Clinical scores and risk factors to predict patient outcomes: how useful are they?

KC Chong *, PhD, **SY Chan,** BSc, **Katherine M Jia,** BSc

School of Public Health and Primary Care, The Chinese University of Hong Kong, Shatin, Hong Kong

* Corresponding author: marc@cuhk.edu.hk

Clinical scores and risk factors for a prediction of patient outcomes are useful for improving patient care. Famous examples include the response evaluation criteria in solid tumours (RECIST) score for guidance of treatment and the Framingham Risk Score for risk assessment of cardiovascular and related diseases. One great potential of clinical scores is accelerating diagnosis and providing timely treatment. In the case of pregnant women with pre-eclampsia, the result of spot urine protein-to-creatinine ratio test is highly correlated with that of the usual diagnostic criteria—over 300 mg of protein in a 24-hour urine sample.[1] This allows prompt response or follow-up in positive cases and increases management efficiency. In addition, simpler detection methods with similar accuracy can encourage more people to take a test or complement existing tests to reduce errors, as seen with non-invasive prenatal testing after its introduction in Hong Kong in 2011.[2]

Risk factors can also be used to estimate the risk of mortality. In a study of Chinese geriatric patients who had received hip fracture operations, Lau et al[3] combined the Charlson Comorbidity Index with score weighting that reflects age to form the total Charlson comorbidity score of patients. The authors found this score to be significantly associated with 30-day and 1-year mortality risk in geriatric patients.[3] With information like this available, patients and health care providers can make better informed decisions. Better information can reassure patients and their families, and relieve their usual fear and stress in response to the uncertainty of undergoing surgery with co-morbidities. In addition, practitioners can quickly identify higher-risk patients and take these risks into consideration when providing treatment and follow-ups. Furthermore, managers can utilise clinical scores to perform needs assessments and to plan for resource allocation. For example, a scale for predicting length of hospital stay after primary total knee replacement based on the risk factors was verified in Hong Kong in 2017,[4] but its value reaches beyond just estimating the length of stay. The predictive factors also provide information on how the quality of health care can be improved if the factors are non-biological and controllable, such as urinary catheterisation in this case.

Further analysing the health outcomes of multiple treatment routines, clinical scores could be applied to estimate the health effect of a certain treatment and its alternatives for individual patients. This prediction power would be particularly valuable in complex conditions where differences in individual factors, such as pharmacokinetics, could play a significant role in affecting the outcome. For example, in a clinical trial in 2015, Mulvenna et al[5] found no significant difference in survival or quality-adjusted life years among 538 patients who received optimal supportive care only or additional whole-brain radiation therapy, suggesting the presence of very heterogeneous tumour behaviour. In contrast, a study of frameless stereotactic radiosurgery found that prognostic scoring identified patients who would benefit more from the treatment.[6] In the current development direction of personalised care, clinical scores could be used to enhance informed clinical decision making or as a transitional alternative for precision medicine.

A useful clinical prediction instrument not only helps improving patient care, but also reduces wasting health care resources owing to misdiagnosis. In the current issue of the *Hong Kong Medical Journal*, Cheung et al[7] have validated and refined the existing Ottawa subarachnoid haemorrhage (SAH) rule to improve its sensitivity for SAH diagnosis. The results of that study indicate the sensitivity of Ottawa SAH rule can be increased to 100% by adding two more predictors—vomiting and SBP >160 mm Hg—while retaining a specificity of 13.1%. The authors conclude that unnecessary costs (ie, 11.8% of computed tomographic scans in this study population) can likely be reduced.

Some caution is warranted when interpreting the performance of a clinical prediction instrument, and therefore its usefulness. Missing values are a common limitation for developing a clinical prediction rule, as acknowledged by Cheung et al.[7] Some patients might be positive for certain symptoms but be misclassified as negative due to missing values. Differential misclassification can cause the odds ratios of predictors (the symptoms) to be biased away from the null hypothesis, jeopardising the validity of symptoms found to be associated or not associated with a disease.[8] Caution is also needed when applying performance

metrics to a clinical prediction instrument. For example, 'accuracy' is a specific measure of ability of a predictive test in identifying cases from non-cases; one measure of accuracy involves dividing the sum of true positive and true negative results by the total population size. Using the study from Cheung et al[7] as an example, the prediction accuracy of the original Ottawa SAH rule was 39% (ie, [47+148]/500) which is higher than that of the modified Ottawa SAH rule (ie, [50+59]/500=21.8%). Thus, assessing the prediction performance based on multiple metrics are essential for judging the usefulness of a prediction rule. Last but not least, a useful clinical prediction tool should be subject to external validation, ie, with independent cohorts and data that have not been used in the model development.[9] This validation process is able to help examine the heterogeneousness of the model predictions, ie, whether it is reliable or accurate enough to be used in a wider population. Most proposed prediction models in the literature involve only internal validations; relatively few models have been through external validations, primarily because of a lack of data.[10] Future development and evaluations of clinical scores and risk factors should take such factors into consideration, and proposed models should be followed up with external validation. Under this framework, we anticipate that research and development on clinical scores and risk factors will be more useful in real-world settings. This may have an positive effect on patient care and clinical outcomes, such as patient survival and quality of life.

Declaration

As the statistical advisor of the *Hong Kong Medical Journal*, KC Chong was not involved in the peer review process of this article. Other authors have disclosed no conflicts of interest. All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Author contributions

SY Chan and KM Jia contributed to the concept of this article.

KC Chong drafted the manuscript and provided critical revision for important intellectual content.

References

1. Cheung HC, Leung KY, Choi CH. Diagnostic accuracy of spot urine protein-to-creatinine ratio for proteinuria and its association with adverse pregnancy outcomes in Chinese pregnant patients with pre-eclampsia. Hong Kong Med J 2016;22:249-55.
2. Kou KO, Poon CF, Kwok SL, et al. Effect of non-invasive prenatal testing as a contingent approach on the indications for invasive prenatal diagnosis and prenatal detection rate of Down's syndrome. Hong Kong Med J 2016;22:223-30.
3. Lau TW, Fang C, Leung F. Assessment of postoperative short-term and long-term mortality risk in Chinese geriatric patients for hip fracture using the Charlson comorbidity score. Hong Kong Med J 2016;22:16-22.
4. Lo CK, Lee QJ, Wong YC. Predictive factors for length of hospital stay following primary total knee replacement in a total joint replacement centre in Hong Kong. Hong Kong Med J 2017;23:435-40.
5. Mulvenna P, Nankivell M, Barton R, et al. Dexamethasone and supportive care with or without whole brain radiotherapy in treating patients with non-small cell lung cancer with brain metastases unsuitable for resection or stereotactic radiotherapy (QUARTZ): results from a phase 3, non-inferiority, randomised trial. Lancet 2016;388:2004-14.
6. Mok ST, Kam MK, Tsang WK, et al. Frameless stereotactic radiosurgery for brain metastases: a review of outcomes and prognostic scores evaluation. Hong Kong Med J 2017;23:599-608.
7. Cheung HY, Lui CT, Tsui KL. Validation and modification of the Ottawa subarachnoid haemorrhage rule in risk stratification of Asian Chinese patients with acute headache. Hong Kong Med J 2018;24:584-92.
8. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. Sources of Systematic Error or Bias: Information Bias. ERIC Notebook. 2nd ed. Chapel Hill (NC): The University of North Carolina at Chapel Hill; 2015.
9. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691-8.
10. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353:i3140.