REVIEW ARTICLE
Systematic review of risk prediction scores for surgical site infection or periprosthetic joint infection following joint arthroplasty

S. K. KUNUTSOR*, M. R. WHITEHOUSE, A. W. BLOM and A. D. BESWICK

Musculoskeletal Research Unit, School of Clinical Sciences, University of Bristol, Learning & Research Building (Level 1), Southmead Hospital, Southmead Road, Bristol, BS10 5NB, UK

Received 30 November 2016; Final revision 2 February 2017; Accepted 5 February 2017

SUMMARY
Accurate identification of individuals at high risk of surgical site infections (SSIs) or periprosthetic joint infections (PJIs) influences clinical decisions and development of preventive strategies. We aimed to determine progress in the development and validation of risk prediction models for SSI or PJI using a systematic review. We searched for studies that have developed or validated a risk prediction tool for SSI or PJI following joint replacement in MEDLINE, EMBASE, Web of Science and Cochrane databases; trial registers and reference lists of studies up to September 2016. Nine studies describing 16 risk scores for SSI or PJI were identified. The number of component variables in a risk score ranged from 4 to 45. The C-index ranged from 0·56 to 0·74, with only three risk scores reporting a discriminative ability of >0·70. Five risk scores were validated internally. The National Healthcare Safety Network SSIs risk models for hip and knee arthroplasties (HPRO and KPRO) were the only scores to be externally validated. Except for HPRO which shows some promise for use in a clinical setting (based on predictive performance and external validation), none of the identified risk scores can be considered ready for use. Further research is urgently warranted within the field.

Key words: Bone infections, epidemiology, risk assessment.

INTRODUCTION
Surgical site infections (SSIs) which can be classified as superficial wound infections, deep wound infections, or periprosthetic joint infections (PJIs) [1], are uncommon but serious complications of total joint replacements [2, 3]. PJIs can result in severe pain, functional deficits and even death [4–6]; and their management is a huge financial burden to health care systems [7, 8]. With increasing life expectancy and a growing indication for primary joint replacements [9], there will be a proportionate rise in the number of patients who will be affected by PJIs. An approach to tackle the increasing incidence of PJIs is to identify those people at high risk and offer appropriate interventions. Early and accurate identification of individuals at high risk of PJI influences clinical decisions and development of targeted preventive strategies, and helps to optimise resources required for detection of PJI. Several factors such as characteristics of the patient, surgical procedure and postoperative care, have been found to influence the risk of developing PJI [10, 11], however their potential utility for PJI risk assessment remains uncertain.

A risk score or prognostic model is a statistical equation that predicts an individual’s disease risk based on a combination of the values of multiple
predictors or risk factors [12]. Risk prediction scores are ideally developed using data from long-term follow-up of large population-based cohorts of individuals without a history of the event of interest (SSI or PJI in this case) at baseline. The dataset is used to identify important predictors and the model equation is developed [13]. Using the derivation sample, the score’s apparent performance is evaluated in a process known as internal evaluation. The next stage is external validation, which examines the generalisability of the model using new data. Risk prediction scores first emerged in the area of cardiovascular disease (CVD) prevention and have been widely used globally in clinical and public health practice. Well known amongst them is the Framingham CVD risk score [14] (a risk score which assesses an individual’s risk of a cardiovascular event within 10 years), which is a commonly used algorithm in clinical practice and accepted tool in preventive medicine.

Prevention of SSIs or PJIs is a high policy priority and there has been an increasing interest in the development of risk prediction tools for SSI or PJI over the last decade. However, unlike the substantial progress made in CVD prevention using risk scores, the amount of progress made in the area of SSIs or PJIs is uncertain. There is therefore a need for objective data on the development of risk scores (including their component variables), their discriminative abilities, whether they have been externally validated, and whether their clinical effectiveness have been assessed in well-designed randomised controlled trials (RCTs). In this context, using systematic review methodology, we aimed to: (i) identify and summarise studies reporting the development of risk prediction scores for SSI or PJI; (ii) assess clinical variables selected for model inclusion and the predictive performance of these models; (iii) assess if identified models have been externally validated and their performances compared; (iv) assess if the impact or clinical effectiveness of these risk scores have been evaluated in appropriate RCTs and (v) finally to identify gaps in the existing evidence and whether further research is needed in the field.

METHODS

This review was conducted using a predefined protocol, which has been registered in the PROSPERO prospective register of systematic reviews (CRD42016042158), and in line with PRISMA guidelines [15] (Supplementary Material 1). We searched MEDLINE, EMBASE, Web of Science and the Cochrane Library electronic databases up to 30 September 2016. The publicly available trial registers ClinicalTrials.gov, UKCRN (UK Clinical Research Network) Study Portfolio Database, and the WHO International Clinical Trials Registry Platform were also searched. The search strategy combined free and MeSH search terms and combination of key words relating to risk prediction (e.g., ‘predict’, ‘risk score’, ‘sensitivity’), SSI or PJI (e.g., ‘periprosthetic joint infection’, ‘deep infection’, ‘surgical site infection’), and joint replacement (e.g., ‘hip replacement’, ‘knee replacement’, ‘hip arthroplasty’, ‘knee arthroplasty’). No restrictions were placed on publication dates and only articles published in English were considered. Reference lists of retrieved articles and relevant review articles identified on the topic were manually scanned for all relevant additional studies. Detailed description of all Materials and Methods, as well as the Literature Search Strategy are available in Supplementary Materials 2 and 3.

RESULTS

Study identification and selection

Figure 1 shows the flow of studies through the review. Our literature search strategy identified 1802 potentially relevant articles. After the initial screening of titles and abstracts, 15 articles remained for further evaluation. Following detailed evaluation which included full-text reviews, six articles were excluded because (i) they were studies of diagnostic scores (n = 2) and (ii) they were studies of risk scores for outcomes such as readmission, infection eradication and treatment outcome of PJI (n = 4). The remaining nine articles met the inclusion criteria and were included in the review [16–24].

Study characteristics and quality assessment

Table 1 summarises characteristics of the studies in the sample. Studies were published between 2006 and 2016, with all but one appearing in 2011–2016. One study was reported as a published conference abstract [16]. Overall, the studies involved 482 877 joint replacements, including 6968 SSIs or PJIs. For studies that reported age data, the baseline age of participants ranged from 56 to 81 years. The sample size of cohorts ranged from 217 to 172 055 and follow-up for infection outcomes ranged from 30 days to 2 years. For
the assessment of infections, the majority of the studies used Centre for Disease Control or Infectious Diseases Society of America criteria. Studies classified infection outcomes as SSI or PJI specifically. One study employed both SSI and PJI outcomes [21] and another study used PJI recurrence [24]. Quality assessment using PROBAST showed evidence of high overall risk of bias throughout the included studies. Five risk scores had unclear concern for overall applicability and only two scores were deemed to be usable in the targeted individuals and context (the National Healthcare Safety Network (NHSN) SSIs risk models for hip and knee arthroplasties (HPRO and KPRO)) [18] (Supplementary Material 4).

**Model description and development**

Table 2 provides details of risk scores included in eligible studies: their component predictors, statistical properties, measures of discrimination and/or calibration, and
Table 1. Summary characteristics of studies included in the review

<table>
<thead>
<tr>
<th>Lead author, publication date</th>
<th>Location</th>
<th>Baseline year</th>
<th>Study design</th>
<th>Population/sampling frame</th>
<th>Mean/median age</th>
<th>Name of risk tool</th>
<th>Specific outcome reported</th>
<th>Sample size</th>
<th>Number of events</th>
<th>Duration of follow-up</th>
<th>Ascertainment of outcome (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mu et al. [18]</td>
<td>USA</td>
<td>2006–2008</td>
<td>Retrospective cohort</td>
<td>NHSN data Total primary, partial primary, partial revision, total revision arthroplasty</td>
<td>NR</td>
<td>HPRO</td>
<td>SSI (superficial incisional, deep incisional, and organ/space) Deep incisional and organ/space SSIs</td>
<td>13 879</td>
<td>1855</td>
<td>30 days for superficial incisional and 1 year for deep incisional and organ/space infections</td>
<td>CDC definition</td>
</tr>
<tr>
<td>Mu et al. [18]</td>
<td>USA</td>
<td>2006–2008</td>
<td>Retrospective cohort</td>
<td>NHSN data Primary or revision arthroplasty</td>
<td>NR</td>
<td>KPRO</td>
<td>SSI (superficial incisional, deep incisional, and organ/space) Deep incisional and organ/space SSIs</td>
<td>172 055</td>
<td>1723</td>
<td>30 days for superficial incisional and 1 year for deep incisional and organ/space infections</td>
<td>CDC definition</td>
</tr>
<tr>
<td>Paxton et al. [16]</td>
<td>USA</td>
<td>2001–2009</td>
<td>Retrospective cohort</td>
<td>Kaiser Permanente’s Total Joint Replacement Registry. Patients who underwent total knee replacement</td>
<td>NR</td>
<td>NS</td>
<td>Deep infection</td>
<td>38 094</td>
<td>241</td>
<td>1 year</td>
<td>CDC definition</td>
</tr>
<tr>
<td>Berbari et al. [19]</td>
<td>USA</td>
<td>2001–2006</td>
<td>Prospective case-control</td>
<td>Patients who underwent THA or TKA</td>
<td>NR</td>
<td>Baseline Mayo PJI risk score</td>
<td>PJI</td>
<td>617</td>
<td>301</td>
<td>NR</td>
<td>CDC definition</td>
</tr>
<tr>
<td>Berbari et al. [19]</td>
<td>USA</td>
<td>2001–2006</td>
<td>Prospective case-control</td>
<td>Patients who underwent THA or TKA</td>
<td>NR</td>
<td>1-month-post-surgery Mayo PJI risk score</td>
<td>PJI</td>
<td>574</td>
<td>258</td>
<td>NR</td>
<td>CDC definition</td>
</tr>
<tr>
<td>Bozic et al. [20]</td>
<td>USA</td>
<td>1998–2009</td>
<td>Retrospective cohort</td>
<td>Medicare patients with primary THA (Administrative claims data) Patients with procedures performed at Rochester Mayo Clinic Primary or revision hip replacement</td>
<td>NR</td>
<td>NS</td>
<td>PJI</td>
<td>53 252</td>
<td>1102</td>
<td>2 years</td>
<td>NR</td>
</tr>
<tr>
<td>Lewallen et al. [21]</td>
<td>USA</td>
<td>2002–2009</td>
<td>Retrospective Cohort</td>
<td>Patients with procedures performed at Rochester Mayo Clinic Primary or revision hip replacement</td>
<td>65-6</td>
<td>HPRO</td>
<td>SSI and PJI</td>
<td>10 869</td>
<td>426*</td>
<td>1 year</td>
<td>Infectious Diseases Society of America criteria</td>
</tr>
<tr>
<td>Lewallen et al. [21]</td>
<td>USA</td>
<td>2002–2009</td>
<td>Retrospective Cohort</td>
<td>Patients with procedures performed at Rochester Mayo Clinic Primary or revision hip replacement</td>
<td>67-4</td>
<td>KPRO</td>
<td>SSI and PJI</td>
<td>11 072</td>
<td>426*</td>
<td>1 year</td>
<td>Infectious Diseases Society of America criteria</td>
</tr>
<tr>
<td>Lead author, publication date</td>
<td>Location</td>
<td>Baseline year</td>
<td>Study design</td>
<td>Population/sampling frame</td>
<td>Mean/median age</td>
<td>Name of risk tool</td>
<td>Specific outcome reported</td>
<td>Sample size</td>
<td>Number of events</td>
<td>Duration of follow-up</td>
<td>Ascertainment of outcome(s)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>Australia</td>
<td>2001–2012</td>
<td>Retrospective Cohort</td>
<td>Australian Department of Veterans’ Affairs database. Primary unilateral total hip replacement</td>
<td>80.9</td>
<td>RxRisk-V; Elixhauser; and Charlson comorbidities coding algorithm</td>
<td>PJI after THA</td>
<td>11 848</td>
<td>364</td>
<td>90 days</td>
<td>ICD-10-AM diagnostic codes T81-4, T84-5, T85-79 or hospitalisations with the procedure ICD-10-AM procedure codes 930 301 (lavage of hips); or the initiation of the antibiotics gentamicin (ATC code J01GB03) or vancomycin (ATC code J01XA01).</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>Australia</td>
<td>2001–2012</td>
<td>Retrospective Cohort</td>
<td>Australian Department of Veterans’ Affairs database. Primary total knee replacement</td>
<td>79.8</td>
<td>RxRisk-V; Elixhauser; and Charlson comorbidities coding algorithm</td>
<td>PJI after TKA</td>
<td>18 972</td>
<td>648</td>
<td>90 days</td>
<td>ICD-10-AM diagnostic codes T81-4, T84-5, T85-79 or hospitalisations with the procedure ICD-10-AM procedure codes 4950030 (lavage of knees); or the initiation of the antibiotics gentamicin (ATC code J01GB03) or vancomycin (ATC code J01XA01).</td>
</tr>
<tr>
<td>Maradit Kremers et al. [23]</td>
<td>USA</td>
<td>2002–2009</td>
<td>Retrospective cohort</td>
<td>Patients who underwent primary or revision THA at Rochester Mayo Clinic</td>
<td>64.6</td>
<td>Claims-based risk model for THA</td>
<td>SSI after THA</td>
<td>9720</td>
<td>192*</td>
<td>1 year</td>
<td>Infectious Diseases Society of America Criteria</td>
</tr>
<tr>
<td>Maradit Kremers et al. [23]</td>
<td>USA</td>
<td>2002–2009</td>
<td>Retrospective cohort</td>
<td>Patients who underwent primary or revision TKA at Rochester Mayo Clinic</td>
<td>67.7</td>
<td>Claims-based risk model for TKA</td>
<td>SSI after TKA</td>
<td>10 451</td>
<td>192*</td>
<td>1 year</td>
<td>Infectious Diseases Society of America Criteria</td>
</tr>
<tr>
<td>Tikhilov et al. [24]</td>
<td>Russia</td>
<td>2008–2012</td>
<td>Retrospective cohort</td>
<td>Patients treated for PJI of the hip</td>
<td>56.1</td>
<td>NS</td>
<td>PJI recurrence</td>
<td>217</td>
<td>78</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

ATC, Anatomic, Therapeutic and Chemical Classification; CDC, Centre for Disease Control; HPRO, National Healthcare Safety Network surgical site infections risk model for hip arthroplasty; KPRO, National Healthcare Safety Network surgical site infections risk model for knee arthroplasty; NHSN, National Healthcare Safety Network; NR, not reported; NS, not stated; PJI, periprosthetic joint infection; SSI, surgical site infections; THA, total hip arthroplasty; TKA, total knee arthroplasty; ICD-10-AM, International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification.

* Indicates the total number of SSIs for both THA and TKA.
Table 2. Key characteristics of risk prediction scores for PJI included in the review

<table>
<thead>
<tr>
<th>Lead author, publication date</th>
<th>Name of risk tool</th>
<th>Statistical model</th>
<th>Predictors used</th>
<th>Number of predictors</th>
<th>Discrimination (C-index)</th>
<th>Calibration (HL goodness-of-fit test)</th>
<th>Internal validation</th>
<th>External validation</th>
<th>Performance comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geubbels et al. [17]</td>
<td>THA-specific risk model for SSI</td>
<td>Logistic regression</td>
<td>Age, duration of preoperative hospital stay, postdischarge surveillance and number of discharge diagnoses</td>
<td>4</td>
<td>0.64</td>
<td>Satisfactory goodness of fit reported</td>
<td>Cross-validation C-index = 0.62</td>
<td>None</td>
<td>Compared with the NNIS index (C-index = 0.56; P &lt; 0.001)</td>
</tr>
<tr>
<td>Mu et al. [18]</td>
<td>HPRO</td>
<td>Logistic regression</td>
<td>Age, anaesthesia, ASA, procedure duration, type of surgery (total primary, partial primary, partial revision, total revision), bed size and trauma</td>
<td>7</td>
<td>0.66 for SSI (superficial incisional, deep incisional and organ/space) 0.67 for deep incisional and organ/space SSIs</td>
<td>NR</td>
<td>Bootstrapping resampling</td>
<td>Externally validated in Lewallen et al. [21]</td>
<td>Compared with traditional NHSN risk index (C-index = 0.61) P-value for comparison of the two models &lt;0.0001</td>
</tr>
<tr>
<td>Mu et al. [18]</td>
<td>KPRO</td>
<td>Logistic regression</td>
<td>Age, anaesthesia, ASA, procedure duration, gender, type of surgery (revision vs. primary), bed size and trauma</td>
<td>8</td>
<td>0.64 for SSI (superficial incisional, deep incisional and organ/space) 0.65 for deep incisional and organ/space SSIs</td>
<td>NR</td>
<td>Bootstrapping resampling</td>
<td>Externally validated in Lewallen et al. [21]</td>
<td>Compared with traditional NHSN risk index (C-index = 0.60) P-value for comparison of the two models &lt;0.0001</td>
</tr>
<tr>
<td>Paxton et al. [16]</td>
<td>NS</td>
<td>Cox regression</td>
<td>Age, sex, race, indication for knee replacement, diabetes and its complications, and BMI</td>
<td>6</td>
<td>NR</td>
<td>P = 0.80</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Berbari et al. [19]</td>
<td>Baseline Mayo PJI risk score</td>
<td>Logistic regression</td>
<td>BMI, prior other operation on the index joint, prior arthroplasty, immunosuppression, ASA score and procedure duration</td>
<td>6</td>
<td>Original: 0.722 Bias-corrected: 0.690</td>
<td>Satisfactory model calibration</td>
<td>None</td>
<td>None</td>
<td>Compared with traditional NHSN risk index (C-index = 0.638; P &lt; 0.001)</td>
</tr>
<tr>
<td>Berbari et al. [19]</td>
<td>1-month-postsurgery Mayo PJI risk score</td>
<td>Logistic regression</td>
<td>BMI, prior other operation on the index joint, prior arthroplasty, immunosuppression, ASA score and procedure duration and postoperative wound drainage</td>
<td>7</td>
<td>Original: 0.716 Bias-corrected: 0.680</td>
<td>NR</td>
<td>None</td>
<td>None</td>
<td>Compared with traditional NHSN risk index (C-index = 0.633; P &lt; 0.001)</td>
</tr>
<tr>
<td>Bozic et al. [20]</td>
<td>NS</td>
<td>Logistic regression</td>
<td>29 comorbid conditions, age, sex and socioeconomic status</td>
<td>32</td>
<td>NR</td>
<td>NR</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Lewallen et al. [21]</td>
<td>HPRO</td>
<td>Logistic regression</td>
<td>Age, anaesthesia, ASA, procedure duration, type of surgery (total primary, partial primary, partial revision, total revision), bed size and trauma</td>
<td>7</td>
<td>0.695 for SSI 0.737 for PJI</td>
<td>P = 0.323 for SSI P = 0.606 for PJI</td>
<td>N/A</td>
<td>N/A</td>
<td>Modest improvement in discrimination on addition of morbid obesity and diabetes mellitus to the model C-index = 0.706 for SSI C-index = 0.746 for PJI</td>
</tr>
</tbody>
</table>

6 S. K. Kunutsor and others
<table>
<thead>
<tr>
<th>Lead author, publication date</th>
<th>Name of risk tool</th>
<th>Statistical model</th>
<th>Predictors used</th>
<th>Number of predictors</th>
<th>Discrimination (C-index)</th>
<th>Calibration (HL goodness-of-fit test)</th>
<th>Internal validation</th>
<th>External validation</th>
<th>Performance comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lewallen et al.† [21]</td>
<td>KPRO</td>
<td>Logistic regression</td>
<td>Age, anaesthesia, ASA, procedure duration, gender, type of surgery (revision vs. primary), bed size and trauma</td>
<td>8</td>
<td>0.592 for SSI 0.645 for PJI</td>
<td>$P = 0.121$ for SSI $P = 0.072$ for PJI</td>
<td>N/A</td>
<td>N/A</td>
<td>Modest improvement in discrimination on addition of morbid obesity and diabetes mellitus to the model C-index = 0.623 for SSI C-index = 0.669 for PJI</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>RxRisk-V for THA</td>
<td>Logistic regression</td>
<td>42 comorbid conditions plus age, gender and primary diagnoses for surgery</td>
<td>45</td>
<td>0.60</td>
<td>0.793</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>Elixhauser for THA</td>
<td>Logistic regression</td>
<td>30 comorbid conditions plus age, gender and primary diagnoses for surgery</td>
<td>33</td>
<td>0.59</td>
<td>0.744</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>Charlson for THA</td>
<td>Logistic regression</td>
<td>17 comorbid conditions plus age, gender and primary diagnoses for surgery</td>
<td>20</td>
<td>0.58</td>
<td>0.905</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>RxRisk-V for TKA</td>
<td>Logistic regression</td>
<td>42 comorbid conditions plus age, gender and primary diagnoses for surgery</td>
<td>45</td>
<td>0.57</td>
<td>0.057</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>Elixhauser for TKA</td>
<td>Logistic regression</td>
<td>30 comorbid conditions plus age, gender and primary diagnoses for surgery</td>
<td>33</td>
<td>0.58</td>
<td>0.827</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Inacio et al. [22]</td>
<td>Charlson for TKA</td>
<td>Logistic regression</td>
<td>17 comorbid conditions plus age, gender and primary diagnoses for surgery</td>
<td>20</td>
<td>0.56</td>
<td>0.513</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Maradit Kremers et al. [23]</td>
<td>Claims-based risk model for THA</td>
<td>Cox regression</td>
<td>Age, sex, type of surgery (primary vs. revision), and 16 individual Charlson index comorbidities</td>
<td>19</td>
<td>Original: 0.662 Bias-corrected: 0.629</td>
<td>Reported as 'reasonably calibrated'</td>
<td>Bootstrap resampling</td>
<td>None</td>
<td>On addition of four clinical predictors (morbid obesity, prior surgeries on the same joint, ASA score and length of operative time) Original C-index: 0.706 Bias-corrected C-index: 0.665 Difference in C statistic: 0.043 ($0.012$-$0.074$) Aggregated IDE: 0.37% ($0.12$-$0.62$%)</td>
</tr>
<tr>
<td>Maradit Kremers et al. [23]</td>
<td>Claims-based risk model for TKA</td>
<td>Cox regression</td>
<td>Age, sex, type of surgery (primary vs. revision), and 16 individual Charlson index comorbidities</td>
<td>19</td>
<td>Original: 0.621 Bias-corrected: 0.585</td>
<td>Reported as 'reasonably calibrated'</td>
<td>Bootstrap resampling</td>
<td>None</td>
<td>On addition of four clinical predictors (morbid obesity, prior surgeries on the same joint, ASA score, and length of operative time) Original C-index: 0.648 Bias-corrected C-index: 0.606 Difference in C statistic: 0.027 ($0.007$-$0.047$) Aggregated IDE: 0.69% ($-0.02$ to 0.21%)</td>
</tr>
</tbody>
</table>
reports of any validation and performance comparisons made. A total of 16 risk scores were described in the nine eligible studies. Five of these scores had separate models for hip and knee replacement patients [18, 22, 23]. Four studies described the development of two or more risk scores [18, 19, 22, 23]. All 16 risk scores were derivations of risk models on a base population and two of them were also externally validated on new populations [21]. Except for one study that developed the risk score based on a cohort recruited prospectively for the surveillance of SSIs [17], all studies used datasets retrospectively that had been established for different purposes. Except for the scores that were developed in both knee and hip replacement patients, the component predictors varied from score to score. However, age, sex and type of primary surgery featured in the majority of risk scores. Except for one score that was mainly based on invasive data such as ESR (erythrocyte sedimentation rate), CRP (C-reactive protein) and microbial aetiology [24], all scores were based on data that can be assessed non-invasively such as demographics, anthropometrics, medical and surgical histories, and surgical procedures. The number of component variables in a single score ranged from 4 to 45 (n = 16, median 19, interquartile range 6·5–32·5).

Seven out of the 16 risk scores had 10 or fewer components. Of the 16 risk scores, 15 used regression techniques (logistic or Cox) to develop the score and one used a classification tree [24].

Model diagnostics

Except for three studies (comprising of three risk scores) [16, 20, 24], the C-statistic was reported for 13 risk scores. The C-index ranged from 0·56 to 0·74. Only three risk scores were reported to have a discriminative ability of >0·70 and these were the baseline Mayo and 1-month-postsurgery Mayo PJI risk scores as reported by Berbari et al. [19] and HPRO which was externally validated by Lewallen et al. [21]. Calibration measures were presented for 11 risk scores (including the baseline Mayo PJI risk score) and each was reported to have satisfactory model calibration. Two studies did not report on any measures of discrimination or calibration [20, 24] (Table 2).

Model validation

Only five of the risk scores were validated internally using resampling techniques such as bootstrapping and cross-validation [17, 18, 23].
Risk prediction scores for SSI or PJI

9

Performance comparisons

The performances of five risk scores were compared with existing models in three studies [17–19]. Geubbels et al. compared the predictive performance of their newly developed THA-specific risk score for SSI with the NNIS (National Nosocomial Infection Surveillance) system risk index (which incorporates three risk factors of equal weight namely wound contamination class, American Society of Anesthesiologists (ASA) score, and duration of surgery), and reported better predictive performance for the new risk score (C-index: 0.64 vs. 0.56; \( P < 0.001 \)) [17]. Mu et al. also reported statistically significantly better performances for the HPRO and KPRO risk scores when compared with the traditional NHSN SSI risk model, though the C-statistics were generally low (<0.70) [18]. The baseline Mayo and 1-month-postsurgery Mayo PJI risk scores also performed well compared with the traditional NHSN SSI risk score (C-index: 0.72 vs. 0.64; \( P < 0.001 \)) and (C-index: 0.72 vs. 0.63; \( P < 0.001 \)), respectively [19]. Two studies assessed the incremental prognostic value of adding additional risk factors to their existing models [21, 23]. Lewallen et al. externally validated the HPRO and KPRO risk scores and reported that addition of information on morbid obesity and diabetes mellitus to each score modestly improved discrimination [21]. On addition of four clinical risk factors (morbid obesity, prior non-arthroplasties on the same joint, ASA score and operative time) to their claims-based risk models for THA and TKA, Maradit Kremers et al. reported improved performance (by C-statistics) for both models, though the THA model showed better performance than the TKA model [23]. There was however no noticeable improvement in calibration for both models. Finally, while there was an improvement in IDI (Integrated Discrimination Index) for the THA score, no significant improvement was seen for the TKA score: 0.37% (0.12% to 0.62%) and 0.09% (−0.02% to 0.21%), respectively.

Clinical evaluation of risk scores

None of the studies described the evaluation of the clinical effectiveness of a score in an intervention study or as part of an impact study aimed at changing patient outcomes.

DISCUSSION

Key findings

Using systematic review methods, we have reported the first overview of available risk assessment scores for SSI or PJI following joint replacement. Based on established quality criteria for risk scores [25, 26], none of the risk scores in our review were judged to be promising for use in clinical settings or public health practice, except for the HPRO. The HPRO is a procedure-specific risk score which was adapted from the traditional NHSN risk index using NHSN data and its purpose is for predicting SSI or PJI within 1 year of hip replacement [18]. The HPRO was found to perform better than the traditional NHSN risk index and external validation in an independent cohort showed high discriminative ability [21]. The HPRO also showed higher accuracy for predicting PJI compared with SSI. The data also show that risk prediction models for SSI or PJI have only been developed over the past 5 years. Of the 16 risk scores identified, only seven had 10 or fewer components included in the final score, with a number of scores having between 30 and 45 components. Although all 11 risk scores reporting calibration measures exhibited satisfactory calibration, only three of these risk scores were reported to have a discriminative ability of >0.70. Of all 16 risk scores, HPRO and KPRO were the only risk scores externally validated in an independent population. Quality assessment of the risk scores’ development and validation criteria showed all scores to have a high risk of bias. This was mainly due to the methodology used in assessment of predictors and outcomes, inappropriate handling of missing data, and lack of external validation.

Explanations and implications of findings

Our findings highlight the limited evidence available on appropriate risk scores for predicting SSI or PJI after joint replacement. Given the absence of an
ideal risk score which can be used in a routine clinical setting, it appears that the potential value of risk scores in preventing SSI or PJI may have been underestimated in orthopaedic practice. The findings also highlight the use of poor methodology in the development of some of these risk scores. Although cross-sectional study designs were not included, the included studies were not free from bias and confounding. The majority of the designs were based on retrospective cohorts instead of prospective cohort designs, which are ideal for risk score modelling as predictor information can be ascertained blindly in relation to the outcome or disease [13]. None of the risk scores was developed in a cohort recruited for this sole purpose, which introduced an inherent selection bias. A key methodological issue was the absence of clear and detailed reporting of the treatment of missing data in all studies, which is of utmost importance prior to the development of risk scores [12]. Included studies used complete case analysis in the presence of missing data, which does not represent the entire population and reduces the sample size [13]. It has been shown that risk scores that use multiple imputation, produce more valid results and have better discrimination than tools that ignore such additional analyses [27]. There were also concerns with usability of the risk scores, as the majority of the risk scores had more than 10 variables. It is recognised that the simplicity of the model is an important criteria for developing clinically useful risk scores [28, 29]. Evidence suggests that complex models are more likely to provide overoptimistic predictions, especially when extensive variable selection has been performed [30]. Only five of the risk scores were validated internally using resampling methods, which are techniques which give a good indication of how optimistic the risk score may be [31]. Although internal validation is helpful, it cannot provide information on the model’s performance elsewhere or its generalisability. Before a risk prediction tool can be used in clinical practice or in real-world settings, evaluation of its generalisability (or transportability) requires data from elsewhere – also known as external validation [12]. However, only two risk scores were externally validated in our sample [21]. Finally, none of the risk scores was reported to have been used in an impact study aimed at changing patient outcomes. Before a risk score can be implemented, a vital criterion that needs to be fulfilled is its impact on clinical practice [12]. Among the identified risk scores, only the HPRO was found to be potentially promising for use in a clinical setting. However, it cannot be considered ready for use as its clinical effectiveness is still yet to be evaluated. The unavailability of appropriate existing risk scores for use in the clinical setting is extremely concerning. To add to this challenge is the lack of established uniform criteria for the diagnosis of infection especially PJI, which actually makes it difficult to conduct diagnostic or risk prediction studies for infection. Although hip and knee replacements are successful elective procedures, with SSIs or PJIs being rare complications of these procedures [3, 32]; the incidence of these infections will increase in conjunction with growing healthcare burden due to osteoarthritis [33] and a predicted large rise in the numbers of arthroplasty procedures [34, 35]. To meet this challenge, there should be a clinical drive towards identification of individuals at high risk of SSIs or PJIs using risk prediction engines. The current findings should stimulate research groups to develop and evaluate appropriate infection outcome-specific risk prediction algorithms using robust methodology. The clinical effectiveness of the HPRO also needs to be evaluated before it is implemented. Within our 5 year INFORM (INFection ORthopaedic Management) Programme, the aim is to develop and establish optimum strategies for the prevention and treatment of PJIs within the UK National Health System [36], and which may include the development of appropriate risk prediction engines when the data allows.

Study strengths and limitations
To the best of our knowledge, this is the first systematic review to identify limited progress in the development and validation of risk prediction models for SSI or PJI following joint replacement, using robust systematic methodology. It is also the first review to assess the validity of existing risk scores based on risk of bias and applicability. Our search strategy was comprehensive and spanned multiple databases, making it unlikely that any relevant study was missed. There was variation in the definition of SSIs in the included studies, which did not allow for a head-to-head comparison of risk scores across studies. We were unable to harmonise data from contributing studies to perform a quantitative analysis, due to the heterogeneity in study designs and populations, predictors used, model types, and measures reported. Even though we tried to present the data as robustly as possible using established criteria, our conclusions might be limited due to the
quality of published research and the large variability across study characteristics and methodologies.

CONCLUSION

In conclusion, available risk scores to predict SSI or PJI have been developed using poor methodology and have several limitations. The majority of these risk scores have not been externally validated and are not ideal for use in clinical settings. The HPRO is the only risk prediction tool identified to show some promise for use in a clinical setting (based on its predictive performance and having some external validation); however, it needs further validation using new data and its clinical effectiveness should be evaluated using a RCT design. A potentially effective way of tackling the increasing incidence of SSIs is early and accurate identification of individuals at high risk using established risk prediction scores, an approach which has been very effective in the area of CVD prevention. Further research is urgently warranted within the field to develop and test appropriate outcome-specific risk prediction tools.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0950268817000486

ACKNOWLEDGEMENTS

This publication is part of the INFection ORthopaedic Management (INFORM) Programme. As such it benefits from involvement of the whole INFORM team. The INFORM team includes: Simon Strange, Setor Kunutsor, Kirsty Garfield, Erik Lenguerrand, Rachael Gooberman-Hill, Drew Moore, Amanda Burston, Joanne Simon, Garry King, Michael Whitehouse, Vikki Wylde, Andrew Beswick, Ashley Blom, Sian Noble, Athene Lane, Fran Carroll (Musculoskeletal Research Unit, School of Clinical Sciences, University of Bristol, Southmead Hospital, Southmead Road, Bristol, BS10 5NB, UK); Sian Noble, Athene Lane, Fran Carroll (School of Social and Community Medicine, University of Bristol, Bristol, BS8 2PS, UK); Jason Webb, Alasdair MacGowan (North Bristol NHS Trust, Southmead Hospital, Bristol, BS10 5NB, UK); Stephen Jones (Cardiff and Vale University Health Board, Longcross Street, Cardiff, CF24 0SZ, UK); Adrian Taylor (Oxford University Hospitals NHS Trust, John Radcliffe Hospital, Oxford OX3 9DU, UK); Paul Dieppe (University of Exeter, Medical School, Exeter, EX1 2LU, UK); Andrew Toms, Matthew Wilson (Royal Devon and Exeter NHS Foundation Trust, Newcourt House, Exeter, EX2 7JU, UK); Ian Stockley (Sheffield Teaching Hospitals NHS Trust, Northern General Hospital, Sheffield, S5 7AU, UK); Ben Burston, John-Paul Whittaker (The Robert Jones and Agnes Hunt Orthopaedic Hospital NHS Foundation Trust, Oswestry, Shropshire, SY10 7AG, UK); Tim Board (Wrightington, Wigan and Leigh NHS Foundation Trust, Apple Bridge, Wigan, Lancashire, WN6 9EP, UK); and all the surgeons and nurses from the collaborating centres. This article presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research program (RP-PG-1210-12005). The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

DECLARATION OF INTEREST

M.R.W. reports grants from the National Institute of Health Research during the conduct of the study; grants from British Orthopaedic Association, which were outside the submitted work.

REFERENCES


