT period of 14 days was also used by the algorithms to assess when another UTI episode could be registered when multiple potential UTI episodes were present during one admission. Consequently, during admissions of more than 14 days when multiple UTI episodes more than 14 days apart from each other were present, one CA-UTI and/or one or more HA-UTIs could potentially be classified during one single admission.

Fever could be present as a numeric measurement value recorded in the EHR, but also described in free-text medical notes. The UTI symptoms, i.e. urgency, frequency, dysuria, suprapubic tenderness, or costovertebral angle pain or tenderness (last only used for CDC algorithms), and presence of CAD were only described in free-text medical notes. NLP was used to detect fever, UTI symptoms, and CAD presence in these medical notes. Regular expressions (regex) were built to automatically detect phrases and terms in sentences of medical notes. These were based on the annotated medical notes from the development dataset for UTI symptoms and for fever and CAD presence on the expert opinion of medical doctors. Negation cues present in the same sentence were used to determine whether UTI symptoms were negated (e.g. 'no dysuria present', 'urgency not mentioned by patient'). A more complex NLP approach based on the annotated medical notes was explored. As this increased the sensitivity to a small extent, but largely decreased the specificity, this approach was eventually not used in the study.

Statistical analysis

Data handling and algorithm calculations were performed using Python version 3.7. The programme pyConTextNLP version 0.7.0.0 was used for the NLP to detect UTI-symptom regex (with negation cues), and fever and CAD regex (without negation cues) [13,14]. Statistical analyses were performed using R version 3.6.1 [15]. Continuous variables are presented as median with interquartile ranges and categorical variables as numbers with percentages. For algorithm performance, the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were assessed. The confidence intervals (CIs) for these estimates in the validation dataset were calculated using the asymptotic variance with Wilson score method [16].

The results of the algorithms from the validation dataset were extrapolated to the validation period cohort of January–March 2012 to obtain correct performance estimates of the algorithms in the intended target population who are all hospitalized patients (Figure 1) [17]. This extrapolation was done based on the sampling proportion of the validation dataset from the validation period cohort within the separate admission-groups. The CIs for these estimates were calculated as the 2.5th and 97.5th percentiles of point estimates obtained from 10,000 bootstrap samples for each of the five groups using the R package 'boot' [18]. To account for uncertainty, the bootstrapping was performed before extrapolating the proportions from the validation dataset to the validation period cohort. Area under the receiver operating characteristic (ROC) curve (AUC) was determined using the R package 'pROC' [19]. Finally, the best performing algorithm was applied to the data of 2011–2013 to show how the incidence of HA-UTI by continuous surveillance would look.

Results

In the validation dataset among the 533 admissions with at least one positive UCx, 821 UCx were performed. In the 200 admissions with only negative UCx, 237 UCx were performed. Together with 200 admissions with no UCx performed, this amounted to 1258 potential UTI episodes in the validation dataset used for the calculations of algorithm performance (Table I and Figure 1). In the admissions of the validation period cohort (N =15,986) of January to March 2012 there were 16,712 potential UTI episodes used for the calculation of extrapolated algorithm performance (Figure 1). The patients within the validation dataset with a positive UCx were older and consisted of more females than the patients with negative or no UCx (Table I). The length of hospital stay, ICU admissions, and in-hospital mortality were highest in admissions with at least one positive UCx and lowest in admissions with no UCx.

No UTIs, according to the ECDC UTI-A or CDC definition, were identified during manual record review in admissions with only negative UCx or no UCx. Among the 821 potential UTI episodes in admissions with positive UCx in the validation dataset, 163 UTI episodes (19.9%) fulfilled ECDC HA-UTI (UTI-A) definition and 166 UTI episodes (20.2%) fulfilled CDC HA-UTI definition. The five algorithms classified 391 (47.6%), 150 (18.3%), 189 (23.0%), 194 (23.6%), and 153 (18.6%) UTI episodes, respectively, as ECDC HA-UTI. In Appendix Tables A.2

Statistical Classification of Diseases and Related Health Problems 10th Revision; RIT, Repeat Infection Timeframe; UCx, urine culture; UTI, urinary tract infection. UTI symptoms: urgency, frequency, dysuria, and/or suprapubic tenderness (ECDC)/urgency, frequency, dysuria, suprapubic tenderness and/or costovertebral angle pain or tenderness (CDC). Positive UCx: urine culture with $\leq$2 pathogens and with at least one pathogen having $>10^5$ colony-forming units (cfu) per millilitre of urine. UTI episode: all performed UCx were regarded as potential UTI episodes during an admission, and admissions with no UCx counted as one potential UTI episode.

Table I
Characteristics of admissions or patients in validation cohort

Characteristics	All	Positive UCx	Negative UCx	No UCx
No. of UTI episodes/no. of urine cultures (% of all)	1258 (100)	821 (65.3)	237 (18.8)	200 (15.9)
No. of admissions (% of all)	933 (100)	533 (57.1)	200 (21.4)	200 (21.4)
No. of patients (% of all)	915 (100)	525 (57.4)	192 (21.0)	198 (21.6)
No. of females (% of patients)	555 (60.7)	356 (67.8)	95 (49.5)	104 (52.5)
Age (years), median (IQR) of admissions	72 (59–81)	76 (65–85)	69.5 (55–79)	62.5 (47–73)
Length of stay (days), median (IQR) of admissions	8 (4–16)	12 (6–22)	7 (3–13)	4 (2–7)
No. of ICU admissions (% of admissions)	109 (11.7)	81 (15.2)	19 (9.5)	9 (4.5)
No. of in-hospital mortalities (% of admissions)	36 (3.9)	28 (5.5)	7 (3.5)	1 (0.5)
No. of ECDC HA-UTIs (% of episodes)	163 (13.0)	163 (19.9)	0	0
No. of ECDC CAD-associated HA-UTIs (% of episodes)	109 (8.7)	109 (13.3)	0	0
No. of ECDC CA-UTIs (% of episodes)	62 (4.9)	62 (7.6)	0	0
No. of CDC HA-UTIs (% of episodes)	166 (13.1)	166 (20.2)	0	0
No. of CDC CAD-associated HA-UTIs (% of episodes)	95 (7.6)	95 (11.6)	0	0
No. of CDC CA-UTIs (% of episodes)	62 (4.9)	62 (7.6)	0	0
No. of ECDC and/or CDC HA-UTIs (% of episodes)	181 (14.4)	181 (22.0)	0	0
No. of ECDC and/or CDC CAD-associated HA-UTIs (% of episodes)	125 (9.9)	125 (15.2)	0	0
No. of ECDC and/or CDC CA-UTIs (% of episodes)	71 (5.6)	71 (8.6)	0	0

UCx, urine culture; UTI, urinary tract infection; IQR, interquartile range; ICU, intensive care unit; ECDC, European Centre for Disease Prevention and Control; HA, healthcare-associated; CAD, urinary catheter (catheter à demeure) CA, community-acquired; CDC, US Centers for Disease Control and Prevention.

and A.4–A.6 the number of true positives, false positives, false negatives, and true negatives within the different groups and for the five algorithms are shown for the ECDC HA-UTI, CDC HA-UTI, ECDC CAD-associated HA-UTI and CDC CAD-associated HA-UTI outcome, respectively. The reasons for the differences in classification between algorithms 4 and 5 for the ECDC HA-UTI outcome is shown in Appendix Table A.3.

The algorithm based on positive UCx only (algorithm 1) had a high sensitivity but a low PPV to classify HA-UTI according to the ECDC definition (Table II). The algorithm based on positive UCx and UTI-related ICD-10 codes (algorithm 2) had a low sensitivity although the PPV increased compared to the previous algorithm. In the algorithm that used positive UCx with UTI-specific antibiotics (algorithm 3), the sensitivity and PVV increased compared to algorithm 2. The algorithm that was based on adding fever and UTI symptoms through text-mining of free-text to positive UCx (algorithm 4) had a sensitivity of 0.721 (95% CI: 0.655–0.782) and PPV of 0.613 (0.532–0.694) in the validation period cohort. The algorithm that added to algorithm 4 a negating rule for cases with fever without

Table II
Performance characteristics of five rule-based algorithms for classifying healthcare-associated urinary tract infection according to ECDC definition UTI-A

Algorithm	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC (95% CI)
Validation dataset with admissions with at least one positive UCx (UTI episodes $N = 821$)					
1	0.933 (0.883–0.962)	0.637 (0.599–0.673)	0.389 (0.342–0.438)	0.974 (0.955–0.986)	0.785 (0.758–0.811)
2	0.399 (0.327–0.475)	0.871 (0.843–0.894)	0.443 (0.357–0.513)	0.854 (0.825–0.879)	0.635 (0.595–0.673)
3	0.577 (0.500–0.650)	0.856 (0.827–0.880)	0.497 (0.427–0.568)	0.891 (0.864–0.913)	0.716 (0.676–0.757)
4	0.730 (0.657–0.792)	0.886 (0.859–0.908)	0.613 (0.543–0.679)	0.930 (0.907–0.947)	0.808 (0.772–0.844)
5	0.675 (0.600–0.742)	0.935 (0.913–0.951)	0.719 (0.643–0.784)	0.921 (0.898–0.939)	0.805 (0.768–0.842)
Extrapolated results to validation period cohort (UTI episodes $N = 16,712$)					
1	0.921 (0.885–0.952)	0.986 (0.984–0.987)	0.389 (0.356–0.422)	0.999 (0.999–1.000)	0.953 (0.933–0.974)
2	0.394 (0.321–0.467)	0.995 (0.994–0.996)	0.433 (0.335–0.535)	0.994 (0.993–0.995)	0.694 (0.657–0.732)
3	0.570 (0.491–0.648)	0.994 (0.993–0.995)	0.497 (0.413–0.585)	0.996 (0.995–0.996)	0.782 (0.744–0.820)
4	0.721 (0.655–0.782)	0.995 (0.994–0.997)	0.613 (0.532–0.694)	0.997 (0.997–0.998)	0.858 (0.824–0.893)
5	0.667 (0.594–0.733)	0.997 (0.996–0.998)	0.719 (0.624–0.807)	0.997 (0.996–0.997)	0.832 (0.796–0.868)

ECDC, European Centre for Disease Prevention and Control; UTI, urinary tract infection; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic (ROC) curve; CI, confidence interval; UCx, urine culture.
The extrapolated results of the algorithms from the validation dataset to the validation period cohort (Jan–Mar 2012) were based on the sampling proportion of potential UTI-episodes from the five different groups: (1) admissions with a positive urine culture (UCx) only present on day 3 or later of admission; (2) admissions with a positive UCx both on day 1 or 2 and on day 3 or later of admission; (3) admissions with a positive UCx only present on day 1 or 2 of admission; (4) admissions with only negative UCx and; (5) admissions without a UCx.

UTI symptoms by non-correspondent positive BCx or relevant ICD-10 codes (algorithm 5) resulted in slightly decreased sensitivity of 0.667 (0.594–0.733) but an increase in PPV to 0.719 (0.624–0.807). In the validation period cohort (extrapolated data), the specificity and NPV was high for all five algorithms.

The sensitivity of algorithm 4 to detect HA-UTI according to CDC definition compared to ECDC definition was slightly higher, while the PPV was similar (Appendix Table A.7). For algorithm 5, sensitivity and PPV were somewhat lower to detect CDC HA-UTI compared to ECDC HA-UTI. The sensitivity and PPV of algorithm 4 and 5 to detect CAD-associated HA-UTI were lower compared to HA-UTI regardless of CAD presence, both for ECDC and CDC HA-UTI (Appendix Tables A.8 and A.9).

Algorithm 5 was used to show what the incidence rate of ECDC HA-UTI over the period 2011–2013 would look like (Appendix Figure A.1). The algorithm-determined hospital-wide incidence rate ranged from 0.92 to 1.73 per 1000 bed-days and was quite stable over time. When analysing specific ward specialties, it can be seen that, compared to the hospital-wide trend, the incidence rate was similar in internal medicine wards, similar but more fluctuating in surgery wards, and higher in geriatric wards.

Discussion

This study shows that by using EHR data and NLP it is possible to develop an acceptable and fully automated algorithm for microbiologically confirmed symptomatic HA-UTI. Such an algorithm may be used for surveillance purposes and could be implemented to replace manual surveillance by record review.

This study explored whether simple algorithms, based on (i) microbiology only or (ii) a combination of microbiology results with UTI ICD-10 codes or UTI-specific antibiotics, could be sufficient to assess HA-UTIs or whether more sophisticated algorithms, e.g. by using NLP, would be necessary. The former kind of algorithms would be easier to implement in different kinds of systems and circumstances than the latter. However, the results indicate that an algorithm relying on microbiology only would lead to an overestimation of the incidence of HA-UTI, whereas relying on the combination of microbiology and ICD-10 codes would lead to an underestimation of the incidence. An algorithm relying on microbiology and antibiotics performed slightly better but would result in overestimation of the incidence. These findings correspond with the results of a recent systematic review on the performance of different algorithms for automated surveillance systems of HCAIs, including UTI, concluding that ICD coding has a low sensitivity and that microbiology results alone are typically not sufficient [20]. More sophisticated algorithms using more parameters and techniques such as NLP had better performances, which was also the case in this study. For epidemiological surveillance, higher specificity and PPV are of importance and the combination of microbiology with presence of UTI symptoms (algorithm 4) reduced the number of false-positives. The algorithm adding a negating rule for cases with fever and without UTI symptoms (algorithm 5) had a sensitivity that was only slightly reduced compared to algorithm 4, but further improved the PPV.

The algorithms performed similarly for the detection CDC HA-UTI compared to ECDC HA-UTI; only the algorithms for CAD-associated CDC HA-UTIs performed slightly worse compared to the ECDC version. This might be explained by the fact that the timeframe to look for CAD was based on ECDC criteria rather than CDC criteria, whereas UTI symptoms for ECDC versus CDC algorithms were based on their respective differences in definitions. However, these results show that the definitions of ECDC and CDC are close enough for algorithms to have similar performance for surveillance purposes.

Studies have tried to develop surveillance algorithms for (CAD-associated) HA-UTI [21–38]. Compared to these studies, the specificity and NPV of the best algorithm in this study was similar to that of the other studies. Compared to semi-automated surveillance algorithms, the sensitivity and PPV was lower in this study [21–34]. By contrast, compared to other fully automated surveillance algorithms, the sensitivity and PPV of algorithms 4 and 5 in this study were in general slightly better but in a similar range [35–38]. Fully automated algorithms with both high sensitivity and specificity are difficult to develop, and the algorithms developed in our study are at present useful for surveillance but not for diagnostic purposes. However, the development of diagnostic algorithms with both high sensitivity and specificity would be beneficial for decision support systems.

A major strength of this study is that it is based on a large dataset that is representative of the clinical population for which the algorithm is designed. Furthermore, comprehensive EHR data were used, facilitating implementation in a real clinical setting. Also, the NLP method used is one of the simplest methods and should be less troublesome to apply in different settings than more complicated methods. Finally, the manual annotation has been done according to the internationally recognized definitions of ECDC and CDC, increasing the comparability with other studies. Limitations include that the developed algorithms have not been validated in obstetric and paediatric populations. Yet in the obstetric admissions hardly any positive UCx were present and in children the rate of HA-UTI is lower compared to adults, so they would probably not greatly influence the results [39]. Even though the algorithms were developed in a copy of EHR data, they have not yet been implemented and evaluation in real-life settings is warranted. Additionally, the algorithms need to be tested in other hospitals/EHR systems to assess generalizability. Although a simple method of NLP has been used for the algorithms, this algorithm is not directly applicable to other languages and the regex would need to be adapted accordingly and the algorithms revalidated. As the best-performing algorithm partly uses ICD-10 codes for the negation of fever without UTI symptoms, and ICD-10 codes are mostly recorded in the EHR at discharge, this reduces the applicability of this algorithm to some extent for real-time surveillance of HA-UTI. Although the developed fully-automated surveillance algorithm could replace manual surveillance, this system would need readjustment and validation as registration and reporting of elements within the algorithm change over time. The performance of algorithms will also depend on how well the different elements are recorded in the EHR system, but this also applies to manual surveillance by record review. There were no quality assurance data available for the different components used for the record review and algorithm classification in this study to assess the possible error arising from this. Finally, the algorithms were developed to assess microbiologically confirmed symptomatic HA-UTIs, so they do not apply to microbiologically non-confirmed symptomatic HA-UTIs (UTI-B definition of ECDC).

Semi-automated algorithms are likely to outperform fully automated algorithms, but this still means that time-consuming manual annotation will be necessary and subjectivity in assessing UTIs or other HCAIs will remain present. If one wants to use the system for correct assessment of HCAIs on the individual patient level for diagnostic and clinical reasons, then semi-automated approaches are probably the most suitable. However, if the individual patient's treatment is not the focus, then fully automated algorithms are an option. This can be used for continuous and real-time surveillance on HCAIs, which will give more direct feedback to wards about their disease burden and will help to assess the effect of infection prevention and control (IPC) interventions more effectively. Furthermore, fully automated surveillance is easier to scale up, can be used on larger populations or in surveillance networks, and can be more standardized and objective if thoroughly validated in the settings where it will be used. This is very useful for public reporting or other situations when one wants to compare different healthcare facilities. Fully automated surveillance ideally should be developed for all the major HCAIs and thereby preferably not only focus on device-associated HCAI, e.g. UTI vs CAD-associated UTI, to be useful for quality improvement of healthcare and supporting IPC in a broad sense. The use of NLP was useful in this study and developments in machine learning may further improve the performance of fully automated algorithms [40]. Future research should explore these possibilities.

In conclusion, a fully automated surveillance algorithm for HA-UTI – based on positive urine culture combined with simple NLP for UTI symptom detection and negation of fever when being the sole symptom, and using EHR data – performed well compared to manual record review. This algorithm can be used for continuous surveillance of HA-UTI and may replace manual surveillance with the benefit of quick, continuous, standardized and objective feedback to healthcare personnel and effective evaluation of interventions to reduce the incidence of HA-UTI.

Acknowledgements

We would like to thank A. Frej and A. Sandberg for their help with the study.

Conflict of interest statement
None declared.

Funding sources
The work was supported by Sweden's Innovation Agency (Vinnova grant 2016-00563). P.N. was supported by Region Stockholm (clinical research appointment). J.K.V. was supported by Region Stockholm (combined clinical residency and PhD training programme).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhin.2021.01.023.

References

[1] Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. Infect Control Hosp Epidemiol 2011;32:101–14. https://doi.org/10.1086/657912.

[2] Cassini A, Plachouras D, Eckmanns T, Abu Sin M, Blank H-P, Ducomble T, et al. Burden of six healthcare-associated infections on European population health: Estimating incidence-based disability-adjusted life years through a population prevalence-based modelling study. PLoS Med 2016;13:e1002150. https://doi.org/10.1371/journal.pmed.1002150.

[3] Suetens C, Latour K, Kärki T, Ricchizzi E, Kinross P, Moro ML, et al. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. Euro Surveill 2018;23:pii=1800516. https://doi.org/10.2807/1560-7917.ES.2018.23.46.1800516.

[4] van Mourik MSM, Perencevich EN, Gastmeier P, Bonten MJM. Designing surveillance of healthcare-associated infections in the era of automation and reporting mandates. Clin Infect Dis 2018;66:970–6. https://doi.org/10.1093/cid/cix835.

[5] Storr J, Twyman A, Zingg W, Damani N, Kilpatrick C, Reilly J, et al. Core components for effective infection prevention and control programmes: new WHO evidence-based recommendations. Antimicrob Resist Infect Control 2017;6:6. https://doi.org/10.1186/s13756-016-0149-9.

[6] Mitchell BG, Hall L, Halton K, MacBeth D, Gardner A. Time spent by infection control professionals undertaking healthcare associated infection surveillance: a multi-centred cross sectional study. Infect Dis Health 2016;21:36–40. https://doi.org/10.1016/j.idh.2016.03.003.

[7] Russo PL, Barnett AG, Cheng AC, Richards M, Graves N, Hall L. Differences in identifying healthcare associated infections using clinical vignettes and the influence of respondent characteristics: a cross-sectional survey of Australian infection prevention staff. Antimicrob Resist Infect Control 2015;4:29. https://doi.org/10.1186/s13756-015-0070-7.

[8] Haut ER, Pronovost PJ. Surveillance Bias in Outcomes Reporting. JAMA 2011;305:2462. https://doi.org/10.1001/jama.2011.822.

[9] Sips ME, Bonten MJM, van Mourik MSM. Automated surveillance of healthcare-associated infections: state of the art. Curr Opin Infect Dis 2017;30:425–31. https://doi.org/10.1097/QCO.0000000000000376.

[10] Dalianis H, Henriksson A, Kvist M, Velupillai S, Weegar R. HEALTH BANK – A Workbench for Data Science Applications in Healthcare. In: Proceedings of the CAiSE-2015 Industry Track co-located with 27th Conference on Advanced Information Systems Engineering, vol. 1381; 2015. p. 18–21.

[11] European Centre for Disease Prevention and Control. Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals – protocol version 5.3. Stockholm: ECDC; 2016. Available at: https://doi.org/10.2900/374985 [last accessed February 2021].

[12] Centers for Disease Control and Prevention. National Healthcare Safety Network (NHSN) patient safety component manual. 2019. Available from: https://www.cdc.gov/nhsn/pdfs/validation/2019/pcsmanual_2019-508.pdf [last accessed February 2021].

[13] Chapman B, Chapman WW, Dayton G, Mowery D. pyConTextNLP. 2019. Available from: https://pypi.org/project/pyConTextNLP/ [last accessed February 2021].

[14] Chapman BE, Lee S, Kang HP, Chapman WW. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. J Biomed Inform 2011;44:728–37. https://doi.org/10.1016/J.JBI.2011.03.011.

[15] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available from: https://www.R-project.org/ [last accessed February 2021].

[16] Brown LD, Cai TT, Dasgupta A. Interval estimation for a binomial proportion. Statist Sci 2001;16:101–17.

[17] Karlsson Valik J, 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;29:735–45. https://doi.org/10.1136/bmjqs-2019-010123.

[18] Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-22. 2019. Available from: https://CRAN.R-project.org/package=boot [last accessed February 2021]

[19] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77. https://doi.org/10.1186/1471-2105-12-77.

[20] Streefkerk HRA, Verkooijen RP, Bramer WM, Verbrugh HA. Electronically assisted surveillance systems of healthcare-associated infections: a systematic review. Euro Surveill 2020;25: pii=1900321. https://doi.org/10.2807/1560-7917.ES.2020.25.2.1900321.

[21] 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;24:51–5. https://doi.org/10.1086/502115.

[22] 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;125:34–9. https://doi.org/10.1309/502AUPR8VE67MBDE.

[23] 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;79:38–43. https://doi.org/10.1016/J.JHIN.2011.05.006.

[24] 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;32:757–62. https://doi.org/10.1086/661103.

[25] 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;51:509–16. https://doi.org/10.1097/MLR.0b013e31828d1210.

[26] Lo Y-S, Lee W-S, Liu C-T. Utilization of electronic medical records to build a detection model for surveillance of healthcare-associated urinary tract infections. J Med Syst 2013;37:9923. https://doi.org/10.1007/s10916-012-9923-2.

[27] de Bus L, Diet G, Gadeyne B, Leroux-Roels I, Claeys G, Steurbaut K, et al. Validity analysis of a unique infection surveillance system in the intensive care unit by analysis of a data warehouse built through a workflow-integrated software application. J Hosp Infect 2014;87:159–64. https://doi.org/10.1016/J.JHIN.2014.03.010.

[28] Venable A, Dissanaike S. Is automated electronic surveillance for healthcare-associated infections accurate in the burn unit? J Burn Care Res 2013;34:591–7. https://doi.org/10.1097/BCR.0b013e3182a2aa0f.

[29] Stamm AM, Bettacchi CJ. A comparison of 3 metrics to identify health care-associated infections. Am J Infect Control 2012;40:688–91. https://doi.org/10.1016/J.AJIC.2012.01.033.

[30] Streefkerk RHRA, Borsboom GJJM, van der Hoeven CP, Vos MC, Verkooijen RP, Verbrugh HA. Evaluation of an algorithm for electronic surveillance of hospital-acquired infections yielding serial weekly point prevalence scores. Infect Control Hosp Epidemiol 2014;35:888–90. https://doi.org/10.1086/676869.

[31] 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;35:685–91. https://doi.org/10.1086/676429.

[32] Hsu HE, Shenoy ES, Kelbaugh D, Ware W, Lee H, Zakroysky P, et al. An electronic surveillance tool for catheter-associated urinary tract infection in intensive care units. Am J Infect Control 2015;43:592–9. https://doi.org/10.1016/J.AJIC.2015.02.019.

[33] Streefkerk HRA, Lede IO, Eriksson JL van, Meijling MG, van der Hoeven CP, Wille JC, et al. Internal and external validation of a computer-assisted surveillance system for hospital-acquired infections in a 754-bed general hospital in the Netherlands. Infect Control Hosp Epidemiol 2016;37:1355–60. https://doi.org/10.1017/ice.2016.159.

[34] Marra AR, Alkatheri M, Edmond MB. Catheter-associated urinary tract infection: utility of the ICD-10 metric as a surrogate for the national healthcare safety network (NHSN) surveillance metric. Infect Control Hosp Epidemiol 2017;38:506–7. https://doi.org/10.1017/ice.2016.335.

[35] Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural language processing for real-time catheter-associated urinary tract infection surveillance: results of a pilot implementation trial. Infect Control Hosp Epidemiol 2015;36:1004–10. https://doi.org/10.1017/ice.2015.122.

[36] Condell O, Gubbels S, Nielsen J, Espenhain L, Frimodt-Møller N, Engberg J, et al. Automated surveillance system for hospital-acquired urinary tract infections in Denmark. J Hosp Infect 2016;93:290–6. https://doi.org/10.1016/J.JHIN.2016.04.001.

[37] 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;91:231–6. https://doi.org/10.1016/J.JHIN.2015.05.011.

[38] Tanushi H, Kvist M, Sparrelid E. Detection of healthcare-associated urinary tract infection in Swedish electronic health records. Stud Health Technol Inform 2014;207:330–9. https://doi.org/10.3233/978-1-61499-474-9-330.

[39] Zingg W, Hopkins S, Gayet-Ageron A, Holmes A, Sharland M, Suetens C, et al. Health-care-associated infections in neonates, children, and adolescents: an analysis of paediatric data from the European Centre for Disease Prevention and Control point-prevalence survey. Lancet Infect Dis 2017;17:381–9. https://doi.org/10.1016/S1473-3099(16)30517-5.

[40] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347–58. https://doi.org/10.1056/NEJMra1814259.