KNEE

Improving quality of care in total knee arthroplasty using risk prediction: a narrative review of predictive models and factors associated with their implementation in clinical practice

Daniel J Gould,[1*] Michelle M Dowsey,[1,2] Tim Spelman,[1] James A Bailey,[3] Samantha Bunzli,[4] Peter FM Choong,[1,2]

[1] University of Melbourne, Department of Surgery at St Vincent's Hospital, Melbourne, Australia
[2] Department of Orthopaedics at St Vincent's Hospital, Melbourne, Australia
[3] University of Melbourne School of Computing and Information Systems, Melbourne, Australia
[4] School of Health Sciences and Social Work, Griffith University, Nathan Campus, Queensland, Australia

**Corresponding author: daniel.gould@unimelb.edu.au*

Citation: Gould DJ, Dowsey MM, Spelman T, Bailey JA, Bunzli S, Choong PFM. Improving quality of care in total knee arthroplasty using risk prediction: a narrative review of predictive models and factors associated with their implementation in clinical practice. SA Orthop J. 2024;23(1):15-22. http://dx.doi.org/10.17159/2309-8309/2024/v23n1a3

Editor: Prof. Michael Held, University of Cape Town, Cape Town, South Africa

Received: May 2023

Accepted: August 2023

Published: March 2024

Abstract

With the growing capacity of modern healthcare systems, predictive analytics techniques are becoming increasingly powerful and more accessible. Careful consideration must be given to the whole process of prognostic model development and implementation to improve patient care in orthopaedics. Using the example of risk prediction models for total knee arthroplasty outcomes, the literature was reviewed to identify evidence and examples of factors associated with successfully taking predictive models from the computer and implementing them in the clinical environment where they can influence patient outcomes. There were 164 articles included after screening 439 abstracts, 37 of which reported models which had been implemented in the clinical environment. Six of these 37 articles reported some form of clinical impact evaluation, and five of the six evaluated the Risk Assessment and Prediction Tool (RAPT) for arthroplasty. These models demonstrated some positive impacts on clinical outcomes, such as decreased length of stay. However, the findings of this review demonstrate that only a small proportion of developed risk prediction models have been successfully implemented in the clinical environment where they can achieve this positive clinical impact.

Level of evidence: Level 5

Keywords: knee, arthroplasty, risk, predictive model, machine learning

Funding: No funding was received for this study.

Conflict of interest: The authors declare they have no conflicts of interest that are directly or indirectly related to the research.

Introduction

From patient selection to discharge planning, the shared decision-making process between patient and surgeon can benefit from clinically informed multivariable prognostic models.[1,2] A prognostic model is a statistical formula that takes patient characteristics and predicts an outcome, such as the Risk Assessment and Prediction Tool (RAPT),[3] which predicts discharge destination following total knee arthroplasty (TKA) and total hip arthroplasty (THA).

A powerful model predicts the outcome accurately, according to a range of metrics each measuring a specific aspect of predictive performance. Statistical predictive models can outperform clinicians' predictive capability.[4] This is potentially due to predictive models being less prone to clinicians' biases, such as characteristics considered risk factors based on prior personal clinical experience alone. Predictive models are also able to process a greater amount of complex data and generate a prediction where clinicians would be unable to process the sheer volume and complexity of information.[5]

Surgeons appreciate the benefit of using decision aids to enhance shared decision-making but have concerns over what to do with the information.[6,7] Patients have demonstrated the ability to interpret even relatively complex information from decision aids if the information is presented in an informative and user-friendly manner.[8] All key stakeholders must be engaged in a process of co-creation[9] of a system that is created to identify appropriateness of care, predict outcomes, and guide treatment strategies, if predictive analytics is to make its promised impact on healthcare.[10] These stakeholders include researchers, clinicians, statisticians/data scientists, hospital administrative and management staff, and patients/community members.

The objective of risk prediction in the clinical context should be to achieve better patient outcomes by improving quality of care. Too often, research stops at predictive model development.[11] To improve quality of care, this is not enough. Models need to be taken from the computer and implemented in the clinical setting, then evaluated and consistently re-evaluated to inform the process of updating the model to optimise its impact.

This review was contextualised by considering predictive models developed for TKA patients. TKA is a highly effective treatment for advanced osteoarthritis of the knee joint.[12,13] The number of TKA procedures being performed each year continues to grow. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) Annual Report documented 61 154 TKA procedures performed in Australia in 2018.[14] This reflected a 3.8% increase in primary TKA procedures from the previous year, and a 156.2% increase since 2003. It is projected that the volume of TKA procedures will increase by 146% from 2013 to 2046 in Australia, based on a conservative estimate.[15] The international estimates are even more impressive, with the demand for primary TKA procedures expected to increase by 673% in the 25 years leading up to 2030 in the United States.[16] Patients consistently report improved quality of life, reduced pain and better function following the procedure.[17] However, a substantial proportion of patients report dissatisfaction following TKA for a variety of reasons, including persistent pain and functional limitation.[18,19] These are just some examples of outcomes for which predictive models can be developed as part of efforts to mitigate risk of unsatisfactory outcome and maximise the chance of a successful procedure and postoperative course.[20-22]

The aim of this review was to identify evidence and examples of studies in the literature capturing the critical factors associated with TKA risk prediction models making the leap from desk to bedside and having a positive impact on clinical outcomes. Modern computing techniques can process a diverse range of modalities, including tabular data, images, video and audio. The focus of this review is on tabular data. This review builds upon recent work[23] by exploring risk prediction models for TKA developed using machine learning as well as traditional statistical techniques.

Structure of this review

This review is divided into the following sections:

- Literature search and inclusion criteria
- Overview of TKA outcome prediction models
- TKA outcome prediction models implemented in clinical practice
- TKA outcome prediction models evaluated for their impact on quality of care
- Concluding remarks

Literature search and inclusion criteria

A broad literature search was conducted in PubMed to identify studies reporting predictive models developed for TKA outcomes, using the following search strategy: ((knee replacement[Title/Abstract]) OR (knee joint replacement[Title/Abstract]) OR (knee arthroplasty [Title/Abstract]) OR (knee joint arthroplasty[Title/Abstract])) AND ((risk prediction[Title/Abstract]) OR (predictive model[Title/Abstract]) OR (prediction model[Title/Abstract]) OR (prediction[Title/Abstract])). This included both predictive model development studies as well as studies reporting on existing predictive models. Titles and abstracts were screened for studies which reported on predictive models for clinical outcomes following TKA. Studies which used only postoperative risk factors were excluded, as were studies which reported on outcomes related to implant design without a clinical focus such as implant failure. There was no restriction on date of publication. This was not a systematic review, therefore, the search strategy was intentionally broad and inclusive.

Overview of TKA outcome prediction models

The total number of references retrieved was 439, with 164 included following title and abstract screening. The findings of the screening processes are presented in *Table I*. Outcomes were grouped into categories. Some studies reported multiple models for separate outcomes, or single models that were developed for multiple outcomes, hence some studies appear in the table multiple times.

There has been increasing interest in machine learning (ML) throughout recent years in orthopaedics, generally,[24,25] and specifically in TKA.[26] This is illustrated in *Figure 1*. To generate this figure, a distinction was made between ML and non-machine learning models, often referred to as 'traditional statistical models', that had been used in prior literature.[27] If a paper included both ML and non-machine learning models, it was counted as a machine learning paper for the purpose of *Figure 1*.

With the explosion of research and development in artificial intelligence (AI) and ML over the past two decades, there is great excitement about their potential to outperform traditional statistical techniques, such as logistic regression, in terms of discriminative capability.[28] ML is a subset of AI and refers to the process of enabling computers to learn from information and achieve a desired output without rules programmed explicitly by humans. The nature of ML is such that it enables model developers to arbitrarily include a vast range of predictors, which may improve the predictive ability of the model but may not be feasible to obtain manually in the clinical setting for most patients in diverse clinical contexts. Modern AI algorithms, specifically various forms of artificial neural networks, have shown immense potential in image-based[29] and time-series analyses.[30] As such, it is no surprise that there has been growing interest in applying these techniques to clinical risk prediction in healthcare.[31,32]

As patients journey through the modern healthcare system, data of increasing volume and granularity are being generated.[33] This enables more accurate modelling of patients' healthcare utilisation patterns – so-called 'utilomics'[33] – while generating a more detailed profile of individual patients.[34] This facilitates individualised risk

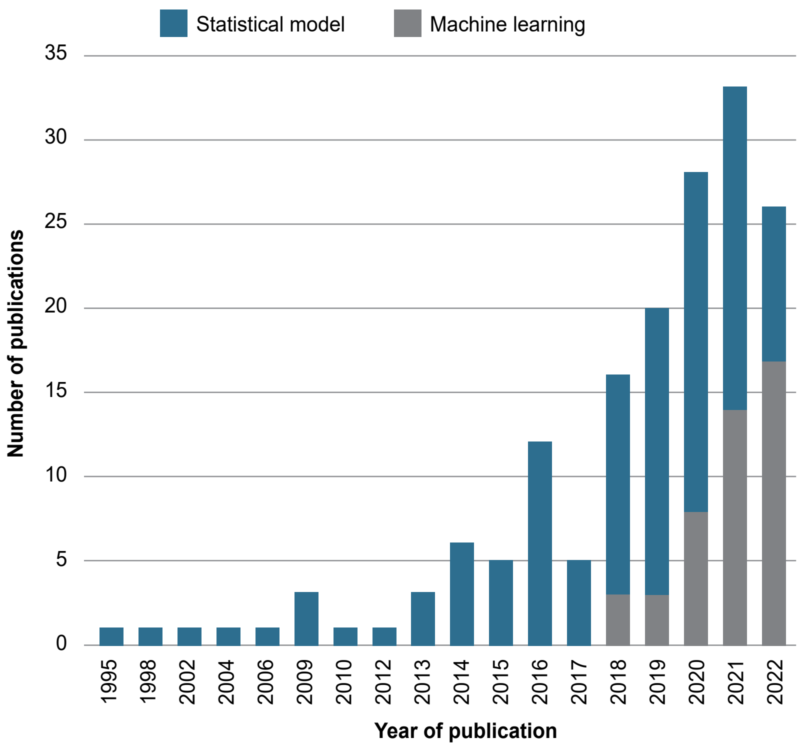

Figure 1. Total knee arthroplasty predictive model studies over time

prediction for more common outcomes based on subtle differences between the unique characteristics of each patient, contrasting traditional population-level risk scores.[35] Another potential advantage is improved predictive accuracy for rarer outcomes.[36] ML is a promising avenue for predictive modelling, but it is not guaranteed to improve predictive performance. In a comprehensive review, ML and AI techniques did not outperform logistic regression in studies at low risk of bias.[27] The power of ML lies in its ability to capture complex interactions and non-linear relationships in the data.[21,37] Surgical risk may possess such qualities[38] but not necessarily for every predictive model development task based on real-world datasets, which may not capture these underlying complexities in a computable form. Furthermore, ML facilitates the inclusion of many predictors, but it may be prudent to include as few variables as possible while retaining strong model performance; this because such models are more likely to be implementable in diverse clinical contexts, thus facilitating external validation and implementation in resource-poor settings.[39] The key is to ensure clinically relevant and readily available predictors are retained in parsimonious models[40] and to be clear from the outset what is the main priority: best statistical fit of the model to the data, or clinical applicability. Superior statistical fit may be achieved with a purely data-driven approach using ML algorithms without clinical insight, but if the gain in predictive performance is only slight then clinical applicability may be preferable.

As with surgery, in predictive modelling it is important to use the right tool for the job. Although prior literature can provide loose guidance when selecting which type of model to use, there is no single model architecture that will be best suited for every task.[41-43] As such, trial and error are necessary. Multiple candidate models are compared using a range of metrics while accounting for other factors such as interpretability, and computational power required for model training. Most ML techniques are available in freely accessible packages for statistical software,[44,45] enabling researchers to trial and tune different models with relative ease such that they do not need to decide which one to use a priori.

Table I: Total knee arthroplasty predictive model studies by outcome category

Outcome category	Studies (surname of first author)
Length of stay (20 papers)	Anis et al.;[1] Johannesdottir et al.;[2] Jorgensen et al.;[3] Carr et al.;[4] Turcotte et al.;[5] Ong et al.;[6] Wei et al.;[7] Poitras et al.;[8] Moore et al.;[9] Cizmic et al.;[10] [a]Gkagkalis et al.;[11] Gronbeck et al.;[12] Han et al.;[13] Li et al.;[14] [a]McCann-Spry et al.;[15] Lopez et al.;[16] [a]Oeding et al.;[17] Klemt et al.;[18] Melfi et al.;[19] Donovan et al.[20]
Pain (or pain relief) (16 papers)	Pua et al.;[21] Li et al.;[22] Sanchez-Santos et al.;[23] Shim et al.;[24] Larsen et al.;[25] Luna et al.;[26] Vogel et al.;[27] Lungu et al.;[28] Tilbury et al.;[29] Buus et al.;[30] [a]Twiggs et al.;[31] Anis et al.;[1] Harris et al.;[32] Arden et al.;[33] Tolk et al.;[34] Dowsey et al.[35]
Function (35 papers)	Pua et al.;[21] Kim et al.;[36] Twiggs et al.;[37] Sanchez-Santos et al.;[23] Shim et al.;[24] Poitras et al.;[8] Amano et al.;[38] Verbeek et al.;[39] Pua et al.;[40] Vissers et al.;[41] Pua et al.;[42] Vogel et al.;[27] Schurman et al.;[43] Hoogeboom et al.;[44] Lungu et al.;[28] Chew et al.;[45] Meessen et al.;[46] Tilbury et al.;[29] Bloomfield et al.;[47] Buus et al.;[30] Nankaku et al.;[48] Braaksma et al.;[49] Pua et al.;[50] [a]Katakam et al.;[51] Turcotte et al.;[52] Li et al.;[53] Rondon et al.;[54] Anis et al.;[1] Harris et al.;[32] Klemt et al.;[55] Chen et al.;[56] Barlow et al.;[57] Arden et al.;[33] Tolk et al.;[34] Dowsey et al.[35]
Satisfaction (or dissatisfaction) (13 papers)	Liu et al.;[58] Van Onsem et al.;[59] Kunze et al.;[60] Zabawa et al.;[61] Calkins et al.;[62] Itou et al.;[63] Garriga et al.;[64] Neuprez et al.;[65] Van Onsem et al.;[66] Kunze et al.;[67] Munn et al.;[68] Barlow et al.;[57] [a]Arden et al.[33]
Revision (7 papers)	Andersen et al.;[69] [a]Starr et al.;[70] Inacio et al.;[71] Mak et al.;[72] Cuthbert et al.;[73] Aram et al.;[74] El-Galaly et al.[75]
Discharge destination (20 papers)	[a]Goltz et al.;[76] Dibra et al.;[77] Dibra et al.;[78] [a]Cohen et al.;[79] Pronk et al.;[80] [a]Hansen et al.;[81] Klemt et al.;[82] [a]Gkagkalis et al.;[11] [a]Lu et al.;[83] Coudeyre et al.;[84] Kugelman et al.;[85] [a]Oldmeadow et al.;[86] Ayala et al.;[87] Hadad et al.;[88] Gholson et al.;[89] Zeng et al.;[90] Zalikha et al.;[91] Menendez et al.;[92] Ottenbacher et al.;[93] Kapoor et al.[94]
Complications (28 papers)	Yayac et al.;[95] Meyer et al.;[96] Abdul-Muhsin et al.;[97] Klausing et al.;[98] Mačiulienė et al.;[99] Zhong et al.;[100] Wang et al.;[101] [a]Hildén et al.;[102] Harris et al.;[103] Rudasil et al.;[104] Seo et al.;[105] Jorgensen et al.;[106] Mulhall et al.;[107] Devana et al.;[108] Chen et al.;[109] Xie et al.;[110] Xiao et al.;[111] Wuerz et al.;[112] Wu et al.;[113] Chotanaphuthi et al.;[114] [a]Jung et al.;[115] Lewallan et al.;[116] Inacio et al.;[117] Inacio et al.;[118] Xie et al.;[119] Jaremko et al.;[120] [a]Ko et al.;[121] [a]Wang et al.;[122] Yeo et al.;[123] Wang et al.;[124] Trivedi et al.;[125] Kapoor et al.;[94] Hyer et al.[126]
Readmission (8 papers)	Anis et al.;[1] Ayers et al.;[127] Mohammadi et al.;[128] Mesko et al.;[129] Goltz et al.;[130] Bhavnani et al.;[131] Kapoor et al.;[94] Hyer et al.;[126] [a]Oeding et al.[17]
Transfusion (9 papers)	To et al.;[132] Pempe et al.;[133] [a]Jo et al.;[134] Huang et al.;[135] Ahmed et al.;[136] Hu et al.;[137] Huang et al.;[138] Yeh et al.;[139] Donovan et al.[20]
Quality of life (6 papers)	Anis et al.;[1] Tanaka et al.;[140] Harris et al.;[32] Huber et al.;[141] Yakobov et al.;[142] Arden et al.[33]
Mortality (8 papers)	Trela-Larson et al.;[143] Penfold et al.;[144] Williams et al.;[145] Harris et al.;[103] Zalikha et al.;[91] [a]Wang et al.;[122] Villacorta Junior et al.;[146] Melfi et al.;[19] Kapoor et al.;[94] Hyer et al.[126]
Postoperative care needs (rehabilitation and ICU) (5 papers)	Zeng et al.;[147] Ditton et al.;[148] Takagawa et al.;[149] [a]Kamath et al.;[150] Dauty et al.[151]
Implant failure (4 papers)	Gudnason et al.;[152] Zhang et al.;[153] Cuthbert et al.;[154] [a]Ellison et al.[155]
Periprosthetic joint infection (4 papers)	Klemt et al.;[156] Del Toro et al.;[157] Sabry et al.;[158] Zhang et al.[159]
Duration of operation (2 papers)	Moteshare et al.;[160] Hinterwimmer et al.[161]
Other (5 papers)	Inpatient payments – Karnuta et al.[162] Sciatic nerve block – Babazade et al.[163] Postoperative opioid use – Klemt et al.[164] Medicare 'super-use' – Hyer et al.[126] KOOS (symptoms, recreation, Junior) – Harris et al.[32]

[a] Embedded model – i.e. implemented for use in the clinical setting (see *Table II*)
References: Appendix A (available online https://saoj.org.za/index.php/saoj/article/view/691)

The important thing is to document, with accompanying computer code, the process of developing the model and the main candidate models from which the final model was selected, taking care to explain the rationale for this selection.[46] Another important consideration is when the model will be used. For example, will the model be used at the point of consent or immediately prior to discharge? What data are available in the system at each of these time points? One must account for any delays between initial entry of raw data into the system and subsequent availability of processed data that can be used by the model.

A more subtle and pervasive problem is class imbalance, in which the cases that the model is being developed to predict comprise a small minority of the observations in the dataset. This is a common feature of real-world medical datasets and ML algorithms are prone to problems with predictive accuracy for the minority class, which typically comprises the high-risk patients.[47] The consequences of failing to properly account for this common challenge has been illustrated in prior literature.[48] In this example, failure to account for class imbalance – there were 260 positive cases and 10 923 negative controls – led to severely imbalanced accuracy, with almost 100% accuracy in predicting the majority (negative) class and only 0–10% for the minority class. The consequence is that up to 90% of patients with cancer would be misclassified (i.e. misdiagnosed) as not having cancer, which is obviously disastrous. Implementing a model without interrogating the results and accounting for such issues is therefore critical, and one must be transparent about how these issues are handled.

TKA outcome prediction models implemented in clinical practice

It is difficult to implement predictive models in clinical practice. This is highlighted by the fact that only a minority (37/164 = 23%) of the models identified in this review (*Table I*) have been deployed in the clinical environment (see *Table II*). In keeping with the focus on inclusiveness prioritised throughout this review, the term 'implemented' was used broadly to refer to any form of clinical implementation, including but not limited to the following: automatic data retrieval and risk calculation in the electronic medical record (EMR), online risk calculators, and depiction of nomograms in the publication reporting the predictive tool. In *Table II*, the specific outcomes predicted by the model detailed in the publication were listed, rather than the outcome categories in *Table I*.

It is important to note that there are many valid reasons for which models in *Table I* were not, or could not, be implemented in clinical practice. For example, the model developers may have determined that their model did not perform well enough to justify implementation, or the teams developing these models were still in the process of engaging administrative, clinical and technical staff to implement the models. In any case, publishing the predictive model study prior to implementation is prudent because it enhances transparency around the specifications and performance of the model.[49] It is also important that external validation of a model is conducted in a population similar to the target population for implementation.[50,51] In the case of online risk calculators, published nomograms, and studies which include the full model coefficients or other specifications required for full independent implementation, it may be possible for interested parties to validate the model on their own data prior to implementation. However, readers should be wary of model developers who have a financial conflict of interest in the uptake of their model, especially in cases where the details required to fully reproduce the model are proprietary.[52-54]

In addition to engaging patients and clinicians, and working with them to build a predictive model both are willing to use in clinical decision-making, engaging members of the hospital administrative and information technology (IT) departments is equally important. Without the proper infrastructure in place, decision support tools can be seen as potentially useful but prohibitively unwieldy.[55,56] The more overtly negative outcome is that they are seen as an untrustworthy, potentially dangerous nuisance. Clinical staff may then use workarounds to maintain the status quo and avoid using the tool altogether.[57] Ideally, the tool is integrated seamlessly into the existing clinical workflow and EMR systems.[58] If this is not possible and changes must be made, such as alterations in the configuration of the tool or development of a separate application outside of the EMR system, then these changes should be minimal. The system must be user-friendly and accessible to users with varied levels of computer literacy. This is where education and training are critical in terms of what the tool can offer and what are the technical aspects involved in using it. This requires strong advocacy and leadership from clinicians, researchers and administrative staff.

TKA outcome prediction models evaluated for their impact on quality of care

This is arguably the most important step. The aim of clinical predictive models should ultimately be to improve patient care. There are many different metrics available to assess statistical performance of predictive models.[59] Compromises often need to be made to optimise performance on a selection of these metrics. For example, for some models it may be more important to sacrifice specificity in preference for higher sensitivity if the outcome of interest is life-threatening and therefore must be detected even if this results in the detection of a relatively high number of false positives.[39]

Contrasting statistical evaluation is evaluating the model's impact on quality of care. An obvious target for evaluation includes event rate, such as a reduction in mortality following deployment of a mortality prediction model.[39] However, a more nuanced and comprehensive evaluation might be more informative.[60] For example, hypothetically, if mortality does not decrease but using the model saves clinicians time in the discharge planning process and assists in the allocation of palliative care services to patients with the highest mortality risk, then the model might be useful even in the absence of reduced mortality. Accurate predictive models could also potentially reduce cognitive load for clinicians by giving them a better understanding of the patient's risk profile even if the model does not predict the outcome with greater accuracy than the clinicians themselves. The model can be formally evaluated in a randomised controlled trial,[46] but this is expensive and logistically challenging. Decision curve analysis can be used to provide a clinically informed, robust estimate of net benefit from using the model, and a clinical trial could be avoided if the model is unlikely to improve clinical outcomes.[61]

A minority of studies in *Table II* underwent some sort of model evaluation (6/37 = 17%). These are depicted in *Table III*. Five of these six studies evaluated the RAPT.[3]

Another way in which models can be useful is by improving the quality of shared decision-making between patient and surgeon,[62-64] in part due to the increased amount of accurately calculated and quantified information pertaining to risks and outcomes. The patient and clinician could have more time in the consultation to discuss what is important to the patient based on their balanced risk of achieving a good outcome compared with experiencing complications. This informs their choice to proceed with surgery and how best to optimise their condition beforehand, maximising benefit and mitigating risk, and how to best plan for discharge and post-discharge follow-up.

Continued after Table III on page 20

Table II: Total knee arthroplasty predictive models implemented into clinical practice

Study	Outcomes	Method of implementation
Goltz et al.[1]	Readmission	Online tool: www.surgicalriskpredictions.com
[a]Twiggs et al.[2]	Pain at 12 months postop	Patients filled out information on web-based '360 Knee System' application on iPad in waiting room
[a]Cohen et al.[3]	Discharge destination	Nursing staff recorded RAPT scores at preadmission appointment ~3 to 4 weeks before surgery. Surgeons then discussed scores with patients in terms of impact on discharge disposition. Case management and physical therapy staff also accessed RAPT scores to further optimise discharge planning. Nomogram depicted in paper
Starr et al.[4]	Revision	Online tool: http://www.bit.do/tka
Hansen et al.[5]	Discharge destination	RAPT scores were prospectively captured, preoperatively, by the nurse case manager
Jo et al.[6]	Transfusion	Online tool: http://safetka.net
[a]Gkagkalis et al.[7]	Discharge destination; length of stay	CAS: Patients were assessed daily with a French version of the CAS by a member of a physiotherapy team RAPT: Patients completed RAPT questionnaire
Hildén et al.[8]	Complications	POSSUM and P-POSSUM scores were administered prospectively. • Preoperatively: orthopaedic surgeon collected the following variables when assessing the patient prior to surgery: cardiac and respiratory signs, ECG and surgical wound • Postoperatively: orthopaedic surgeon assessed operation severity after surgery All other POSSUM variables were collected automatically using integrated software. The software retrieved ICD-10 codes, surrogate variables, and personal information such as sex and age from the patient's medical record.
Lu et al.[9]	Non-routine discharge	Online tool: https://sportsmed.shinyapps.io/Nonroutine_Discharge_UKA/
Ellison et al.[10]	Implant failure	Online tool (no URL provided)
[a]McCann-Spry et al.[11]	Length of stay (LOS)	Patients were sent the RAPT questionnaire to complete at home prior to attending a class in which individual guided conversations were held to discuss patient scores. Patients with scores less than 6 were encouraged to discuss specific discharge planning needs and preferences after class to ensure the appropriate arrangements could be made regarding discharge planning.
[a]Oldmeadow et al.[12]	Discharge destination (specifically, the need for extended inpatient rehabilitation)	Physiotherapist assessed risk using RAPT score at or before admission
Jung et al.[13]	Delirium	Online tool: https://safetka.connecteve.com
Kamath et al.[14]	ICU monitoring	Unclear (bespoke predictive model implemented at a single institution)
Arden et al.[15]	Pain; function; quality of life; satisfaction; implant failure	Implemented in two hospitals as part of a prospective observational study Preoperative visits: After deciding to participate in the study, patients are contacted by a member of the research team. Patients bring a completed 'self-assessment for inpatient surgery' form to the research appointment, and during this appointment they sign a consent form. The following additional tests are undertaken: whole-body DEXA, physical assessment, blood test, urine sample Inpatient data and sample collection: COASt collects inpatient data and intraoperative samples
Ko et al.[16]	Acute kidney injury	Online tool: https://safetka.net
Wang et al.[17]	Major complications (deep wound infection, pneumonia, renal insufficiency or failure, cerebrovascular accident, cardiac arrest, myocardial infarction, pulmonary embolism, sepsis, or death)	Scoring systems depicted in the publication
[a]Oeding et al.[18]	Inpatient vs outpatient status; readmission	On file in EMR at the study institution
Katakam et al.[19]	Function	Online tool: https://sorg-apps.shinyapps.io/tka_koos_mcid/

[a] Impact evaluation carried out (see *Table III*); RAPT = Risk Assessment and Prediction Tool; CAS = Cumulated Ambulation Score; POSSUM = Physiological and Operative Severity Score for the numeration of Mortality and morbidity; P-POSSUM = Portsmouth-POSSUM; ECG = electrocardiogram; DEXA = dual-energy X-ray absorptiometry; COASt = Clinical Outcomes in Arthroplasty Study; EMR = electronic medical record
References: Appendix B (available online https://saoj.org.za/index.php/saoj/article/view/691)

Table III: Total knee arthroplasty predictive models which have undergone impact evaluation

Study	Impact evaluation methodology	Impact per outcome
Twiggs et al.[1]	Consecutive case series validation: First cohort = patients filled out the questionnaire. The surgeon and patient were blinded to the outputs of the tool and consulted for TKA surgery per normal practice. Second cohort = the surgeon and patient were exposed to the outputs of the tool during the consultation for TKA surgery. Statistical validation: A data audit was conducted of all patients who had been consulted with the prediction tool since February 2016, selected for surgery, and consented and answered a 12-month postoperative KOOS questionnaire. There were two predictions of interest: absolute change in pain score predicted and achieved following surgery (analysed as a correlation), and the binary prediction of change greater than or equal to the MCID of KOOS pain.	Use of the tool's predictions did not significantly change the number of patients booked for TKA surgery. However, before introduction of the tool there was no difference between patients booked for surgery and those not booked for surgery in terms of patient-communicated pain state, while after introduction of the tool there was a significant difference in patient-communicated pain between those booked for surgery and those not booked for surgery.
Oldmeadow et al.[2]	Impact of RAPT score implementation was evaluated according to three outcomes: discharge destination, length of stay (LOS), and readmission rates. The score for the prospective group was calculated during the preadmission evaluation. Data collected at discharge included discharge destination (home or rehabilitation) and LOS (days from admission to discharge).	LOS: There was a decrease in LOS in each RAPT risk level. Discharge destination: Logistic regression analysis was used to examine the effects of age, sex, arthroplasty type, LOS, and RAPT score on discharge destination. For cohort 2, only one factor – the RAPT score – had a significant effect on discharge destination. Readmission rate: There was no increase in readmission rates.
McCann-Spry et al.[3]	RAPT implementation was part of a broader interdisciplinary effort to reduce LOS. A list of local subacute rehabilitation and home care agencies was provided to all patients, along with a 'frequently asked questions' sheet to help patients explore options for facilities. Decisions made at the class were communicated to hospital care management staff for reference when the patient arrived at the inpatient unit postoperatively. The impact of the interdisciplinary programme was evaluated in four ways: LOS, cost, early postoperative ambulation, and patient likelihood to recommend the CJR.	Decrease in LOS: Decrease in LOS in January 2014 for TKA patients – this was after pilot implementation of POD 0 physical rehabilitation programme – and April 2014, which was after all providers and staff were fully educated about new LOS expectations. Ten continuous months of LOS data below the previous 2013 average resulted in a process shift in average LOS of 0.5 days. Complications and readmission rates were tracked as separate performance improvement measures – these did not increase. Cost: The average decrease of 0.5 days per patient generated cost savings of approximately $400 per patient after all interventions were implemented. Early postoperative ambulation: In a 1-month follow-up study conducted in June 2015, physical therapists were able to evaluate patients an average of 2 hours sooner than before, and patients stated that they felt less pain on the subsequent day after receiving POD 0 therapy. Also, some patients were able to receive a second therapy session on the day of surgery. Patient likelihood to recommend CJR: Decreasing the LOS did not result in a decrease in patients' likelihood to recommend the CJR. The CJR has consistently exceeded expectations for the question, 'would you recommend this hospital to your friends and family?'
Gkagkalis et al.[4]	Analysis conducted to detect significant difference in hospital LOS between patients with CAS < 11 and CAS ≥ 11. Comparison was also made regarding discharge destination based on this score cut-off. RAPT was analysed as a continuous variable for its association with hospital LOS and discharge destination.	Discharge destination: CAS: Most (85.7%) patients with CAS < 11 were sent to a rehabilitation centre on discharge. In contrast, only 24.1% of patients with CAS ≥ 11 were sent to a rehabilitation centre ($p < 0.001$). RAPT: The association between discharge destination and RAPT as a continuous variable was strongly significant ($p < 0.001$). Length of stay: CAS: The relationship between the CAS 11 score and hospital LOS was not significant ($p = 0.107$). RAPT: The relationship between the RAPT score and hospital LOS was also not significant ($p = 0.64$).
Cohen et al.[5]	Primary TJA patients at a single academic centre before (pre-RAPT) and after (post-RAPT) implementation of the RAPT score were compared.	Implementation of RAPT score significantly decreased hospital LOS from 2.22 (pre-RAPT) to 1.82 days (post-RAPT). The proportion of patients discharged to a facility was also significantly reduced (from 21.8% to 15.2%) without an increased rate of readmission or adverse events.

Table III: Continued

Study	Impact evaluation methodology	Impact per outcome
Oeding et al.[6]	To evaluate the predictive ability of RAPT for inpatient vs outpatient designations and assess how changes to the inpatient-only list may have influenced this potential relationship, RAPT scores were compared between inpatient and outpatient TKA. For TKA and THA patients with RAPT scores on file, mean and median RAPT scores were compared between those with and without 90-day readmission.	Predictive ability of RAPT for inpatient versus outpatient designation: There were significant ($p < 0.001$) differences with decreasing RAPT scores between same-day outpatient, next-day outpatient, and inpatient TKA. Same-day outpatient mean RAPT was 10.35 ± 1.59 for period A, 10.34±1.66 for B, and 10.31 ± 1.54 for C. Next-day outpatient had mean RAPT of 10.07 ± 1.79, 9.70 ± 1.87, and 9.27 ± 2.05 for periods A, B, and C. Inpatient had the lowest mean RAPTs for periods A: 8.74 ± 2.26, B: 8.32 ± 2.35, and C: 8.21 ± 2.27.

TKA = total knee arthroplasty; KOOS = Knee injury and Osteoarthritis Outcome Score; MCID = minimum clinically important difference; RAPT = Risk Assessment and Prediction Tool; LOS = length of stay; POD 0 = postoperative day zero; CJR = Center for Joint Replacement at Spectrum Health; CAS = Cumulated Ambulation Score; TJA = total joint arthroplasty; THA = total hip arthroplasty
References: Appendix C (available online https://saoj.org.za/index.php/saoj/article/view/691)

Beyond the interaction between patient and surgeon, predictive models could positively impact discussions between clinicians. Multidisciplinary team meetings held for high-risk patients with complex surgical needs or complicated risk profiles, could benefit from the utilisation of patient-specific predictive models designed to give a better understanding of the patient's risks for specific outcomes.[65] This is especially pertinent in light of recent evidence demonstrating that a digital platform depicting patient-specific information had a positive effect on efficiency in multidisciplinary team meetings for prostate cancer patients.[66]

In this age of big data and immensely powerful ML models, there is increasing interest in powerful and generalisable predictive models that can be relied upon when implemented in a broad variety of clinical settings, such as hospitals and clinics of various sizes with different computer system infrastructures. However, the benefit of bespoke risk prediction models, taking advantage of the unique aspects of a single institution's data collection, has been demonstrated.[67] The growing availability of large clinical and administrative datasets is an opportunity for the advancement of ML in clinical predictive analytics, but concerns have been raised that too little focus is currently given to the quality of data used to build such models.[68] Having a well-described, robust data collection method boosts confidence in the reliability of predictive models built using the data, as there is transparency around the data collection and auditing processes. It also facilitates implementation and external validation of the model in different clinical settings by enabling comparison of the data used to build the model with that which is available in different settings.[46]

Conclusion

There has been an increase in the number of predictive model development studies for post-surgery outcomes following TKA over the past 20 years, with machine learning increasingly being utilised. However, only a minority of models have been implemented in the live clinical environment, and fewer still have been evaluated to determine whether they are having a positive or deleterious impact on clinical outcomes or shared clinical decision-making. This demonstrates how difficult it can be to implement risk prediction models in the clinical setting, and then evaluate their impact in a nuanced and comprehensive manner. Implementing models, and evaluating their impact, requires stakeholder engagement at multiple levels, including hospital administrative and IT staff, clinicians, patients and research staff experienced in clinical trials. As important as it is to optimise algorithmic performance and drive the technical development of risk prediction techniques, it is time to shift focus in the clinical setting to implementation and ongoing evaluation of such tools. This is necessary to enable them to do what they are supposed to do, which is improve quality of care and increase efficiency of personalised data processing so the patient and clinician can focus on the human connection.

Ethics statement

The authors declare that this submission is in accordance with the principles laid down by the Responsible Research Publication Position Statements as developed at the 2nd World Conference on Research Integrity in Singapore, 2010.
This is a review, therefore ethical approval was not required. Informed written consent was not required.

Declaration

The authors declare authorship of this article and that they have followed sound scientific research practice. This research is original and does not transgress plagiarism policies.

Author contributions

DJG: conceptualisation and design of the review; developed and executed the search strategy, carried out title and abstract screening and subsequent full text screening, extracted data from included articles, synthesised findings and prepared the manuscript; approved the final version
MMD: conceptualisation and design of the review; provided comments on the manuscript and recommendations to improve the synthesis and presentation of key findings; approved the final version
TS: conceptualisation and design of the review; provided comments on the manuscript and recommendations to improve the synthesis and presentation of key findings; approved the final version
JAB: conceptualisation and design of the review; provided comments on the manuscript and recommendations to improve the synthesis and presentation of key findings; approved the final version
SB: conceptualisation and design of the review; provided comments on the manuscript and recommendations to improve the synthesis and presentation of key findings; approved the final version
PFMC: conceptualisation and design of the review; provided comments on the manuscript and recommendations to improve the synthesis and presentation of key findings; approved the final version

ORCID

Gould DJ https://orcid.org/0000-0002-0423-5822
Dowsey MM https://orcid.org/0000-0002-9708-5308
Spelman T https://orcid.org/0000-0001-9204-3216
Bailey JA https://orcid.org/0000-0002-3769-3811
Bunzli S https://orcid.org/0000-0002-5747-9361
Choong PFM https://orcid.org/0000-0002-3522-7374

References

1. Shariat SF, Kattan MW, Vickers AJ, et al. Critical review of prostate cancer predictive tools. Future oncology (London, England). 2009;5(10):1555-84.
2. Varghese J, Kleine M, Gessner SI, et al. Effects of computerized decision support system implementations on patient outcomes in inpatient care: a systematic review. J Am Med Inform Assoc. 2018;25(5):593-602.
3. Oldmeadow LB, McBurney H, Robertson VJ, et al. Targeted postoperative care improves discharge outcome after hip or knee arthroplasty. Arch Phys Med Rehabil. 2004;85(9):1424-27.
4. Kattan MW, Yu C, Stephenson AJ, et al. Clinicians versus nomogram: predicting future technetium-99m bone scan positivity in patients with rising prostate-specific antigen after radical prostatectomy for prostate cancer. Urology. 2013;81(5):956-61.
5. Merath K, Hyer JM, Mehta R, et al. Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. J Gastrointest Surg. 2020;24(8):1843-51.
6. Bunzli S, Nelson E, Scott A, et al. Barriers and facilitators to orthopaedic surgeons' uptake of decision aids for total knee arthroplasty: a qualitative study. BMJ Open. 2017;7(11):e018614.

7. Shelton C, Smith A, Mort M. Opening up the black box: an introduction to qualitative research methods in anaesthesia. Anaesthesia. 2014;69(3):270-80.
8. Mathieu E, Barratt A, Davey HM, et al. Informed choice in mammography screening: a randomized trial of a decision aid for 70-year-old women. Arch Intern Med. 2007;167(19):2039-46.
9. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving research impact through co-creation in community-based health services: literature review and case study. The Milbank Quarterly. 2016;94(2):392-429.
10. Cosgriff CV, Stone DJ, Weissman G, et al. The clinical artificial intelligence department: a prerequisite for success. BMJ health care inform. 2020;27(1).
11. Mateen BA, Liley J, Denniston AK, et al. Improving the quality of machine learning in health applications and clinical research. Nat Mach Intell. 2020:1-3.
12. Ethgen O, Bruyere O, Richy F, et al. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. J Bone Joint Surg Am. 2004;86-a(5):963-74.
13. Association AO. Annual Report 2017. 2017.
14. Association AO. Annual Report. Adelaide, Australia: South Australian Health and Medical Research Institute; 2019.
15. Inacio MCS, Graves SE, Pratt NL, et al. Increase in total joint arthroplasty projected from 2014 to 2046 in Australia: A conservative local model with international implications. Clin Orthop Relat Res. 2017;475(8):2130-37.
16. Kurtz S, Ong K, Lau E, et al. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am. 2007;89(4):780-85.
17. Neuprez A, Neuprez AH, Kaux JF, et al. Total joint replacement improves pain, functional quality of life, and health utilities in patients with late-stage knee and hip osteoarthritis for up to 5 years. Clin Rheumatol. 2020 Mar; 39(3):861-71.
18. DeFrance M, Scuderi G. Are 20% of patients actually dissatisfied following total knee arthroplasty? A systematic review of the literature. J Arthroplasty. 2023 Mar;38(3):594-99.
19. Gunaratne R, Pratt DN, Banda J, et al. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. J Arthroplasty. 2017;32(12):3854-60.
20. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty. 2020;35(11):3117-22.
21. Hinterwimmer F, Lazic I, Suren C, et al. Machine learning in knee arthroplasty: specific data are key—a systematic review. Knee Surg Sports Traumatol Arthrosc. 2022:1-13.
22. Batailler C, Lording T, De Massari D, et al. Predictive models for clinical outcomes in total knee arthroplasty: a systematic analysis. Arthroplasty Today. 2021;9:1-15.
23. Ogink PT, Groot OQ, Karhade AV, et al. Wide range of applications for machine-learning prediction models in orthopedic surgical outcome: a systematic review. Acta Orthop. 2021;92(5):526-31.
24. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. Front Bioeng Biotechnol. 2018;6:75.
25. Lalehzarian SP, Gowd AK, Liu JN. Machine learning in orthopaedic surgery. World J Orthop. 2021;12(9):685.
26. Lopez CD, Gazgalis A, Boddapati V, et al. Artificial learning and machine learning decision guidance applications in total hip and knee arthroplasty: a systematic review. Arthroplast Today. 2021;11:103-12.
27. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22.
28. Artetxe A, Beristain A, Grana M. Predictive models for hospital readmission risk: A systematic review of methods. Comput Methods Programs Biomed. 2018;164:49-64.
29. Noguerol TM, Paulano-Godino F, Martín-Valdivia MT, et al. Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning applications in radiology. J Am Coll Radiol. 2019;16(9):1239-47.
30. Lin Y-W, Zhou Y, Faghri F, et al. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PloS one. 2019;14(7):e0218942.
31. Kerr R. The future of surgery. Bull R Coll Surg Engl. 2019;101(7):264-67.
32. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-31.
33. Ehlers AP, Roy SB, Khor S, et al. Improved risk prediction following surgery using machine learning algorithms. eGEMs. 2017;5(2).
34. Harris AHS, Kuo AC, Weng Y, et al. Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? Clin Orthop. 2019;477(2):452-60.
35. Jalali A, Lonsdale H, Do N, et al. Deep learning for improved risk prediction in surgical outcomes. Sci Rep. 2020;10(1):1-13.
36. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. N Engl J Med. 2016;375(13):1216.
37. Merath K, Hyer J, Mehta R, et al. Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. J Gastrointest Surg. 2020;24(8):1843-51.
38. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HM. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. Ann Surg. 2018;268(4):574-83.
39. Caruana R, Lou Y, Gehrke J, et al., eds. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining; 2015.
40. Steyerberg EW. Clinical prediction models. Springer. 2019.
41. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res. 2014;15(1):3133-81.
42. Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE transactions on evolutionary computation. 1997;1(1):67-82.
43. Wolpert DH. The lack of a priori distinctions between learning algorithms. Neural Comput. 1996;8(7):1341-90.
44. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825-30.
45. Kuhn M, Wickham H. Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles. Boston, MA, USA (accessed 10 December 2020). 2020.
46. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-73.
47. Rahman MM, Davis DN. Addressing the class imbalance problem in medical datasets. Int J Mach Learn. 2013;3(2):224.
48. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on knowledge and data engineering. 2009;21(9):1263-84.
49. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circulation. 2015 Jan 13;131(2):211-19.
50. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69:245-47.
51. Naufal ER, Wouthuyzen-Bakker M, Babazadeh S, et al. Methodological challenges in predicting periprosthetic joint infection treatment outcomes: A narrative review. Front Rehabil Sci. 2022 Jul 11;3:824281.
52. Cristea IA, Cahan EM, Ioannidis JP. Stealth research: lack of peer-reviewed evidence from healthcare unicorns. Eur J Clin Invest. 2019;49(4):e13072.
53. Liu Y, Cohen ME, Ko CY, et al. Considerations in releasing equations for the American College of Surgeons NSQIP surgical risk calculator: in reply to Wanderer and Ehrenfeld. J Am Coll Surg. 2016;223(4):674-75.
54. Ford N, Norris SL, Hill SR. Clarifying WHO's position on the FRAX® tool for fracture prediction. Bull World Health Organ. 2016;94(12):862.
55. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assn. 2017;24(2):246-50.
56. Brufau SR, Wyatt KD, Boyum P, et al. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inform. 2019:104072.
57. Ser G, Robertson A, Sheikh A. A qualitative exploration of workarounds related to the implementation of national electronic health records in early adopter mental health hospitals. PloS one. 2014;9(1).
58. Goltz DE, Ryan SP, Attarian DE, et al. A preoperative risk prediction tool for discharge to a skilled nursing or rehabilitation facility after total joint arthroplasty. J Arthroplasty. 2021;36(4):1212-19.
59. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 2020;21(1):1-13.
60. Klarenbeek SE, Weekenstroo HH, Sedelaar J, et al. The effect of higher level computerized clinical decision support systems on oncology care: a systematic review. Cancers. 2020;12(4):1032.
61. Martin MS, Wells GA, Crocker AG, et al. Decision curve analysis as a framework to estimate the potential value of screening or other decision-making aids. Int J Methods Psychiatr Res. 2018;27(1):e1601.
62. Couet N, Desroches S, Robitaille H, et al. Assessments of the extent to which health-care providers involve patients in decision making: a systematic review of studies using the OPTION instrument. Health Expect. 2015;18(4):542-61.
63. Henry KE, Kornfield R, Sridharan A, et al. Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. NPJ Digit Med. 2022;5(1):1-6.
64. Barlow T, Griffin D, Barlow D, Realpe A. Patients' decision making in total knee arthroplasty: a systematic review of qualitative research. Bone Jt Res. 2015;4(10):163-69.
65. Whiteman A, Dhesi J, Walker D. The high-risk surgical patient: a role for a multi-disciplinary team approach? Oxford University Press. 2016. p. 311-14.
66. Ronmark E, Hoffmann R, Skokic V, et al. The effect of digital-enabled multidisciplinary therapy conferences on efficiency and quality of the decision making in prostate-cancer care. medRxiv. 2022.
67. Morgan DJ, Bame B, Zimand P, et al. Assessment of machine learning vs standard prediction rules for predicting hospital readmissions. JAMA Network Open. 2019;2(3):e190348-e.
68. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearb Med Inform. 2019;28(1):128-34.

Appendices available online at https://saoj.org.za/index.php/saoj/article/view/691