Clinical considerations for risk prediction tools to predict morbidity prior to surgery

About this document

Risk prediction tools play a critical role in preoperative care by estimating the likelihood of adverse outcomes, including mortality, morbidity, and postoperative complications in individuals undergoing surgery. This document has been produced to help provide assistance to surgical teams when considering which risk prediction tool to use prior to surgery.

This work was produced as part of the Health and Care Research Wales Evidence Centre collaboration with the Public Health Wales Evidence Service for Planned Care Wales. The evidence contained within this document was identified in a recent rapid evidence review (which uses abbreviated systematic evidence synthesis methods to provide a description of the distribution and volume of the available evidence) for preoperative risk prediction tools that predict morbidity risk in adults undergoing surgery and an in-depth summary for a subset of tools.
The clinical context and considerations have been added by Consultant Anaesthetists from Planned Care Wales, NHS Productivity and Improvement, Dr Claire Dunstan (National Lead Anaesthetic Clinical Implementation Network), and Dr Catherine Cromey (Deputy Lead Anaesthetic Clinical Implementation Network).

The aim of the evidence review was to identify and map the evidence on the external validation of preoperative surgical risk prediction tools currently used in Wales in any elective, or non-emergency surgical settings, and to provide a more in-depth look at the effectiveness of a selection of the tools deemed to be the most applicable on a population level (ACS NSQIP, P-POSSUM, RCRI, ASA classification system). For more information, the full report is available here.

The evidence review focused on external validation studies of risk prediction models for assessing the risk of postoperative morbidity and complications. **External validation studies** evaluate both **discrimination** and **calibration** to determine a tool's performance (how accurately it predicts a risk) (Collins et al., 2014). A glossary of key terms can be found at the end of this document.

Clinical considerations and implications

Evaluating postoperative risk at the preoperative stage to inform decisions about surgery is a complex process which requires specialist and expert clinical input. Each specific tool needs to be used in context. Out of context use can yield widely different results/outcomes. The use of single tools (e.g. ASA, NSQIP) in isolation is not recommended, but used appropriately, they can contribute to the clinical decision making and subsequent risk discussion with patients and family. There needs to be an evidence-based shift in practice to align with the research included in this evidence map and summary.

This work suggests that risk prediction would be better performed by a digital solution which can pull the required information on surgical procedure and the patient's medical conditions to ascertain the most appropriate tools for the individual.

Risk prediction tools considerations

Overall, no one tool was identified that adequately predicted complications across all surgical specialties, so it may be likely that some tools are better suited for specific surgery types or that a combination of risk prediction tools may be needed to adequately assess an individual's level of risk. There was considerable heterogeneity among the included studies in which surgical specialties the risk prediction tools were used for, how complications were defined, and which outcome measures were used to determine a tool's predictive ability. This makes direct comparisons very challenging. In addition, the methodological quality of these studies were not assessed, so any findings reported here and in the full evidence review should be interpreted with caution. A summary of the risk prediction tools and the evidence base identified can be seen in Tables 1 and 2. Table 1 includes the four risk prediction tools selected for an in-depth summary (ACS NSQIP, P-POSSUM, RCRI, and the ASA classification system), their recommended uses, their outputs, the number of external validation studies identified across the differing surgical specialties and a reflection of the discriminative ability of the four risk prediction tools. Table 2 includes the remaining risk prediction tools included in the evidence review and outlines their recommended uses, outputs, and the number of external validation studies identified across the differing surgical specialties. The discriminative ability of the tools in Table 2 have not been assessed.

A summary of the key findings for the four risk prediction tools is reported below. The predictive ability of the tools varied across different surgical specialties. However, the certainty of the findings for the different surgical specialities may be limited due to a very small evidence base available for each surgical specialty.

ACS NSQIP

The ACS NSQIP surgical risk calculator is suitable for any surgical procedure. It provides Individual risk scores for up to 19 different outcomes within 30-days following surgery, including a predicted length of stay. It is important to note that ACS NSQIP allows the clinician to adjust the predicted risk if they feel the patient has risk factors that have not been included by the tool which may impact how well the tool was determined to perform (ACS NSQIP, 2025).

ACS NSQIP was found to have a **poor predictive ability for composite complications across all studies.** There is limited evidence to suggest that ACS NSQIP tool had variable ability to predict postoperative morbidity risk across the following specific surgical disciplines:

- Mixed surgery - **Excellent predictive ability**
- Thoracic surgery and Plastic surgery **- Fair predictive ability**
- General surgery, Neurosurgery, or Gynaecology surgery **- Poor predictive ability**
- Orthopaedic surgery, urology surgery or vascular surgery - **Very poor predictive ability**

ASA classification system

The ASA classification system is suitable for use as part of a comprehensive preoperative assessment for patients undergoing surgery, and not in isolation (Saklad, 1941; American Society of Anesthesiologists, 2020). It provides an ASA classification where a higher grade indicates a higher preoperative risk.

The ASA classification system was found to have a **poor predictive ability for composite complications across all studies.** There is limited evidence to suggest that the ASA classification system had variable ability to predict postoperative morbidity risk across the following specific surgical disciplines:

- Mixed surgery – **Fair predictive ability**
- General surgery or orthopaedic surgery - **Poor predictive ability**
- Urology surgery or vascular surgery – **Very poor predictive ability**

P-POSSUM

The P-POSSUM tool is recommended to be used in emergency and elective general surgical procedures (Prytherch et al.,1998; MDCALC, 2024). It provides a single overall score for predicted morbidity.

P-POSSUM was found to have a **poor predictive ability for composite complications across all studies.** There is limited evidence to suggest that P-POSSUM had variable ability to predict postoperative morbidity risk across the following specific surgical disciplines:

- ENT surgery – **Fair predictive ability**
- General surgery – **Poor predictive ability**
- Gynaecology surgery – **Very poor predictive ability**

RCRI

The RCRI is used in elective non-cardiac surgery or urgent/semi-urgent (non-emergent) non-cardiac surgery (Lee et al.,1999; MDCALC, 2024). It provides a single overall score for predicted postoperative cardiac complications.

The RCRI was found to have a **fair predictive ability for composite complications across all studies.** There is limited evidence to suggest that the RCRI had variable ability to predict postoperative morbidity risk across the following specific surgical disciplines:

- Mixed surgery, vascular surgery, or orthopaedic surgery – **Fair predictive ability**
- Urology surgery – **Poor predictive ability**

Summary of the evidence gaps

- It is unclear whether study findings would be generalisable to the UK.
- No quality appraisal of included studies was conducted.
- No evidence was identified assessing the predictive ability of two tools: the Carlisle Risk Calculator and the NELA PRS.

Table 1. Summary of risk prediction tools, the available evidence, and discriminative ability findings

No tool should be used in isolation for clinical decision making												
Tool name	Recommended surgical disciplines	Output	Number of external validation studies and discriminative ability by surgical type*									
			ENT	GE	GY	M	N	OMF	O	P	U	T	V
ACS NSQIP	Any procedure, in most surgical subspecialties,	Individual risk scores for up to 19 different outcomes within 30-days following surgery, including a predicted length of stay.		12	1	3[1]	4		6	2	3	1	1
ASA Classification system	Not specified however it is not recommended to be used alone to predict risk of morbidity.	Provides an ASA classification where a higher grade suggests higher preoperative risk.		1		2[2]			3		1		2
P-POSSUM	Emergency and elective general surgical procedures.	A single overall score for predicted morbidity.	1	5	1								
RCRI	Elective non-cardiac surgery or urgent/semi-urgent (non-emergent) non-cardiac surgery.	A single overall score for predicted postoperative cardiac complications.				9[3]			2		1		1

Abbreviations: ACS NSQIP = The American College of Surgeons National Surgical Quality Improvement Program Tool; ASA Classification system = The American Society of Anesthesiologists Physical Status Classification System; P-POSSUM - The Portsmouth Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity; RCRI - The Revised Cardiac Risk Index; ENT = Ear, nose and throat surgery; GE = General surgery; GY = gynaecological surgery; M = mixed surgical specialties; N = Neurosurgery; OMF = Oral and Maxillofacial surgery; O = Orthopaedic surgery; P = Plastic surgery; T = thoracic surgery; U = Urology surgery; V = vascular surgery.

[1] = Mixed surgeries include: abdominal, anorectal surgeries, breast surgery, ENT surgery, excision and incision biopsies of superficial masses, head and neck surgeries, laparoscopic, neurosurgical procedures, orthopaedic surgeries, obliterative and other (urethral diverticulum excision, fistula repair, or vaginal cyst removal); open, laparoscopic and percutaneous abdominal surgeries, thoracic surgery, thyroid surgeries, urologic surgeries, vaginal, vascular surgeries, wound debridement.

[2] = Mixed surgeries include open major upper elective (partial/total colectomy; Hartmann's procedure; total/partial gastrectomy; liver resection; and pancreatic-duodenectomy) and lower (nephrectomy, prostatectomy or hysterectomy) abdominal surgery, and a range of non-cardiothoracic surgeries including, lower limb surgery, upper limb surgery, neck surgery, other, pelvic surgery,

[3] = Mixed surgeries include a range of non-cardiac surgeries including: Abdominal aortic aneurysm repair, abdominal (bowel), abdominal (nonbowel), abdominal surgery, open abdominal surgery, other abdomen, laparoscopic abdominal surgery, anorectal, aortic, breast, bariatric, brain, breast, carotid endarterectomy, colectomy, cystectomy, ear nose and throat surgeries, endocrine, esophagectomy, excision/incision biopsy superficial, eye, lower extremity bypass, female reproductive, foregut/hepatopancreatobiliary, GBAAS/intestinal, lower gastrointestinal, upper gastrointestinal, gastrectomy, genitourinary, gynaecology, head and neck, hernia, hepatobiliary, hysterectomy, intercranial, lower limb surgery, upper limb surgery, male reproductive, neck surgery, neurosurgery, nephrectomy, nonarterial vessels, non-cardiac, non-vascular surgery, obstetrics, orthopaedic, other, major orthopaedic, minor orthopaedic, otolaryngology, pancreatectomy, pelvic surgery, peripheral vascular, pigtail insertion, plastics, pneumonectomy, prostatectomy, skin, spine, thoracic, thyroid, total hip replacement, total knee replacement, trauma, urologic, vascular, vein, wound debridement.

***Key for summarising discriminative ability:** – Worse than chance, very poor or poor discriminative ability; – Fair discriminative ability; – Good or excellent discriminative ability.

Table 2. Summary of other risk prediction tools and the available evidence

No tool should be used in isolation for clinical decision making												
Tool name	Recommended surgical disciplines	Output	Number of external validation studies by surgical type*									
			ENT	GE	GY	M	N	OMF	O	P	U	T	V
ARISCAT	Surgery performed under general, neuraxial, or regional anaesthesia.	A single overall score for predicted postoperative pulmonary complications.		2								2	
CPET	Abdominal, colorectal, urological, hepatobiliary, liver, bariatric, vascular, thoracic, oesophageal gastric.	Scores for VO^2 peak and anaerobic threshold.	1	11		1					1	2	1
CFS	N/A not originally designed for surgery.	Provides an overall frailty score.				3					1		
DASI	Not specified but used during preoperative assessments for surgery and before and during exercise programmes.	A single overall score for functional capacity, VO^2 max, and metabolic equivalents.				1							
NRS-2002	N/A not originally designed for surgery.	Provides an overall score for nutritional risk.		3								2	
PONV	Any surgery performed under general anaesthesia.	Provides an overall score for the risk of experiencing PONV.		2		6						1	
POSSUM	See P-POSSUM	See P-POSSUM		9		1	1		1			1	1
SORT	Surgery not specified, but for use in adults only.	Provides an overall risk score.		1	1								

Abbreviations: ARISCAT = The Assess Respiratory Risk in Surgical Patients in Catalonia; CPET = Cardiopulmonary exercise testing; CFS = The Clinical Frailty Scale; DASI = The Duke Activity Status Index; NRS-2002 = The Nutritional Risk Screening 2002; PONV = The Apfel score for Postoperative Nausea and Vomiting; POSSUM = The Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity; SORT = The Surgical Outcome Risk Tool; ENT = Ear, nose and throat surgery; GE = General surgery; GY = gynaecological surgery; M = mixed surgical specialties; N = Neurosurgery; OMF = Oral and Maxillofacial surgery; O = Orthopaedic surgery; P = Plastic surgery; T = thoracic surgery; U = Urology surgery; V = vascular surgery.
*Discriminative ability was not assessed for these tools.

Glossary

Term	Definition
Accuracy	Accuracy of risk prediction tools is commonly assessed in the literature using the 'Brier Score'. This is a simultaneous measure of calibration and discrimination, reported as a score between 0 and 1. A score of 0 indicates no difference between the outcome predicted by the tool and actual outcome, thus indicating the best possible result. A score of 1 indicates that the test did not predict the outcome.
Calibration	Calibration assesses the agreement between the predicted and observed outcomes, typically presented graphically as observed risks versus predicted risks or an observed/expected (O/E) ratio (Collins et al., 2014). An O/E ratio of 1 implies the tool accurately predicted complications, a ratio below 1 suggests the tool underpredicted complications, whereas a ratio over 1 suggests the tool overpredicted complications (Hammond et al., 2021). Following the approach utilised by NICE, an O/E ratio of between 0.9 to1.1 would be considered a fair level of calibration.
Composite outcomes	Rather than reporting individual outcomes, some studies combined two or more outcomes into a single measure. For example, 'All complications' could include, pneumonia; sepsis, infection etc. Where outcomes are combined, this is known as a composite outcome.
Discrimination	Discrimination measures how well a tool differentiates between patients who do and do not experience an event, quantified using the area under the receiver operating characteristic (ROC) curve (AUC) or c-statistic. Within the literature, the following c-statistic scores relate to a tools performance: • 90% or greater -considered an excellent level of discriminative ability, • 80% or greater is considered a good level of discriminative ability, • 70% or greater is considered a fair level of discriminative ability, • 60% or greater is considered a poor level of discriminative ability • 50% or greater is considered a very poor level of discriminative ability (or no better than chance) (NICE 2020; Çorbacıoğlu and Aksel, 2023).
External validation	In this context external validation refers to the evaluation of a risk prediction tool's predictive performance in a dataset that was not used to develop the tool, which is known as internal validation (Collins et al., 2014).
Median	The median is the middle value in a group of numbers ranked in order of size, in this way it is not sensitive to outliers or skewed data.
Mixed surgery	In this context mixed surgery refers a study that included a dataset or sample with a range of surgeries across more than one surgical specialty
Rapid Evidence Map	Rapid Evidence Maps (REMs) use abbreviated systematic mapping or scoping review methods to provide a description of the nature, characteristics and volume of the available evidence for a particular policy domain or research question. The methods have been developed as part of the Health and Care Research Wales Evidence Centre Collaboration.
Risk prediction tools	Tools, usually a set of clinical or personal information that is used to calculate the estimated likelihood of adverse outcomes, including mortality, morbidity, and postoperative complications.

References

ACS NSQIP. (2025) Risk Calculator Frequently Asked Questions. Available at: https://riskcalculator.facs.org/RiskCalculator/faq.html

American Society of Anesthesiologists (2020) Statement on ASA Physical Status Classification System. Available at: https://www.asahq.org/standards-and-practice-parameters/statement-on-asa-physical-status-classification-system

Collins, G. S., de Groot, J. A., Dutton, S., et al. (2014). External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC medical research methodology, 14, 1-11. Available at: https://link.springer.com/article/10.1186/1471-2288-14-40

Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. Turkish Journal of Emergency Medicine, 23(4),195. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10664195/

Hammond, G., Johnston, K., Huang, K., et al. (2020). Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death. Circulation: Cardiovascular Quality and Outcomes, 13(6), e006752. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC7299736/

Lee, T. H., Marcantonio, E. R., Mangione, C. M., et al. (1999). Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation, 100(10), 1043–1049. Available at: https://www.ahajournals.org/doi/10.1161/01.CIR.100.10.1043

MDCALC (2024). All Calculators. Available at: https://www.mdcalc.com/

NICE (2020) Perioperative care in adults [C] Evidence review for preoperative risk stratification tools. Available at: https://www.nice.org.uk/guidance/ng180/evidence/c-preoperative-risk-stratification-tools-pdf-8833151056

Prytherch, D. R., Whiteley, M. S., Higgins, B., et al. (1998). POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. *The British journal of surgery*, *85*(9), 1217–1220. Available at: https://doi.org/10.1046/j.1365-2168.1998.00840.x

Saklad, M. (1941, May). Grading of patients for surgical procedures. In The Journal of the American Society of Anesthesiologists (Vol. 2, No. 3, pp. 281-284). The American Society of Anesthesiologists. Available at: https://journals.lww.com/anesthesiology/citation/1941/05000/grading_of_patients_for_surgical_procedures.4.aspx

Acknowledgement to Public Health Wales NHS Trust to be stated.

ISBN: 978-1-83766-619-5

This work was funded for this work by the Health and Care Research Wales Evidence Centre, itself funded by Health and Care Research Wales on behalf of Welsh Government.

Health and Care Research Wales Evidence Centre

Canolfan Dystiolaeth Ymchwil Iechyd a Gofal Cymru

Ymchwil Iechyd a Gofal Cymru Health and Care Research Wales

Ariennir gan Lywodraeth Cymru Funded by Welsh Government