



Predicting periprosthetic joint infection: external validation of preoperative prediction models

Seung-Jae Yoon¹, Paul C. Jutte¹, Alex Soriano², Ricardo Sousa³, Wierd P. Zijlstra⁴, and Marjan Wouthuyzen-Bakker⁵

¹Department of Orthopaedic Surgery, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

²Infectious Diseases Service, Clínic Barcelona, University of Barcelona, Barcelona, Spain

³Porto Bone Infection Group (GRIP), Orthopaedic Department, Centro Hospitalar Universitário do Porto, Porto, Portugal

⁴Department of Orthopaedic Surgery, Medical Center Leeuwarden, Leeuwarden, the Netherlands

⁵Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Correspondence: Marjan Wouthuyzen-Bakker (m.wouthuyzen-bakker@umcg.nl)

Received: 8 June 2024 – Revised: 21 August 2024 – Accepted: 29 August 2024 – Published: 25 October 2024

Abstract. Introduction: Prediction models for periprosthetic joint infections (PJIs) are gaining interest due to their potential to improve clinical decision-making. However, their external validity across various settings remains uncertain. This study aimed to externally validate promising preoperative PJI prediction models in a recent multinational European cohort.

Methods: Three preoperative PJI prediction models – by Tan et al. (2018), Del Toro et al. (2019), and Bülow et al. (2022) – that have previously demonstrated high levels of accuracy were selected for validation. A retrospective observational analysis of patients undergoing total hip arthroplasty (THA) and total knee arthroplasty (TKA) at centers in the Netherlands, Portugal, and Spain between January 2020 and December 2021 was conducted. Patient characteristics were compared between our cohort and those used to develop the models. Performance was assessed through discrimination and calibration.

Results: The study included 2684 patients, 60 of whom developed a PJI (2.2 %). Our cohort differed from the models' original cohorts with respect to demographic variables, procedural variables, and comorbidity prevalence. The overall accuracies of the models, measured with the *c* statistic, were 0.72, 0.69, and 0.72 for the Tan, Del Toro, and Bülow models, respectively. Calibration was reasonable, but the PJI risk estimates were most accurate for predicted infection risks below 3 %–4 %. The Tan model overestimated PJI risk above 4 %, whereas the Del Toro model underestimated PJI risk above 3 %.

Conclusions: The Tan, Del Toro, and Bülow PJI prediction models were externally validated in this multinational cohort, demonstrating potential for clinical application in identifying high-risk patients and enhancing preoperative counseling and prevention strategies.

1 Introduction

Periprosthetic joint infection (PJI) is a devastating complication of total hip arthroplasty (THA) and total knee arthroplasty (TKA). Treatment of a PJI involves multiple operations and prolonged antibiotic therapy, which are associated with reduced quality of life and high healthcare costs (Cahill et al., 2008; Hackett et al., 2015; Kurtz et al., 2008). With the increasing demand for arthroplasties in aging populations and PJIs being the leading cause of revision arthroplasties, the burden that PJIs place on healthcare systems is expected to rise (Premkumar et al., 2021; Dale et al., 2012; Bourne et al., 2004).

Improving and tailoring the prevention of PJIs can be facilitated by knowing individual patients' risks. There has been a growing interest in prediction models for PJI, which have the potential to improve patient counseling and clinical decision-making. However, most published PJI prediction models have not undergone external validation or have only been externally validated in one additional healthcare setting, typically in proximity to the institution at which they were developed (Kunutsor et al., 2017; Sweerts et al., 2023; Tan et al., 2018; Bülow et al., 2022). Considerable heterogeneity exists with respect to factors influencing the occurrence of PJI across institutions and countries, including patient characteristics, infection prevention practices, diagnostic criteria for PJI, and definitions of predictors (Gromov et al., 2014; Franklin et al., 2017; Paxton et al., 2019). All of these factors may cause prediction models to be inaccurate and potentially harmful in certain settings (Van Calster et al., 2023). Moreover, changes in preventive strategies over time can diminish the accuracy of models developed from outdated patient cohorts. As such, no prediction model can truly be considered externally valid unless it has been tested across multiple regions and over time (Van Calster et al., 2023).

Therefore, the aim of this study was to externally validate the most promising preoperative PJI prediction models using a recent European patient cohort undergoing THA or TKA.

2 Materials and methods

This external validation study was conducted and reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement (Collins et al., 2015).

2.1 Identification of prediction models

We conducted a literature search to identify all preoperative PJI prediction models for THA or TKA reported until December 2022. Of the retrieved models, those that had previously demonstrated high levels of accuracy (discrimination or calibration) were selected for validation. Models were excluded if they included predictors that were not obtainable prior to surgery, predicted the risk of surgical site infection

(SSI) or recurrent PJI, demonstrated poor external validity, or did not report performance metrics. Our search revealed three promising models developed by Tan et al. (2018) (hereafter referred to as "Tan"), Del Toro et al. (2019) (hereafter referred to as "Del Toro"), and Bülow et al. (2022) (hereafter referred to as "Bülow"), respectively. The Tan and Del Toro models were developed for both THAs and TKAs, whereas the Bülow model is only intended for primary THAs. These models showed good predictive performance in the United States (US), Spain, and Sweden, respectively. Moreover, the Tan and Bülow models were externally valid in respective cohorts from the US and Denmark. These three models were therefore regarded as having high clinical potential.

2.2 Study design and participants

We performed a multicenter retrospective study at four secondary- and tertiary-care hospitals in the Netherlands, Portugal, and Spain. Institutional review board approval was obtained at each hospital. All adult patients who underwent primary or aseptic revision THA or TKA in the period from January 2020 to December 2021, with a follow-up of at least 1 year, were included. The cohort with the diagnosis of aseptic revision were selected based on the postoperative diagnosis. Patients who were diagnosed with a PJI during revision surgery were excluded. As the Bülow model was developed on patients receiving primary THA, it was validated in a subset of our study population comprised solely of such patients.

2.3 Data collection

Patient charts from electronic health records were manually reviewed to collect demographic data, including age, sex, and body mass index (BMI). Procedural variables were also collected, including the affected joint, number of prior surgeries on the joint, type of arthroplasty (primary vs. revision), duration of surgery, anesthesia (regional vs. general), and principal diagnosis for arthroplasty. All comorbidities used by the Tan, Del Toro, and Bülow models were collected (Fig. 1). Comorbidities in the patient charts were assessed using the Elixhauser and Charlson comorbidity indices (Elixhauser et al., 1998; Charlson et al., 1987). If a PJI developed during follow-up, this was recorded as an event. A PJI was defined according to the European Bone and Joint Infection Society (EBJIS) criteria (McNally et al., 2021).

2.4 Missing data

Although the original models were developed via complete case analysis, we did not exclude patients with missing data, as this can lead to reduced power and biased estimates (Harell et al., 1996). After ascertaining that the missing data pattern was consistent with a missing-at-random assumption, multiple imputation was performed using chained equations (White et al., 2011; van Buuren and Groothuis-Oudshoorn,

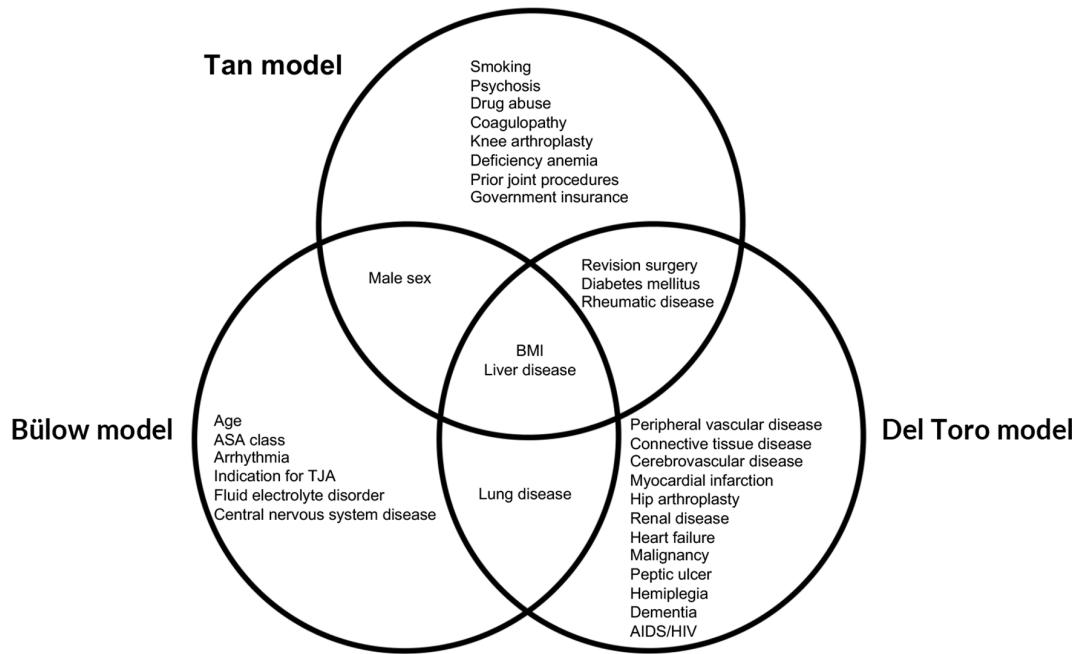


Figure 1. Venn diagram of the predictors included in the Tan, Del Toro, and Bülow models.

2011). A total of 20 imputed datasets were generated for this procedure. Predicted risks and performance measures were estimated in each of the 20 datasets and pooled using Rubin's rules.

2.5 Model validation

To determine the extent to which our cohort differed from the populations in which the models were previously studied, we performed χ^2 tests of homogeneity using patient characteristics. Statistical significance was considered when the p value was less than 0.05.

Model performance was assessed in concordance with a suggested framework for appraising prediction models (Steyerberg et al., 2010): we evaluated discrimination through the c statistic and calibration through the calibration plot, intercept, and slope. The c statistic, also known as the area under the receiver operating characteristic curve (ROC), measures a model's ability to distinguish between patients with an event and patients without an event. The score ranges from 0.5 to 1.0, with scores closer to 1.0 indicating better discrimination (Royston and Altman, 2010). Calibration refers to how closely the predicted risks match the observed rates of the event. This can be visualized through a calibration plot, in which a perfect model has a slope of 1 and an intercept of 0 (Van Calster et al., 2016).

The Tan model was modified to exclude the "government insurance" variable from calculations, as receiving government insurance is an indicator of disease burden, age, or socioeconomic status in the US but not in countries with universal health coverage.

All analyses were performed in R version 4.2.1 (R Foundation for Statistical Computing) using the CalibrationCurves (Van Calster et al., 2016; De Cock et al., 2023), dplyr (Wickham et al., 2023), and mice packages (van Buuren and Groothuis-Oudshoorn, 2011).

3 Results

3.1 Patient characteristics

We included a total of 2684 patients in our cohort, 1528 of whom received primary THA. PJI was observed in 60 (2.2%) patients in the entire cohort and in 33 (2.2%) patients undergoing primary THA. The characteristics of our entire cohort and the primary THA subgroup (second and third columns) are presented alongside the characteristics of the cohorts used to develop the models (fourth to sixth columns) in Table 1.

Our validation cohort for the Tan model had a significantly lower PJI rate (2.2% vs. 3.7%, $p < 0.001$), a higher proportion of female patients (61.9% vs. 55.8%, $p < 0.001$), and a higher rate of hip arthroplasty (64.3% vs. 53.0%, $p < 0.001$) compared with the derivation cohort. Comparison of patient characteristics was limited to the published data.

Our Del Toro validation cohort had significantly more hip arthroplasties (64.3% vs. 40.7%, $p < 0.001$), longer surgery durations ($p < 0.001$), and a higher prevalence of liver disease (2.3% vs. 0.7%, $p < 0.001$) compared with the derivation cohort. On the other hand, fewer female patients (61.9% vs. 68.9%, $p < 0.001$) and cases of diabetes mellitus (13.8% vs. 23.0%, $p < 0.001$) were observed. The PJI rate was slightly

Table 1. Comparison of the patient characteristics between the validation and derivation cohorts.

Patient characteristic	Validation cohort		Derivation cohorts			<i>p</i> value*		
	All patients	Primary THA	Tan	Del Toro	Bülow	Tan	Del Toro	Bülow
Number of patients	2684	1528	27717	2324	88 830			
Inclusion period	2020–2021	2020–2021	2000–2014	2013–2015	2008–2014			
Demographic								
Age (years)	70 (62, 77) ^a	70 (61, 77) ^a	64 ± 12 ^b	71 (64, 78) ^a	70–80 ^c			
Female sex	1661 (61.9)	899 (58.8)	15 468 (55.8)	1601 (68.9)	50 151 (56.5)	<0.001	<0.001	0.07
BMI (kg m ⁻²)	27.9 (24.8, 31.3) ^a	27.1 (24.2, 30.2) ^a	30.0 ± 10.3 ^b		25.0–30.0 ^c			
Procedural								
Type of prosthesis								
Primary hip	1528 (56.9)	1528 (100)	14 700 (53.0) ^d	947 (40.7) ^d	88 830 (100)	<0.001	<0.001	
Revision hip	199 (7.4)	0 (0)			0 (0)			
Primary knee	890 (33.2)	0 (0)	12 092 (43.6) ^d	1377 (59.3) ^d	0 (0)	<0.001	<0.001	
Revision knee	67 (2.5)	0 (0)			0 (0)			
Diagnosis								
Primary osteoarthritis	1917 (71.4)	1083 (70.8)			78 578 (88.5)			<0.001
Secondary osteoarthritis	403 (15.0)	241 (15.8)			52 771 (5.9)			<0.001
Sequelae after childhood hip disease	41 (1.5)	36 (2.4)			1862 (2.1)			0.54
Avascular necrosis	138 (5.1)	130 (8.5)			1895 (2.1)			<0.001
Inflammatory joint disease	34 (1.3)	20 (1.3)			1224 (1.4)			0.91
Other	122 (4.5)	4 (0.3)			1895 (2.1)			<0.001
Duration of surgery								
≤ 120 min	1935 (72.1)	1088 (71.2)		2068 (89.0)				<0.001
> 120 min	471 (17.5)	294 (19.2)		256 (11.0)				<0.001
PJI	60 (2.2)	33 (2.2)	1035 (3.7)	43 (1.9)	2173 (2.5)	<0.001	0.39	0.52
Comorbidities								
Charlson comorbidity index								
0	1397 (52.0)	824 (53.9)		1199 (51.6)	67 437 (75.9)			<0.001
1 or more	1287 (48.0)	704 (46.1)		1125 (48.4)	21 393 (24.0)			<0.001
ASA classification								
1	191 (7.1)	146 (9.6)			21 087 (23.7)			<0.001
2	1381 (51.5)	795 (52.0)			52 798 (59.4)			<0.001
3 or more	1096 (40.8)	586 (38.4)			14 945 (16.8)			<0.001
HIV/AIDS	12 (0.4)	10 (0.7)			16 (0.0)			<0.001
Anemia	69 (2.6)	40 (2.6)			619 (0.7)			<0.001
Arrhythmia	267 (9.9)	144 (9.4)			6368 (7.2)			<0.001
Cancer	198 (7.4)	126 (8.2)			3997 (4.5)			<0.001
CNS disease	162 (6.0)	101 (6.6)			3142 (3.5)			<0.001
Coagulopathy	36 (1.3)	24 (1.6)			329 (0.4)			<0.001
Congestive heart failure	112 (4.2)	52 (3.4)		110 (4.7)				0.37
Diabetes mellitus	371 (13.8)	162 (10.6)		544 (23.0)	5973 (6.7)			<0.001
Drug abuse	17 (0.6)	15 (1.0)			682 (0.8)			0.42
Fluid and electrolyte disorders	35 (1.3)	20 (1.3)			534 (0.6)			<0.001
Kidney disease	182 (6.8)	111 (7.3)			883 (1.0)			<0.001
Liver disease	61 (2.3)	38 (2.5)		16 (0.7)	524 (0.6)			<0.001
Lung and airway disease	275 (10.2)	147 (9.6)			4317 (4.9)			<0.001
Myocardial infarction	122 (4.5)	72 (4.7)			3087 (3.5)			0.01
Peptic ulcer	40 (1.5)	29 (1.9)			536 (0.6)			<0.001
Rheumatic disease	209 (7.8)	109 (7.1)			3926 (4.4)			<0.001

Data are presented as the number and percentage of patients unless otherwise indicated. Characteristics that were not reported by the authors of the models were left blank. The abbreviations used in the table are as follows: THA, total hip arthroplasty; BMI, body mass index; PJI, periprosthetic joint infection; ASA, American Society of Anesthesiologists; and CNS, central nervous system. ^a Median (interquartile range). ^b Mean ± standard deviation. ^c Mode interval. ^d The frequencies of primary and revision surgeries were grouped together by Tan et al. (2018) and Del Toro et al. (2019). * *p* values derive from χ^2 tests of homogeneity between frequencies of characteristic variables in the derivation vs. validation population. Populations that had missing data were excluded from the comparison for that characteristic variable. We considered that a *p* value of 0.05 indicated statistical significance.

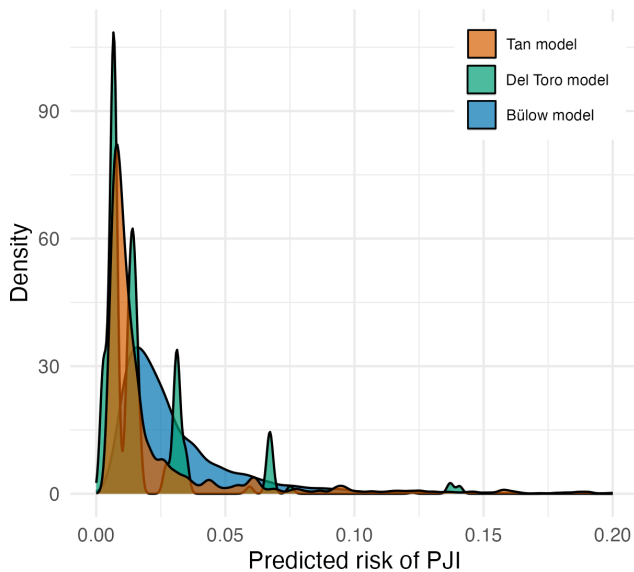


Figure 2. A density plot of the Tan, Del Toro, and Bülow preoperative PJI prediction models, showing the distribution of the predicted risks generated by each model.

higher in our cohort, but statistical significance was not observed ($p = 0.39$).

Our Bülow validation cohort, consisting of only primary THA patients, had higher rates of secondary osteoarthritis and avascular necrosis but a lower rate of primary osteoarthritis ($p < 0.001$ for all) than the derivation cohort. We also observed significantly higher Charlson comorbidity index scores ($p < 0.001$) and higher ASA classification scores ($p < 0.001$) in our validation cohort. Comorbidities were generally more prevalent in our validation cohort; however, the PJI rate did not differ significantly ($p = 0.52$).

3.2 Missing data

The number of patients with missing data was 411 (15%). For the majority of the cases with missing data, only one variable was missing. The variables with the most missing data were duration of surgery, BMI, and ASA classification.

3.3 Distribution of predicted risks

The distribution of predicted risks from all three models were skewed to the right, with predictions above 10% rarely observed (Fig. 2). The Tan model generated risk estimates above this value the most frequently. In contrast to the other models, the Del Toro model displayed distinct peaks in its density plot, as it can only generate 16 discrete risk estimates from its four binary variables.

3.4 Model performance

All three models exhibited comparable, strong discrimination, with c statistics for the Tan, Del Toro, and Bülow models of 0.72 (95% CI: 0.65, 0.78), 0.69 (95% CI: 0.59, 0.78), and 0.72 (95% CI: 0.62, 0.81), respectively (Table 2).

All models displayed reasonable calibration for predicted PJI risks below 3%–4%. The Tan model tended to overestimate the PJI risk above 4%, as indicated by the calibration plot (Fig. 3), a calibration intercept of -0.44 (95% CI: $-0.72, -0.17$), and a calibration slope of 0.51 (95% CI: 0.33, 0.68). For instance, when the Tan model predicted a PJI risk of 5% for a patient, the observed risk of developing a PJI was much lower in our cohort. Conversely, the Del Toro model underestimated the PJI risk above 3%, as shown by the calibration plot and a calibration intercept of 0.19 (95% CI: $-0.07, 0.45$). For example, patients predicted to have a 10% risk of developing a PJI by the Del Toro model actually had a much higher PJI risk. The Bülow model generally overestimated the risk of developing a PJI, reflected by its calibration intercept of -0.35 (95% CI: $-0.70, 0.00$), but showed better calibration at higher risks than the other models. As the Bülow model was specifically developed for primary THA, we only included such procedures in our validation cohort for this model. We performed an additional analysis of the Bülow model in which we included TKA and revision THA as well, but discrimination and calibration were reduced (Fig. 4).

4 Discussion

The implementation of PJI prediction models is often hindered by uncertainty regarding whether they are accurate in new settings. We demonstrated the external validity of three promising preoperative PJI prediction models in a cohort that differed geographically and temporally from the cohorts in which the models were developed. All models showed good discrimination, indicating their clinical utility to identify patients that are at a higher risk of developing a PJI. However, as demonstrated by the calibration plots, the estimated risk of a patient developing a PJI was most accurate for risks up to 3%–4%. Although impact studies are needed, our findings demonstrate the potential of applying the three models in clinical practice as tools for risk stratification.

All three models investigated in this study exhibited performance that aligns with the upper range of performance of preoperative PJI prediction models (Kunutsor et al., 2017; Merrill et al., 2020). While certain models have demonstrated higher accuracy, they have not yet undergone external validation (Klemt et al., 2023; Yeo et al., 2023). To our knowledge, only three other preoperative PJI prediction models have been externally validated; however, their performance has been mixed. One such model, developed at an academic center in the Netherlands, showed poor discrimination (c statistic: 0.55) and calibration during external validation at

Table 2. Statistical performance of the Tan, Del Toro, and Bülow preoperative PJI prediction models.

Performance statistic	Tan model	Del Toro model	Bülow model
<i>c</i> statistic (95 % CI)	0.72 (0.65, 0.78)	0.69 (0.59, 0.78)	0.72 (0.62, 0.81)
Calibration intercept	-0.44 (-0.72, -0.17)	0.19 (-0.07, 0.45)	-0.35 (-0.70, 0.00)
Calibration slope	0.51 (0.33, 0.68)	0.74 (0.59, 0.78)	1.23 (0.76, 1.69)

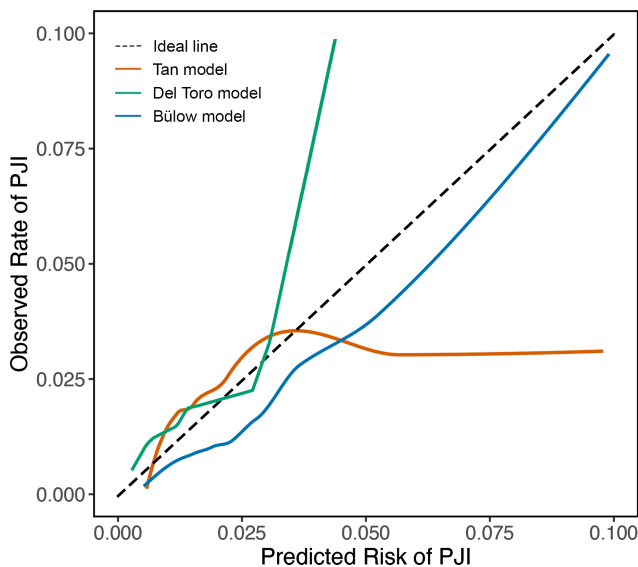


Figure 3. Calibration plots show the agreement between the predicted and observed risks across a range of risks. The dashed line represents perfect calibration. Solid orange, green, and blue lines represent the performance of the Tan, Del Toro, and Bülow preoperative PJI prediction models, respectively. The Tan model overestimated PJI risk above 4 %; for instance, when the model predicted a PJI risk of 5 %, the observed risk was 2.6 %. The Del Toro model underestimated PJI risk above 3 %; for example, a predicted risk of 5 % corresponded to an actual risk of above 20 %. Predicted risks above 0.10 were rarely observed; therefore, they were omitted. The Tan and Del Toro models were evaluated in patients receiving either TKA or THA, whereas the Bülow model was evaluated only in patients who received primary THA.

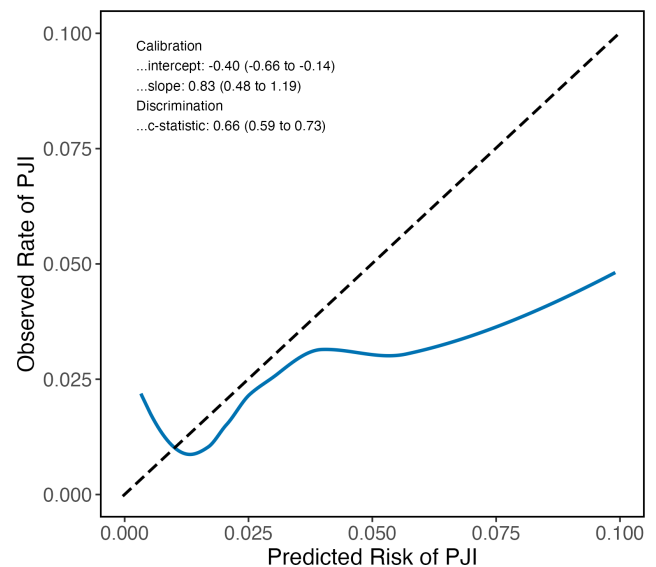


Figure 4. Calibration plot and performance statistics for the Bülow preoperative PJI prediction model for patients receiving THA or TKA. The dashed line represents perfect calibration. Predicted risks above 0.10 were rarely observed; therefore, they were omitted.

a nonacademic center (Sweerts et al., 2023, 2022). Another externally validated model, the American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator (ACS NSQIP SRC), initially showed excellent discrimination with a *c* statistic value of 0.82 (Bilimoria et al., 2013). However, lower *c* statistic values of 0.55, 0.71, and 0.67 were observed in other cohorts (Edelstein et al., 2015; Wingert et al., 2016; Goltz et al., 2018). Moreover, the calibration of the ACS NSQIP SRC has not been assessed, and this calculator has not been validated outside of the US. Espindola et al. (2022) derived a model that showed comparable discrimination to the Tan, Del Toro, and Bülow models; however, its full regression equation has not been published, and the model is not available as an online tool,

rendering its use difficult. Considering these limitations, the Tan, Del Toro, and Bülow models emerge as favorable options in the European setting.

Our findings demonstrate the potential of the Tan, Del Toro, and Bülow models as valuable tools for risk stratification that provide accurate risk estimates in real time. These models are easy to use, requiring a small number of readily available preoperative predictors. Given their ability to classify PJI patients and uninfected patients, the use of prediction models can be beneficial for clinicians and patients during preoperative counseling and for optimization of perioperative preventive strategies. Identifying high-risk patients through prediction models presents a chance to implement additional preventive measures, such as broadening the antibiotic prophylaxis (Iannotti et al., 2020), using dual antibiotic-loaded bone cement (Jenny et al., 2021), or applying negative-pressure wound dressing (Al-Houraihi et al., 2019). Beyond clinical uses, these models hold promise for advancing precision prevention research by providing investigators with more robust methods to identify and study high-risk patients. When selecting which model to use, the advantages and disadvantages of each should be considered.

The Bülow model showed the best calibration, although only for primary THA procedures; when applied to both THA and TKA, calibration was inferior to the other two models (Fig. 4). In contrast, the Tan and Del Toro models are applicable to both THA and TKA; they can classify patients above a 3%–4% predicted risk as having a higher risk than the general population. However, exact risks for individuals above 3%–4% cannot be accurately predicted. The Tan model is accessible as a mobile app and generates a more continuous range of predicted risks than the Del Toro model. Based on these considerations, we believe that the Tan model may be the most practical choice for surgeons routinely performing THA and TKA.

Our results should be interpreted in the light of several limitations. First, the retrospective nature of this study means that the accuracy and completeness of the data may be sub-optimal. Second, we diagnosed PJI and comorbidities using criteria that did not precisely align with those used by the models' authors, which could have introduced biases; nevertheless, this reflects the pragmatic challenges associated with the application of these models. Third, we decided to exclude a variable from the Tan model due to its geographically dependent definition (i.e., health insurance). While this resulted in a loss of information, retaining the variable would have led to elevated predicted risks and greater overestimations. Fourth, our cohort was smaller than the derivation cohorts of two of the models, with relatively few patients developing PJI. This may have limited the precision of the performance measures for discrimination and calibration (Van Calster et al., 2016). Fifth, our study diverged from the follow-up period used by the authors of the Tan and Bülow models. We employed a minimum follow-up period of 1 year, consistent with common practice in assessing PJI risk (Xu et al., 2020) and necessary due to the recency of our cohort. This contrasted with the longer follow-up used by Tan et al. (2018), which may have captured PJI cases with a later onset that we did not observe, resulting in a significantly higher PJI rate in their cohort. Finally, the extent to which our cohort differed from the derivation cohorts could not be comprehensively assessed due to the unavailability of published data.

In conclusion, the Tan, Del Toro, and Bülow preoperative prediction models are valid tools to classify patients at high risk of PJI within Europe. These prediction models hold promise for future clinical application to intensify infection prevention measures in patients at the highest risk of developing a PJI.

Code and data availability. The code and data used in this work are available from the corresponding author upon request.

Author contributions. SJY and MWB: study conception and design; SJY, AS, and RS: data collection; SJY, MWB, PCJ, and WPZ: analysis and interpretation of results; SJY and MWB: manuscript

draft; SJY, MWB, PCJ, WPZ, AS, and RS: manuscript revision and approval. All authors agree to be held accountable for all aspects of the work.

Competing interests. At least one of the (co-)authors is a member of the editorial board of *Journal of Bone and Joint Infection*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

Ethical statement. Institutional review board approval was obtained at each participating hospital before the start of this study. At the coordinating center (University Medical Center Groningen), approval was obtained from the local ethics review board committee, Medische Ethische Toetsingscommissie UMC Groningen (protocol no. 2023/019).

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We would like to thank Katalin Tamasi, medical statistician at the University of Groningen, for her assistance with the statistical analysis.

Review statement. This paper was edited by Antonia Chen and reviewed by two anonymous referees.

References

- Al-Houraihi, R. K., Aalirezaie, A., Adib, F., Anoushiravani, A., Bhashyam, A., Binlaksar, R., Blevins, K., Bonanzinga, T., Chih-Kuo, F., Cordova, M., Deirmengian, G. K., Fillingham, Y., Frenkel, T., Gomez, J., Gundtoft, P., Harris, M. A., Harris, M., Heller, S., Jennings, J. A., Jiménez-Garrido, C., Karam, J. A., Khlopas, A., Klement, M. R., Komnos, G., Krebs, V., Lachiewicz, P., Miller, A. O., Mont, M. A., Montañez, E., Romero, C. A., Schwarzkopf, R., Shaffer, A., Sharkey, P. F., Smith, B. M., Sodhi, N., Thienpont, E., Villanueva, A. O., and Yazdi, H.: General Assembly, Prevention, Wound Management: Proceedings of International Consensus on Orthopedic Infections, *J. Arthroplasty*, 34, S157–S168, <https://doi.org/10.1016/j.arth.2018.09.066>, 2019.
- Bilimoria, K. Y., Liu, Y., Paruch, J. L., Zhou, L., Kmiecik, T. E., Ko, C. Y., and Cohen, M. E.: Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons, *J. Am. Coll. Surg.*, 217, 833–842e3, <https://doi.org/10.1016/j.jamcollsurg.2013.07.385>, 2013.
- Bourne, R. B., Maloney, W. J., and Wright, J. G.: An AOA critical issue. The outcome of the outcomes movement, *J. Bone Joint*

- Surg. Am., 86, 633–640, <https://doi.org/10.2106/00004623-200403000-00026>, 2004.
- Bülrow, E., Hahn, U., Andersen, I. T., Rolfson, O., Pedersen, A. B., and Hailer, N. P.: Prediction of early periprosthetic joint infection after total hip arthroplasty, *Clin. Epidemiol.*, 14, 239–253, <https://doi.org/10.2147/CLEP.S347968>, 2022.
- Cahill, J. L., Shadbolt, B., Scarvell, J. M., and Smith, P. N.: Quality of life after infection in total joint replacement, *J. Orthop. Surg. (Hong Kong)*, 16, 58–65, <https://doi.org/10.1177/230949900801600115>, 2008.
- Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R.: A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, *J. Chronic Dis.*, 40, 373–383, [https://doi.org/10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8), 1987.
- Collins, G. S., Reitsma, J. B., Altman, D. G., and Moons, K. G.: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, *Brit. Med. J.*, 350, g7594, <https://doi.org/10.1136/bmj.g7594>, 2015.
- Dale, H., Fenstad, A. M., Hallan, G., Havelin, L. I., Furnes, O., Overgaard, S., Pedersen, A. B., Källholm, J., Garellick, G., Pulkkinen, P., Eskelinen, A., Mäkelä, K., and Engesaeter, L. B.: Increasing risk of prosthetic joint infection after total hip arthroplasty, *Acta Orthop.*, 83, 449–458, <https://doi.org/10.3109/17453674.2012.733918>, 2012.
- De Cock, B., Nieboer, D., Van Calster, B., Steyerberg, E. W., and Vergouwe, Y.: The CalibrationCurves package: validating predicted probabilities against binary events, Zenodo [code], <https://doi.org/10.5281/zenodo.7801542>, 2023.
- Del Toro, M. D., Peñas, C., Conde-Albarracín, A., Palomino, J., Brun, F., Sánchez, S., and Rodríguez-Baño, J.: Development and validation of baseline, perioperative and at-discharge predictive models for postsurgical prosthetic joint infection, *Clin. Microbiol. Infect.*, 25, 196–202, <https://doi.org/10.1016/j.cmi.2018.04.023>, 2019.
- Edelstein, A. I., Kwasny, M. J., Suleiman, L. I., Khakhkhar, R. H., Moore, M. A., Beal, M. D., and Manning, D. W.: Can the American College of Surgeons Risk Calculator predict 30-day complications after knee and hip arthroplasty?, *J. Arthroplasty*, 30, 5–10, <https://doi.org/10.1016/j.arth.2015.01.057>, 2015.
- Elixhauser, A., Steiner, C., Harris, D. R., and Coffey, R. M.: Comorbidity measures for use with administrative data, *Med. Care*, 36, 8–27, <https://doi.org/10.1097/00005650-199801000-00004>, 1998.
- Espindola, R., Vella, V., Benito, N., Mur, I., Tedeschi, S., Rossi, N., Hendriks, J. G. E., Sorlí, L., Murillo, O., Scarborough, M., Scarborough, C., Kluytmans, J., Ferrari, M. C., Pletz, M. W., McNamara, I., Escudero-Sanchez, R., Arvieux, C., Batailler, C., Dauchy, F. A., Liu, W. Y., Lora-Tamayo, J., Praena, J., Ustianowski, A., Cinconze, E., Pellegrini, M., Bagnoli, F., Rodríguez-Baño, J., and Del Toro, M. D.: ARTHR-IS group: Preoperative and perioperative risk factors, and risk score development for prosthetic joint infection due to *Staphylococcus aureus*: a multinational matched case-control study, *Clin. Microbiol. Infect.*, 28, 1359–1366, <https://doi.org/10.1016/j.cmi.2022.05.010>, 2022.
- Franklin, P. D., Miozzari, H., Christofilopoulos, P., Hoffmeyer, P., Ayers, D. C., and Lübbecke, A.: Important patient characteristics differ prior to total knee arthroplasty and total hip arthroplasty between Switzerland and the United States, *BMC Musculoskelet. Disord.*, 18, 14, <https://doi.org/10.1186/s12891-016-1372-5>, 2017.
- Goltz, D. E., Baumgartner, B. T., Politzer, C. S., DiLallo, M., Bolognesi, M. P., and Seyler, T. M.: The American College of Surgeons National Surgical Quality Improvement Program Surgical Risk Calculator has a role in predicting discharge to post-acute care in total joint arthroplasty, *J. Arthroplasty*, 33, 25–29, <https://doi.org/10.1016/j.arth.2017.08.008>, 2018.
- Gromov, K., Greene, M. E., Sillesen, N. H., Troelsen, A., Malchau, H., Huddleston, J. I., Emerson, R., Garcia-Cimbrelo, E., and Gebuhr, P.; Multicenter Writing Committee: Regional differences between US and Europe in radiological osteoarthritis and self-assessed quality of life in patients undergoing total hip arthroplasty surgery, *J. Arthroplasty*, 29, 2078–2083, <https://doi.org/10.1016/j.arth.2014.07.006>, 2014.
- Hackett, D. J., Rothenberg, A. C., Chen, A. F., Gutowski, C., Jaekel, D., Tomek, I. M., Parsley, B. S., Ducheyne, P., and Manner, P. A.: The economic significance of orthopaedic infections, *J. Am. Acad. Orthop. Surg.*, 23, S1–S7, <https://doi.org/10.5435/JAAOS-D-14-00394>, 2015.
- Harrell, F. E. Jr., Lee, K. L., and Mark, D. B.: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, *Stat. Med.*, 15, 361–387, [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4), 1996.
- Iannotti, F., Prati, P., Fidanza, A., Iorio, R., Ferretti, A., Pérez Prieto, D., Kort, N., Violante, B., Pipino, G., Schiavone Panni, A., Hirschmann, M., Mugnaini, M., and Indelli, F.: Prevention of periprosthetic joint infection (PJI): A clinical practice protocol in high-risk patients, *Trop. Med. Infect. Dis.*, 5, 186, <https://doi.org/10.3390/tropicalmed5040186>, 2020.
- Jenny, J. Y., Hamon, M., Klein, S., Reiter-Schatz, A., Rondé-Oustau, C., Boéri, C., Wisniewski, S., and Gaudias, J.: Cement loaded with high-dose gentamicin and clindamycin reduces the risk of subsequent infection after one-stage hip or knee arthroplasty exchange for periprosthetic infection: a preliminary study, *J. Arthroplasty*, 36, 3973–3978, <https://doi.org/10.1016/j.arth.2021.08.014>, 2021.
- Klemt, C., Yeo, I., Harvey, M., Burns, J. C., Melnic, C., Uzosike, A. C., and Kwon, Y. M.: The use of artificial intelligence for the prediction of periprosthetic joint infection following aseptic revision total knee arthroplasty, *J. Knee Surg.*, 37, 158–166, <https://doi.org/10.1055/s-0043-1761259>, 2023.
- Kunutsor, S. K., Whitehouse, M. R., Blom, A. W., and Beswick, A. D.: Systematic review of risk prediction scores for surgical site infection or periprosthetic joint infection following joint arthroplasty, *Epidemiol. Infect.*, 145, 1738–1749, <https://doi.org/10.1017/S0950268817000486>, 2017.
- Kurtz, S. M., Lau, E., Schmier, J., Ong, K. L., Zhao, K., and Parvizi, J.: Infection burden for hip and knee arthroplasty in the United States, *J. Arthroplasty*, 23, 984–991, <https://doi.org/10.1016/j.arth.2007.10.017>, 2008.
- McNally, M., Sousa, R., Wouthuyzen-Bakker, M., Chen, A. F., Soriano, A., Vogely, H. C., Clauss, M., Higuera, C. A., and Trebše, R.: The EBJIS definition of periprosthetic joint infection, *Bone Joint J.*, 103-B, 18–25, <https://doi.org/10.1302/0301-620X.103B1.BJJ-2020-1381.R1>, 2021.

- Merrill, R. K., Ibrahim, J. M., Machi, A. S., and Raphael, J. S.: Analysis and review of automated risk calculators used to predict postoperative complications after orthopedic surgery, *Curr. Rev. Musculoskelet. Med.*, 13, 298–308, <https://doi.org/10.1007/s12178-020-09632-0>, 2020.
- Paxton, E. W., Cafri, G., Nemes, S., Lorimer, M., Källholm, J., Malchau, H., Graves, S. E., Namba, R. S., and Rolfson, O.: An international comparison of THA patients, implants, techniques, and survivorship in Sweden, Australia, and the United States, *Acta Orthop.*, 90, 148–152, <https://doi.org/10.1080/17453674.2019.1574395>, 2019.
- Premkumar, A., Kolin, D. A., Farley, K. X., Wilson, J. M., McLawhorn, A. S., Cross, M. B., and Sculco, P. K.: Projected economic burden of periprosthetic joint infection of the hip and knee in the United States, *J. Arthroplasty*, 36, P1484–1489.E3, <https://doi.org/10.1016/j.arth.2020.12.005>, 2021.
- Royston, P. and Altman, D. G.: Visualizing and assessing discrimination in the logistic regression model, *Stat. Med.*, 29, 2508–2520, <https://doi.org/10.1002/sim.3994>, 2010.
- Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., and Kattan, M. W.: Assessing the performance of prediction models: a framework for traditional and novel measures, *Epidemiology*, 21, 128–138, <https://doi.org/10.1097/EDE.0b013e3181c30fb2>, 2010.
- Sweerts, L., Hoogeboom, T. J., van Wessel, T., van der Wees, P. J., and van de Groes, S. A. W.: Development of prediction models for complications after primary total hip and knee arthroplasty: a single-centre retrospective cohort study in the Netherlands, *BMJ Open*, 12, e062065, <https://doi.org/10.1136/bmjopen-2022-062065>, 2022.
- Sweerts, L., Dekkers, P. W., van der Wees, P. J., van Susante, J. L. C., de Jong, L. D., Hoogeboom, T. J., and van de Groes, S. A. W.: External validation of prediction models for surgical complications in people considering total hip or knee arthroplasty was successful for delirium but not for surgical site infection, postoperative bleeding, and nerve damage: a retrospective cohort study, *J. Pers. Med.*, 13, 277, <https://doi.org/10.3390/jpm13020277>, 2023.
- Tan, T. L., Maltenfort, M. G., Chen, A. F., Shahi, A., Higuera, C. A., Siqueira, M., and Parvizi, J.: Development and evaluation of a preoperative risk calculator for periprosthetic joint infection following total joint arthroplasty, *J. Bone Joint Surg. Am.*, 100, 777–785, <https://doi.org/10.2106/JBJS.16.01435>, 2018.
- Van Buuren, S. and Groothuis-Oudshoorn, K.: mice: Multivariate imputation by chained equations in R, *J. Stat. Softw.* [code], <https://doi.org/10.18637/jss.v045.i03>, 2011.
- Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M. J., and Steyerberg, E. W.: A calibration hierarchy for risk models was defined: from utopia to empirical data, *J. Clin. Epidemiol.*, 74, 167–176, <https://doi.org/10.1016/j.jclinepi.2015.12.005>, 2016.
- Van Calster, B., Steyerberg, E. W., Wynants, L., and van Smeden, M.: There is no such thing as a validated prediction model, *BMC Med.*, 21, 70, <https://doi.org/10.1186/s12916-023-02779-w>, 2023.
- White, I. R., Royston, P., and Wood, A. M.: Multiple imputation using chained equations: issues and guidance for practice, *Stat. Med.* [code], <https://doi.org/10.1002/sim.4067>, 2011.
- Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D.: dplyr: A Grammar of Data Manipulation, Zenodo [code], <https://doi.org/10.5281/zenodo.7902995>, 2023.
- Wingert, N. C., Gotoff, J., Parrilla, E., Gotoff, R., Hou, L., and Ghanem, E.: The ACS NSQIP risk calculator is a fair predictor of acute periprosthetic joint infection, *Clin. Orthop. Relat. Res.*, 474, 1643–1648, <https://doi.org/10.1007/s11999-016-4717-3>, 2016.
- Xu, C., Tan, T. L., Li, W. T., Goswami, K., and Parvizi, J.: Reporting outcomes of treatment for periprosthetic joint infection of the knee and hip together with a minimum 1-year follow-up is reliable, *J. Arthroplasty*, 35, P1906–1911.E5, <https://doi.org/10.1016/j.arth.2020.02.017>, 2020.
- Yeo, I., Klemm, C., Robinson, M. G., Esposito, J. G., Uzosike, A. C., and Kwon, Y. M.: The use of artificial neural networks for the prediction of surgical site infection following TKA, *J. Knee Surg.*, 36, 637–643, <https://doi.org/10.1055/s-0041-1741396>, 2023.