BJA Open

BJA Open, 3 (C): 100027 (2022)

doi: 10.1016/j.bjao.2022.100027
Advance Access Publication Date: 17 August 2022
Editorial

EDITORIAL

Risk scores in anaesthesia: the future is hard to predict

Daniel James Drayton[1,*], Michael Ayres[2], Samuel D. Relton[3], Matthew Sperrin[4] and Marlous Hall[1]

[1]Leeds Institute for Data Analytics, Leeds, UK, [2]Leeds Institute of Medical Research, Leeds, UK, [3]Leeds Institute of Health Science, University of Leeds, Leeds, UK and [4]Division of Informatics, Imaging & Data Sciences, University of Manchester, Manchester, UK

*Corresponding author. E-mail: d.j.drayton@leeds.ac.uk

Summary: External validation helps to assess whether a given risk prediction model will perform well in a target population. Validation is an important step in maintaining the utility of risk prediction models, as their ability to provide reliable risk estimates will deteriorate over time (calibration drift).

Keywords: anaesthesia; modelling; perioperative medicine; perioperative risk; risk prediction

The number of available risk prediction models has increased over the past decade, and their use in the surgical population has been encouraged. In 2011, the National Confidential Enquiry into Patient Outcome and Death report ‘Knowing the Risk’ recommended that ‘an assessment of mortality risk should be made explicit to the patient and recorded clearly on the consent form and in the medical record’.[1,2] Risk prediction models aim to calculate an individual’s risk of having (diagnostic) or developing (prognostic) an event.[3,4] There are several examples of clinical risk prediction models in anaesthesia, including the Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM), the National Emergency Laparotomy Audit, and Surgical Outcome Risk Tool (SORT).[5–8] Whilst such clinical risk prediction models are designed to be used routinely in clinical practice, they are often applied sporadically and without validation in their target population, which means we do not know if they will perform well for their intended purpose.[2]

In this issue of *BJA Open*, Torlot and colleagues[9] externally validated the following four prediction models: SORT, nzRISK, and POSSUM and its Portsmouth variant, by assessing their performance in 44 031 adult Australian patients from the non-indigenous population. Their well-conducted validation study was reported using the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) checklist, which is the recognised reporting standard for multivariable prognostic or diagnostic prediction models. As the number of available prediction models grows, adherence to reporting guidelines, such as the TRIPOD statement, is essential to ensure the overall scientific integrity of such models.[10]

In their study, Torlot and colleagues[9] uncovered the commonly encountered problem of ‘calibration drift’, which means calibration (how trustworthy the risk estimate is) has become worse over time.[11] In the study of Torlot and colleagues,[9] calibration drift resulted in the 30 day mortality rate being over-predicted when applied to the target population. There are several reasons for calibration drift for any given prediction model, including changes in population demographics, clinical practice patterns, the introduction of new guidelines, or even a change in data recording practices or systems. Although frequently encountered, this is an under-appreciated issue and highlights the importance of ensuring that risk prediction models are accurate for each intended target population before assessing their clinical value. Furthermore, prediction models are frequently developed as a one-time activity, but the reality is that model construction is just a small part of a larger ‘pipeline’, which would be more appropriately viewed as a cycle. Validation is a crucial step in these cycles.[12,13] One way to address this problem is by updating prediction models on a regular basis. The European

DOI of original article: 10.1016/j.bjao.2022.100018.

System for Cardiac Operative Risk Evaluation is one example where an existing prediction model was updated with contemporaneous data and predictors.[13,14]

Static prediction models are always at risk of being outdated.[12,13] Therefore, approaches are needed to determine how to validate them and when they should be formally updated.[12] One solution is proposed by Davis and colleagues,[15] who presented a model-agnostic calibration drift detection system. They used data-driven methods to monitor calibration metrics to detect when model performance deteriorates, providing a guide for when they should be updated. Updating prediction models requires time, money, and infrastructure that are not always readily available, and even prediction models that are periodically updated are at risk of calibration drift between updates. The ideal solution may lie in an approach called dynamic modelling (or online machine learning), in which new data are continuously monitored and the prediction model is continuously updated and validated. Although these methods are deployed in marketing and by social media companies,[16] they are not yet well established in healthcare research and would require significantly more complex infrastructure and ongoing technical expertise, which is not the case for current 'static' prediction models.[13,15,17]

An alternative to updating prediction models is to design bespoke risk prediction models for the target population in whom they will be deployed, which Torlot and colleagues[9] discussed. Whilst this approach has merits, it risks contributing to 'research waste', where old, outdated prediction models are discarded in favour of newer models.[13] In this situation, careful consideration is needed for what prediction task is required. For example, Grant and colleagues[18] highlight that a model designed to benchmark multiple procedures accordingly to adjusted surgical risk will look very different to a model trying to predict a patient's particular outcome for a specific procedure.

As for the future, the increasing collection and availability of Electronic Health Records (EHRs) clearly offer a rich data source of broadly representative data often for large populations from which new prediction models can be derived, but consideration is still required for how the data were collected, and for what purpose, to understand the potential for selection bias.[12] Whilst the extent and sources of bias are yet to be fully described, validating prediction models developed in systems with inherent bias will be difficult.[19]

The precise future of risk prediction models in anaesthesia is hard to forecast. A plethora of policy statements and guidelines have made the case for systematic perioperative risk estimation, but it has not become embedded into routine clinical practice. Closing this translational gap requires funding for qualitative work on patient, clinician, and policy expectations of preoperative risk estimation and funding to support an implementation science-based approach to carefully integrate risk prediction models into the clinical workflow. This will require recognition of the state of healthcare technology relative to other industries and a considered approach to ensure risk prediction models help, not hinder, clinical practice.

Authors' contributions

Editorial conception: DJD, SDR, MS, MH.
Initial draft: DJD, MA; Critical revisions: DJD, MA, SDR, MS.

Acknowledgements

Dr Simon J. Howell provided clinical expertise and reviewed the final draft of this editorial.

Funding

Dr Daniel James Drayton is an Academic Clinical Fellow in Anaesthesia funded by the National Institute for Health and Care Research (NIHR).

Declarations of interest

The authors declare that they have no conflicts of interest.

References

1. Findlay G, Goodwin A, Protopapa K, Smith N, Mason M. *Knowing the risk: a review of the peri-operative care of surgical patients* 2011. Available from https://www.ncepod.org.uk/2011report2/downloads/POC_summary.pdf. [Accessed 26 May 2022]
2. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016; **353**: i2416
3. Dhiman P, Ma J, Andaur Navarro CL, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. *BMC Med Res Methodol* 2022; **22**: 101
4. Verbakel JY, Steyerberg EW, Uno H, et al. ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. *J Clin Epidemiol* 2020; **126**: 207–16
5. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; **78**: 355–60
6. Protopapa KL, Simpson JC, Smith NCE, Moonesinghe SR. Development and validation of the surgical outcome risk tool (SORT). *Br J Surg* 2014; **101**: 1774–83
7. Eugene N, Oliver CM, Bassett MG, et al. Development and internal validation of a novel risk adjustment model for adult patients undergoing emergency laparotomy surgery: the National Emergency Laparotomy Audit risk model. *Br J Anaesth* 2018; **121**: 739–48
8. Bilimoria KY, Liu Y, Paruch JL, et al. 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* 2013; **217**: 833–42. e1–3
9. Torlot F, Chang-Yang Y, Reilly J, et al. The external validity of four risk scores predicting 30-day mortality after surgery. *BJA Open* 2022; **3**, 100018
10. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015; **13**: 1
11. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128–38
12. de Hond AAH, Leeuwenberg AM, Hooft L, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med* 2022; **5**: 2

13. Jenkins DA, Martin GP, Sperrin M, et al. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagn Progn Res* 2021; **5**: 1
14. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; **357**: j2099
15. Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. *J Biomed Inform* 2020; **112**, 103611
16. Amoore L. Machine learning political orders. *Rev Int Stud* 2022: 1—17
17. Carlisle JB. Risk prediction models for major surgery: composing a new tune. *Anaesthesia* 2019; **74**: 7—12
18. Grant SW, Collins GS, Nashef SAM. Statistical primer: developing and validating a risk prediction model. *Eur J Cardiothorac Surg* 2018; **54**: 203—8
19. Huang JY. Representativeness is not representative: addressing major inferential threats in the UK Biobank and other big data repositories. *Epidemiology* 2021; **32**: 189—93