Review

Risk factors for healthcare-associated urinary tract infection and their applications in surveillance using hospital administrative data: a systematic review

C. King*, L. Garcia Alvarez, A. Holmes, L. Moore, T. Galletly, P. Aylin

A The National Centre for Infection Prevention and Management, Imperial College, London, UK
b Infection Prevention and Control, Imperial College Healthcare NHS Trust, London, UK
c Department of Infectious Diseases, Imperial College Healthcare NHS Trust, London, UK
d Dr Foster Unit at Imperial, School of Public Health, Imperial College London, London, UK

SUMMARY

Background: Healthcare-associated urinary tract infections (HCA UTI) account for a large proportion of hospital infections, with recently launched surveillance in the UK focusing on reducing catheter-associated urinary tract infections. However, a wealth of administrative information already collected routinely by hospitals is currently not used to its maximum potential for surveillance.

Aim: To quantify the evidence base of HCA UTI risk factors and to determine their potential for shaping and informing innovative surveillance tools using local hospital data.

Methods: A systematic literature review was undertaken to find established risks for HCA UTI. Population-attributable risk percentages (PAR%) were calculated for these risk factors, generating a hierarchy of risks. Administrative hospital data were subsequently interrogated for these quantified risks.

Findings: Over 30% of the risk factors identified from the systematic literature review were independent predictors of infection. The highest PAR% was associated with urinary catheterization, with the calculation that 79.3% of UTI would be prevented if catheterization was not performed. PAR% calculations were performed for 60% of the independent predictors for HCA UTI. Sixty-five percent of the identified independent risk factors were found to be coded within the administrative hospital dataset, including urinary catheterization.

Conclusion: This work has quantified established HCA UTI risks and demonstrates that there is potential for more effective use of administrative hospital data for risk monitoring and surveillance of HCA UTI.

Introduction

Healthcare-associated urinary tract infections (HCA UTI) contribute a significant burden to healthcare services globally. In the UK, it is estimated that the National Health Service (NHS) incurs 66,160 cases of HCA UTI per year, costing £68 million in excess bed-days. In the 2006 UK and Republic of Ireland prevalence survey of healthcare-associated infections (HCAI),
UTI accounted for 19.9% of HCAI, second in frequency only to gastrointestinal infections (20.6%). The high rate of UTI within hospitals is commonly attributed to the use of urinary catheters, with 31.8% of the NHS patient population having undergone urinary catheterization or bladder instrumentation on the day of the survey or in the preceding seven days.

Traditional surveillance has long been established as an essential component to any infection prevention and control programme, despite the fact that it can be costly to implement and time consuming to conduct. In the UK, there are mandatory surveillance systems for several HCAI, such as *Clostridium difficile* (since January 2004) and meticillin-resistant *Staphylococcus aureus* (MRSA) bloodstream infections (since early 2001). Surveillance of catheter-associated urinary tract infections (CA UTI) is now performed as part of an initiative launched by the UK Department of Health in January 2011 in order to reduce patient harm. However, despite the high burden that HCA UTI represents to the NHS, this surveillance scheme is voluntary, focuses exclusively on CA UTI, and is still in its infancy. The potential utility of innovative HCA UTI surveillance tools is apparent, with current work focusing on the development of automated systems to identify UTI using traditional clinical definitions.

Hospitals routinely collect and store large amounts of administrative data, including information on admissions, microbiology, diagnoses and procedures. Although this wealth of data has not been widely exploited for HCAI surveillance, integrative approaches towards the more effective use of administrative data are being pioneered. Determining the utility of established risk factors for HCA UTI within administrative hospital data presents the opportunity to develop and test innovative surveillance tools with the potential to enhance traditional surveillance of UTI within the healthcare environment.

The aim of this study was to quantify the existing evidence base of risk factors for HCA UTI, create a hierarchy of risk, identify these risks within administrative hospital data, and assess the potential of this data source for informing novel syndromic surveillance tools.

## Methods

### Systematic literature review

A systematic literature review was conducted to establish an evidence base of risk factors for HCA UTI. EMBASE, PubMed/MEDLINE, Web of Science and the Cochrane Database were searched using ‘UTI’, ‘urinary tract infection’ and ‘risk’ as the search terms. The Cochrane Database was included to assess whether or not a comprehensive literature review on HCA UTI risk factors had been undertaken previously. The search strategy and inclusion criteria were discussed and agreed between the authors, and the search was conducted by a single reviewer in April 2011.

The inclusion criteria were: (1) original research papers in English; (2) clear and reasonable study design; (3) demonstrated statistical analysis, and reported odds ratios (OR), relative risks or hazard ratios, and P-values; (4) a clear definition of HCA UTI used to diagnose ‘cases’, either based on published definitions (e.g. US Centers for Disease Control and Prevention) or positive microbiology, antibiotic prescription and clinical data; (5) ‘controls’ clearly stated as not diagnosed with an HCA UTI; and (6) all study participants had been exposed to ‘health care’, defined as either a hospital admission for ≥72 h or a readmission with prior exposure to health care. Those studies investigating risk factors for acquiring an antimicrobial-resistant infection vs an antimicrobial-susceptible infection were excluded as the controls did not meet the inclusion criteria.

Critical Appraisal Skills Programme (CASP) tools, developed by the British Public Health Resource Unit, for case–control studies (available at: http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/cohort%2012%20questions.pdf/view) were employed for critical appraisal of the papers in this study as they were developed specifically for cohort and case–control studies, and have been widely used in review articles. For those papers describing prevalence and surveillance studies, an ‘adjusted CASP cohort tool’ was applied, which included all the aspects considered by the CASP cohort tool with the exception of the follow-up criteria. The maximum and minimum possible CASP scores ranged from 14 to −14 for cohort studies, and from 12 to −12 for case–control, prevalence and surveillance studies.

Those risk factors that were associated with HCA UTI (P < 0.05) at the univariate and multivariate level were selected for further analysis, and were defined as ‘significant’ risks. An ‘independent’ risk factor was considered when its associated P-value was <0.05 at the multivariate level.

### Calculation of population-attributable risk percentages

In order to assess the relative importance of the significant risk factors identified from the literature review, population-attributable risk percentages (PAR%) were calculated. PAR% provides an indication of the proportion of a disease in the total population attributable to a risk factor. If the risk factor was removed from the population, the infection incidence would be reduced by PAR%. In cases where PAR% is negative, the risk factor is protective against infection. The following formula was used to calculate PAR%:

\[
P_{\text{AR}} = \left( \frac{l_i - l_u}{l_i} \right) \times 100
\]

where \(l_i\) is the incidence of HCA UTI in the total population, and \(l_u\) is the incidence of HCA UTI in the unexposed population.

The incidence values were calculated by extracting the number of HCA UTI cases and controls that were exposed and unexposed to each categorical risk factor, and for which published data were available. In those studies where raw data were not published but odds ratios (ORs) and the population size of exposed and unexposed groups were reported, a methodology based on the estimated variance [calculated from the 95% confidence interval (CI) of the OR] described by Pietrantoni was used.

The studies that used prevalence survey methods were excluded from PAR% calculations as this study design is inappropriate for the methodology.

### Identification of risk factors within the available local administrative hospital data

Routine administrative data codes were searched for significant risk factors identified from the systematic literature.
review in order to address the utility of administrative hospital data for the development of surveillance tools.

Imperial College Healthcare NHS Trust (ICHNT) is one of the largest NHS hospital groups in the UK, consisting of five hospitals with 1540 beds and one central laboratory. Over 750 local databases have been identified within ICHNT, comprising a mixture of small departmental datasets, surveillance and larger trust-wide systems. Data on patient demographics, diagnoses [International Classification of Diseases Version 10 (ICD-10)], operations and procedures (OPCS), and admission and discharge data are stored in the large organization-wide patient administration system (PAS). Microbiology and laboratory data are located in the laboratory information management system (LIMS), and further information on antimicrobial prescribing and infection prevention and control is located in locally performed surveillance.

A comprehensive variable list and code book of the hospital data available for this study was created and cross-checked against the list of extracted risk factors for UTI. Keyword searches of the ICD-10 codes were performed, based on the list of HCA UTI risk factors identified during the literature review, using the World Health Organization’s online search tool (http://apps.who.int/classifications/icd10). A Microsoft Excel (2007) reference spreadsheet was generated. ICD-10 codes identified from any included studies were also recorded on the reference spreadsheet. The process was repeated with operation and procedure codes and pathology codes using reference databases provided by ICHNT.

In order to use hospital data for research, ethical approval was granted by St. Mary’s National Research Ethics Committee for the use of administrative hospital data from ICHNT in an anonymized format, i.e. one way anonymization and removal of all patient-identifiable fields prior to data analysis (Ref. No.: 09/H0712/85).

Results

The systematic literature search returned 4109 unique papers. One percent were excluded for not being in English, 8% had titles that were not relevant to the study question, and 6% were excluded as irrelevant after abstract review. The results retrieved from the Cochrane Database returned no relevant studies. In total, 154 studies underwent full-text review and 23 were ultimately included in the study (Figure 1). Of the 23 papers included, three papers were case–control studies, 13 papers were cohort studies, six papers were surveillance and prevalence studies, and one paper used data derived from a large randomized controlled trial database (Table 1).

The majority of studies were performed in high-income settings, of which two were based in the UK. Three studies were undertaken in the middle-income countries of Brazil, India and Egypt. Two studies involved paediatric populations, with the remaining 21 papers addressing adult populations, including patients from general wards, patients in intensive care units (ICUs), surgical patients, stroke patients, orthopaedic patients, hip fracture admissions, spinal injury patients and renal transplant patients.

The CASP scores for the three case–control studies ranged from nine to six, whilst the prevalence studies, using the adjusted CASP cohort study tool, scored between five and 11 out of 12. The cohort studies scored between three and 13 out of 14. Of the 23 papers included, almost 80% had a CASP value within the score’s upper quartile, being six and above for the case–control, prevalence and surveillance studies, and seven and above for the cohort studies.

Of the 143 risk factors assessed in all the studies, 96 (68%) were significant and 43 were independent risk factors for HCA UTI, with six studies not performing multivariate analyses. Of the independent risk factors, 29 were only assessed by one study. Among the remaining 14 risk factors that were assessed by multiple studies, half were found to be significant in more than one of the studies in which they were included (female sex, increasing age, diabetes mellitus, length of hospital stay, prior stroke, urinary catheter and duration of urinary catheter). Female sex was included in all the papers for analysis and was found to be significant in 65% of the studies. One-quarter of the independent risk factors were associated with urinary catheter use and management, including colonization of the drainage bag, having more than two catheters recorded and intermittent catheterization.

In order to assess the relative importance of the significant risk factors for HCA UTI, PAR% values were calculated for those factors where sufficient data were reported in the corresponding studies. Raw data (as described in the Methods) were extracted directly from 14 studies, while ORs were used to obtain raw data in four studies. Two studies did not publish sufficient data to enable PAR% calculations and three papers were not included as these were prevalence studies. Thus, 18 of the 23 papers contributed data to the PAR% calculations. PAR% values were calculated for 35 significant risk factors for HCA UTI, and 26 were also independent risk factors. The remaining 17 independent risk factors did not have sufficient data for PAR% calculations.

Figure 1. Systematic literature review schematic. The numbers represent the number of studies at each stage (e.g. 441 studies underwent abstract review).

<table>
<thead>
<tr>
<th>Criteria not met</th>
<th>Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>34–lacking statistics</td>
<td>23</td>
</tr>
<tr>
<td>60–inappropriate cases/controls</td>
<td>37–not original research</td>
</tr>
</tbody>
</table>
The local hospital data, of which PAR% values were calculated for eight (65%) of the 43 independent risk factors were located in ICHNT, of which PAR% values were calculated for 24. Twenty located within the administrative hospital data stored within (e.g. age, sex, prior surgery and ICU admission). The remaining factors were associated with inpatient admission PAS data associated with ICD-10 or OPCS codes, and 40% of the risk factors included age, female sex, diabetes mellitus, prior stroke, length of stay and urinary catheter.

Of the 35 risk factors for which PAR% values were calculated, functional independence (measured as the level of assistance required with daily living tasks) was found to have a non-significant PAR% (1.47, 95% CI -0.62 to 3.55), despite being identified as an independent predictor of HCA UTI. The remaining 34 significant risk factors were stratified into a hierarchy, with the risk factor with the highest PAR% (79.34%, 95% CI 78.18–80.50) being urinary catheterization (Table II). Five risk factors demonstrated protective properties from the calculated PAR%: creatinine level (<1 (–1.42%), neurological disease (–3.51%), surgical rather than medical admission (–39.22%), antibiotic therapy (–57.47%) and non-elective admission (–62.71%).

From the 96 significant risk factors for HCA UTI, 41% (39) were located within the administrative hospital data stored within ICHNT, of which PAR% values were calculated for 24. Twenty-eight (65%) of the 43 independent risk factors were located in the local hospital data, of which PAR% values were calculated for 18 (Figure 2). All the risk factors identified by multiple studies as independent predictors of HCA UTI (apart from duration of urinary catheterization) were found within the local hospital data. These risk factors included age, female sex, diabetes mellitus, prior stroke, length of stay and urinary catheter.

Thirty-seven percent of the independent risk factors were associated with ICD-10 or OPCS codes, and 40% of the risk factors were associated with inpatient admission PAS data (e.g. age, sex, prior surgery and ICU admission). The remaining variables were identified within the pathology codes (7%), surveillance data (7%) and merging codes to generate new variables (9%); for example, ‘prior surgery’ combines surgical codes and admission data. Of the top 10 risk factors from the PAR% hierarchy, 50% were located or associated with local codes (urinary catheter, prior stroke, meatal colonization, bladder dysfunction and length of stay). The remaining 50%, which were associated with urinary catheterization, excluding urinary catheterization itself, and the American Society of Anesthesiologists’ score, were not identified within the local databases. From all of the significant risk factors for which PAR% were calculated, 71% (24/34) were located within local hospital data (Table II).

### Discussion

The purpose of this study was to quantify the scientific evidence base on risk factors for HCA UTI, and to apply this to locally available administrative hospital data to establish the utility of this data source for further research activities in the development of surveillance tools, including risk stratification and the generation of automated syndromic HCA UTI algorithms.

The systematic literature review returned 23 relevant studies, with 96 significant and 43 independent HCA UTI risk factors identified out of the 143 risk factors that were
investigated by the papers reviewed. The significant risk factors were explored further along with the independent risk factors, as not all studies performed multivariate analysis and these factors may have shown significance on further analysis. Of the independent risk factors, 65% were located within the local administrative hospital databases that were available for this study. However, it must be noted that several limitations were encountered at different stages of this study. The lack of a second reviewer in this study was one such limitation, although inclusion criteria were discussed and a clear systematic methodology was followed to minimize any potential for reviewer bias.

Although the original literature search returned nearly 4109 articles, a surprisingly small percentage of these papers fulfilled the study criteria (0.6%). Only two studies were based in the UK.4,22 The two Canadian studies were the only ones to use linked as well as electronic data, both from local and regional systems.29,30 In the CASP tool evaluations, almost half of the included studies performed multivariate analysis and some were found to be relevant for other HCAI infection outcomes. In addition, some potential, but as yet uninvestigated, risk factors have been identified as relevant for surgical outcomes. The 'available in local data' indicates whether or not the risk factor has been located within the local administrative hospital data. A positive PAR% indicates that this risk factor is responsible for that percentage of infections, and a negative PAR% indicates that the presence of this risk factor prevents that percentage of infections.
site infections,\textsuperscript{42} was not researched specifically as a risk factor for HCA UTI. However, the potential for confounding was reduced by following a systematic approach to scoring the quality of papers and using strict inclusion criteria. Cohort papers scored worst for follow-up time and completeness of follow-up, with less than one-quarter of the papers being awarded a CASP point.

Different study settings, designs and populations, and diversity in local practices (such as antibiotic prescribing policies) may have masked the relevance and importance of certain risk variables. However, the finding that female sex, increasing age, diabetes mellitus, length of stay, urinary catheters and prior stroke are all significant risk factors for acquiring HCA UTI was anticipated, and the fact that these risk factors have been found to be independent predictors of HCA UTI suggests that the methodology was robust. However, some of the other independent and significant risk factors found to be strong risks or protectors in the PAR% calculations were contradictory or questionable. For example, non-elective and surgical admissions were found to be protective against HCA UTI, but surgical patients generally constitute elective admissions. In this case, the population for non-elective admissions was small with only one study contributing to the calculation, suggesting that this association may be specific to that particular study population; the paper itself questioned the finding.\textsuperscript{42} Similarly, neurological disease was found to be protective (–3.51), whereas having had a stroke was associated with the fifth highest PAR% (27.62). In this instance, the population for the neurological disease PAR% calculation was limited, and it is possible that local policies and patient case mix may have distorted the reported associations.

The methodology of PAR% calculations for stratifying HCA UTI risk factors relative to their importance in disease occurrence was appropriate in this context, and has been used previously in both case-control and cohort studies to assess risk factors of disease, and as a descriptive statistic for literature reviews.\textsuperscript{43–45} However, to the authors’ knowledge, this is the first application of PAR% to investigate risk factors for HCAI. Despite the fact that the associated PAR% was only calculated for 37% of the significant risk factors, PAR% was calculated for almost two-thirds of the independent risk factors, and the vast majority of the appropriate studies (90%) contributed data to the PAR% calculations. Further detail from the authors of the included studies would be required in order to produce a more complete assessment of risk importance based on population-attributable risks.

The novel use of this method in this field has shown its potential for providing a more detailed and robust interpretation of the existing literature, and the subsequent utility of administrative hospital data. Three-quarters of the independent risk factors, which had both a PAR% calculation and were located within the local administrative hospital data, had PAR% of more than 10% or less than −10%, indicating a strong potential for these risk factors to be used in further investigation of the use of existing administrative data.

The identification of specific risk factors, which have been preselected through an evidence-based approach, within local administrative databases may be challenging at times. This is due to certain limitations in the terms and codes employed in recording healthcare data. For example, there are no specific codes for risk factors such as ‘underlying debilitation’ and ‘indication for catheterization’, and thus these could not be extracted directly from the local data. However, these restrictions could be addressed through innovative use of alternative codes and terms as proxies for previous undetected variables. Conversely, risk factors might be recorded under a specific code or term within the local hospital databases but these might not necessarily be accurate. For example, the keyword search for ‘unconscious’ returned the ICD-10 code for ‘coma, unspecified’, which might not correspond to the clinical notes of a patient being unconscious at admission. Similarly, the risk factor of ‘meatal colonization’ does not necessarily relate to the code identified for a meatal swab in LIMS.

The quality of coding within the administrative data presents a challenge for the development of innovative surveillance tools. In the UK, clear guidelines offering instructions on
how to code conditions, and information on when they are to be recorded are provided by the NHS (available at: http://www.connectingforhealth.nhs.uk/codingclinic). The most recent guidelines specify a group of comorbidities to be mandatorily coded within administrative hospital data if there is any mention of them in the patient’s medical notes, including urinary retention, renal failure and diabetes. However, despite clear guidance, poor-quality coding is still reported. Issues arise from variability in coders and their experience, and the detail and accuracy of the information contained within the patient’s medical notes. In order to create surveillance tools using administrative data, these quality issues need to be taken into consideration, with an understanding of the potential for the under or over-recording of codes.

In the absence of a real-time medical record, the application of surveillance for HCA UTI based on predictors at an individual level is not plausible. The delay often encountered with administrative data means that data on risks, such as diabetes mellitus, are only available electronically once the patient has been discharged. Centres such as Salt Lake City, where electronic patient records have been implemented for 20 years, have shown the potential for this technology in aiding surveillance of HCAI. Certain hospital data, such as pathology and radiology information, are recorded electronically as they are processed, suggesting that real-time syndromic surveillance could be developed based on these data. This has been implemented successfully in the USA for catheter-associated bloodstream infections. However, the use of predictors to develop surveillance at the hospital population level using administrative data is possible with prospective applications. This includes implementing a risk stratified framework for HCA UTI surveillance, the creation of benchmarking tools for HCA UTI which account for local patient populations, and the opportunity to monitor interventions and control measures.

The use of different administrative hospital data systems to develop novel surveillance systems is still an emerging discipline, and with over 750 individual databases located thus far within ICHNT, further relevant information on risk factors is likely to be stored routinely within the hospital data. Electronic surveillance represents a cost-effective, time-efficient and robust approach to surveillance, although it is subjective to clinical opinion and a changing policy landscape. In the context of HCA UTI, this could potentially provide a system to complement traditional surveillance schemes and infection prevention and control practices across healthcare settings. Despite the methodology in this paper being based upon the administrative data found routinely within the UK NHS, the authors believe that the principle is applicable to any health system with electronic administrative data.

This study has demonstrated the potential of local administrative hospital data for the development of innovative surveillance for HCA UTI by applying the results from a systematic literature review to local administrative data using a novel methodology, despite the various limitations discussed. This work has provided the basis for a framework to further explore the benefits of using hospital data in the development of surveillance tools for HCA UTI. The next steps are to investigate the utility and quality of the codes identified by this review in local data, apply this to developing risk profiles and proxy indicators, and test the value of syndromic algorithms to enhance traditional surveillance and infection control approaches.

Conflict of interest statement
None declared.

Funding sources
The UK Clinical Research Collaboration funds the National Centre for Infection Prevention and Management at Imperial College London, and also supports the UK National Institute for Health Research Biomedical Research Centre funding scheme. The Dr Foster Unit at Imperial is largely funded by a research grant from Dr Foster Intelligence (an independent health service research organization) and is affiliated with the Centre for Patient Safety and Service Quality at Imperial College Healthcare NHS Trust, funded by the National Institute for Health Research.

Acknowledgements
The authors wish to thank the Infection Prevention and Control, Pathology and Microbiology departments and the Information Technology teams within Imperial College NHS Trust for their collaboration.

References