

INTRODUCTION

Clinical scoring systems are algorithms designed to predict outcomes, aid decision making, support treatment options, manage clinical risk, or improve efficiency. The term clinical scoring system is known interchangeably as clinical decision rule, prediction algorithm, clinical prediction tool, risk score, or scoring tool.¹ Medicine is not short of these, with recent searches estimating that over 250 000 are available to use.² Despite the proliferation of algorithms to inform clinical care, this has not been matched with evidence of their utility. The uncertainty among clinicians about their efficiency and accuracy, alongside growing primary care workloads in the limited 10-minute consultation, may contribute to low utility. As demonstrated in the accompanying systematic review by Willis *et al*,³ there is the additional complication of multiple scores being available for the same condition. So how do we decide on whether to use a clinical score and what makes one better than another?

CLINICAL NEED AND CONTEXT

The first consideration is whether a score is needed or can be shown to provide clinical benefit. Scores should reduce uncertainty, prompt missed diagnoses, increase efficiency, and improve outcomes. If a score is not adding to clinical judgment, productivity, or improving outcomes then it becomes a laborious tick-box exercise. It also needs to be relevant to the clinical context in which it is applied. Despite thousands of scores having been developed, less than a handful of these are designed specifically for the UK primary care context. This is highlighted by the Centor and McIsaac criteria that are examined in the accompanying review.³ Both these scores are intended to inform users about the probability of a group A beta-haemolytic streptococcal (GABHS) pharyngitis. There are additional scoring systems for pharyngitis that haven't specifically been considered in the linked

"Despite the proliferation of algorithms to inform clinical care, this has not been matched with evidence of their utility."

review³ such as the Walsh and Fever, Pus, rapid Attendance (illness ≤ 3 days), severe Inflammation, and No cough or coryza (FeverPAIN) score.^{4,5} These should all guide antibiotic prescribing, improve symptom control, prevent complications, and reduce antibiotic use. However, most sore throats in the UK are caused by viral infections. Recent diagnostic cohorts suggest that 34%–40% of patients presenting with sore throats will have pathogenic streptococci, with two-thirds of these being GABHS.⁵ Related complications are rare, with peritonsillar abscesses accounting for $<0.2\%$ of all adults presenting with sore throat and rheumatic fever rates being even lower.^{6,7} Moreover, treatment of GABHS with antibiotics may not prevent suppurative complications such as peritonsillar abscesses.⁸ The ability of the Centor, McIsaac, or Walsh scores to reduce antibiotic prescribing for sore throats is marginal; two trials have demonstrated reduced antibiotic prescribing while two also showed no effect.^{5,9} In terms of symptom control or symptom duration, the Centor and McIsaac scores have not been shown to impact clinical outcomes. Scores that have high sensitivity and specificity to predict outcomes with high levels of discrimination are more likely to be useful.

USER-FRIENDLY

Further considerations might take into account the user-friendliness of a score. These should include routinely recordable variables that can be easily applied to an algorithm. The need for additional measures, such as those from blood results that are not immediately available,

can be perceived as impractical given the workload and through-flow of patients in primary care. Willis *et al* in the linked review³ suggest in their conclusions that additional point-of-care testing might be helpful after the Centor or McIsaac score to distinguish GABHS. A score that requires additional testing is unlikely to be seen as pragmatic by a busy GP. There is also no evidence that the addition of a rapid streptococcal test to the Centor, McIsaac, or FeverPAIN score can improve appropriate use of antibiotics or symptom control.⁵ Apart from inefficiency, it could also be perceived as over-medicalisation of sore throats. This is because carrying out a test for sore throats would suggest to patients that testing is required when in fact there is limited evidence that it adds much to the clinical decision.¹⁰ There is some evidence that simple scores, such as FeverPAIN, without additional testing are more likely to be used by GPs. A free-standing FeverPAIN app has been developed and used several hundred thousand times to date. This is probably because the score is simple and user-friendly, with an easy output to facilitate discussion.

GENERALISABLE

Scoring algorithms need to be able to manage the heterogeneity of patients that are served by primary care. Most scores have been developed and tested in limited cohorts. The availability of big databases with large sample sizes that capture individual-level patient data permit new opportunities to evaluate existing scores and develop new ones within diverse populations. For example, the Centor and McIsaac scores have now been validated in a large dataset of 206 870 participants across a range of areas and sociodemographic backgrounds; however, this is in the US and may not be generalisable to other healthcare systems.¹¹ The increased diversity of patient demographics in UK primary care has been matched by increased heterogeneity in

"Scoring algorithms need to be able to manage the heterogeneity of patients that are served by primary care. Most scores have been developed and tested in limited cohorts."

“Scoring systems are primarily focussed on the consideration of biological measures and symptoms of a disease without allowing for the integration of the patient perspective.”

disease and presentations. Multimorbidity now affects one in four people in the UK and contributes to additional complexity in clinical decision-making.¹² Most existing scores, such as those described above, have been designed to consider a single disease within a single episode (such as, a sore throat in one presentation). Primary care practitioners are increasingly negotiating clinical decisions that traverse multiple concurrent physical, psychological, and social problems. Algorithms alone rarely consider the impact of conflicting advice from a concurrent disease, treatment interaction, and the effect of fatigue on patients due to multiple competing health demands such as polypharmacy or attendance to appointments.

PATIENT-CENTRED

Scoring systems are primarily focussed on the consideration of biological measures and symptoms of a disease without allowing for the integration of the patient perspective. Neither the Centor, McIsaac, or FeverPAIN score include any measure of patient perspective or preferences for treatment. For example, a Centor score might recommend antibiotics, but a patient could have a strong preference against these due to potential side effects such as nausea or diarrhoea. Another example beyond sore throats could be a patient with atrial fibrillation who has a CHA₂DS₂-VASc score that might recommend anticoagulation. A patient could perceive regular international normalised ratio blood testing as a reduced quality of life that might outweigh any potential benefits. Most scores do not predict or take into account the impact of recommendations on functional status or quality of life. This balance between clinical recommendations and patient preferences are rarely captured or explored within scoring systems.

CONCLUSION

Clinical scoring systems do have a role for use in primary care but cannot replace clinical reasoning and judgement. They risk being overly burdensome on the clinician

for limited additional benefit. A previous systematic review suggests that clinical algorithms are rarely superior to clinical judgement.¹³ This is particularly relevant in primary care where clinicians are required to use their clinical experience to collate multiple points of information, balance risk and benefit, and then integrate the patient perspective in a holistic way. Scores are rarely able to consider a patient in totality and tend to be more useful with specific acute illness. Further work is needed to understand the challenges and practicalities of balancing the additional workload burden compared to the potential benefit in everyday primary care practice for both clinicians and patients.

Hajira Dambha-Miller,

GP and National Institute for Health Research Clinical Lecturer in General Practice, Department of Primary Care and Population Health, University of Southampton, Southampton.

Hazel Everitt,

Professor of Primary Care, Department of Primary Care and Population Health, University of Southampton, Southampton.

Paul Little,

Professor of Primary Care, Department of Primary Care and Population Health, University of Southampton, Southampton.

Provenance

Freely submitted; externally peer reviewed.

Competing interests

Paul Little developed the FeverPAIN score for sore throats. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. Hazel Everitt has declared no competing interests.

DOI: <https://doi.org/10.3399/bjgp20X708941>

ADDRESS FOR CORRESPONDENCE

Hajira Dambha-Miller

Department of Primary Care and Population Health, University of Southampton, Southampton SO16 5ST, UK.

Email: hajiradambha@doctors.org.uk

REFERENCES

1. Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagn Progn Res* 2019; **3**: 16.
2. Challener DW, Prokop LJ, Abu-Saleh O. The proliferation of reports on clinical scoring systems: issues about uptake and clinical utility. *JAMA* 2019; **321**(24): 2405–2406.
3. Willis BH, Coommar D, Baragilly M. Comparison of Centor and McIsaac scores in primary care: a meta-analysis over multiple thresholds. *Br J Gen Pract* 2020; DOI: <https://doi.org/10.3399/bjgp20X708833>.
4. McGinn TG, Deluca J, Ahlawat SK, et al. Validation and modification of streptococcal pharyngitis clinical prediction rules. *Mayo Clin Proc* 2003; **78**(3): 289–293.
5. Little P, Hobbs FDR, Moore M, et al. PRISM study: In vitro study, diagnostic cohorts and a pragmatic adaptive randomised controlled trial with nested qualitative study and cost-effectiveness study. *Health Technol Assess* 2014; **18**(6): vii–xxv, 1–101.
6. Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. *Clin Epidemiol* 2011; **3**: 67–84.
7. Little P, Stuart B, Richard Hobbs FD, et al. Predictors of suppurative complications for acute sore throat in primary care: prospective clinical cohort study. *BMJ* 2013; **347**: f6867.
8. Little P, Moore M, Kelly J, et al. Delayed antibiotic prescribing strategies for respiratory tract infections in primary care: pragmatic, factorial, randomised controlled trial. *BMJ* 2014; **348**: g1606.
9. Sanders SL, Rathbone J, Bell KJL, et al. Systematic review of the effects of care provided with and without diagnostic clinical prediction rules. *Diagn Progn Res* 2017; **1**: 13.
10. Chiappini E, Regoli M, Bonsignori F, et al. Analysis of different recommendations from international guidelines for the management of acute pharyngitis in adults and children. *Clin Ther* 2011; **33**(1): 48–58.
11. Fine AM, Nizet V, Mandl KD. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. *Arch Intern Med* 2012; **172**(11): 847–852.
12. Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012; **380**(9836): 37–43.
13. Sanders S, Doust J, Glasziou P. A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. *PLoS One* 2015; **10**(6): e0128233.