COMMENT



IAPN: a simple framework for evaluating whether a population-based risk stratification tool will be cost-effective before implementation



Steven Wyatt^{1*}, Mohammed A. Mohammed^{1,2} and Peter Spilsbury¹

Abstract

Risk prediction tools are widely used in healthcare to identify individuals at high risk of adverse events who may benefit from proactive interventions. Traditionally, these tools are evaluated primarily on statistical performance measures—such as sensitivity, specificity, discrimination, and positive predictive value (PPV)—with minimal attention given to their cost-effectiveness. As a result, while many published tools report high performance statistics, evidence is limited on their real-world efficacy and potential for cost savings. To address this gap, we propose a straightforward framework for evaluating risk prediction tools during the design phase, which incorporates both PPV and intervention effectiveness, measured by the number needed to treat (NNT). This framework shows that to be cost-effective, the per-unit cost of an intervention (I) must be less than the average cost of the adverse event (A) multiplied by the PPV-to-NNT ratio: I < A*PPV/NNT. This criterion enables decision-makers to assess the economic value of a risk prediction tool before implementation.

Keywords Risk prediction, Positive predictive value, Number needed to treat, Cost, Intervention, Design stage evaluation

Background

Healthcare systems face significant challenges in managing rising demand for services and the high costs of new technologies within limited budgets. These pressures are especially pronounced in unplanned care, where increasing emergency department (ED) visits are often viewed as a consequence of unmet patient needs in primary care [1-3]. Unplanned admissions alone cost the English

*Correspondence:

Steven Wyatt

swyatt@nhs.net

¹ The Strategy Unit - NHS Midlands and Lancashire Commissioning Support Unit, Birmingham, UK

² Faculty of Health Studies, University of Bradford, Bradford, UK



National Health Service (NHS) approximately £11 billion annually [4].

The prevailing belief is that timely and appropriate interventions could prevent many of these unplanned episodes and their associated costs. This raises a critical question: how can we identify patients at high risk of adverse events, such as unplanned admissions, to enable earlier interventions and mitigate both the event and its costs? This need underpins the development of numerous population-based risk prediction tools [2, 5] that are widely used in primary care to guide proactive, preventative care.

Risk prediction tools

A recent systematic review [3] identified 28 prediction and stratification tools, such as the Patients At Risk of

© The Author(s) 2024. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Rehospitalisation algorithm [6], the Predictive RIsk Stratification Model, the Nairn Case Finder and the QAdmissions score [7]. The quality of risk prediction tools is usually assessed via a set of statistical performance metrics [8] such as sensitivity, specificity and discrimination described in Table 1.

For instance, the ability of a risk prediction tool to discriminate between patients who do and do not experience the outcome (eg unplanned admission) is a key indicator of performance and is denoted by the *c*-statistic which ranges from 0 to 1; where a value 0.5 is no better than tossing a coin, whilst perfect discrimination has a *c*-statistic of 1. Thus, the higher the *c*-statistic the better the risk prediction tool. In general, values less than 0.7 are considered to show poor discrimination, values of 0.7-0.8 can be described as reasonable, and values above 0.8 suggest good discrimination. A recent review [3] of 28 risk prediction tools reported *c*- statistics that ranged 0.67 to 0.90, although only half of these were externally validated. (Internal validation is where the discrimination of the risk prediction tool is evaluated within the same population in which the model was derived, and external validation uses data from a separate population and is therefore a more stringent test).

The Positive Predictive Value (PPV) of a risk prediction tool is a metric that indicates how likely it is that a person with a positive test result actually has the condition being tested for. It's a measure of the tools accuracy in correctly identifying true positive cases among all positive results. The PPV is crucial because it helps assess the reliability of a positive test result, which can impact clinical decisionmaking and patient care. In general, a risk tool with a higher PPV is preferable. The more common the adverse event in the target population, the higher the PPV tends to be. If the adverse event is rare, the PPV will generally be lower.

Evaluation of risk prediction tools

Whilst the statistical performance of risk prediction tools is often well reported [3], the extent to which they impact on improving outcomes and reduce costs are reported infrequently [1, 3, 4]. One such exception is a well-designed and executed randomised stepped-wedge trial in primary care that measured the effects on service usage, costs, mortality, quality of life and satisfaction of deploying a risk stratification tool, known as Prism, designed to reduce ED usage for use in primary care (32 general practices, 230000 patients) [9]. The intervention was the provision of the risk prediction tool along with training and support for staff in general practices. The primary results showed increases, not decreases, in unplanned admissions, ED attendances and overall healthcare costs.

Indeed, the recent systematic review [3] noted that "The results of real-world evaluation studies present equivocal evidence for the efficacy of these population level interventions. The majority of publications reported no change, or indeed significant increases, in healthcare utilisation within groups targeted by the intervention, with only one-third of reports demonstrating some benefit." [3]. The review concluded that "...there is little evidence to suggest that the identification of high-risk individuals can be translated to improvements in service delivery or morbidity. The available evidence does not support further integration of these types of risk prediction into population healthcare pathways. There is an urgent need to independently appraise the safety, efficacy

 Table 1
 Some of the metrics typically used to report the performance of a risk prediction tool applied to unplanned hospital admission

		True outcome	
		Admitted to hospital	Not admitted to hospital
Predicted outcome	ligh-risk	True positive (TP)	False positive (FP)
L	.ow-risk	False negative (FN)	True negative (TN)
Performance metric formula	Description		
Accuracy=(TP+TN)/(TP+FP+FN+TN)	Accuracy measures how well the risk prediction tool identifies people who were and were not admitted to hospital		
Sensitivity (aka recall) =TP/(TP + FN)	The proportion of high-risk people who were admitted to hospital		
Specificity = $TN/(TN + FP)$	The proportion of low-risk people who were not admitted to hospital		
Positive predictive value = $TP/(TP + FP)$	The proportion of high-risk people who were admitted to hospital		
Negative predictive value = $TN/(TN + FN)$	The proportion of low-risk people who were not admitted to hospital		
Concordance statistic (aka <i>c</i> -statistics or area under receiver operating characteristic):	The probability that a randomly selected person who was admitted to hospital will have a higher mod- elled probability of admission than a randomly selected person who was not admitted to hospital		

and cost-effectiveness of risk prediction systems that are already widely deployed within primary care." [3].

Whilst such empirical evidence is crucial to scientific progress, it is, ironically, relatively late in the day to discover such an antithetical result. It would be useful to find a way to fail faster and safely; by determining the extent to which a risk prediction tool is likely to succeed, at the design stage, before implementation. To address this gap, we propose a straightforward framework for evaluating risk prediction tools at the design stage before implementation.

Desing stage evaluation of risk prediction tools

We propose a straightforward framework for evaluating risk prediction tools during the design phase that integrates both the positive predictive value (PPV) of the tool and the effectiveness of the subsequent intervention. Intervention effectiveness is summarised by the Number Needed to Treat (NNT), which represents the average number of patients who need to receive the intervention to prevent one additional adverse event. A lower NNT, closer to 1, indicates a highly effective intervention, as fewer people need to be treated to achieve a positive outcome for one individual. Higher NNT values suggest less effective treatments, with typical reported ranges from 10 to 100.

Our framework is grounded in the understanding that the real-world impact of risk prediction tools in a given population depends on both the statistical accuracy of the risk prediction tool and the cost-effectiveness of the interventions they guide. Therefore, a comprehensive evaluation of a risk prediction tool's utility requires consideration of its statistical performance—particularly PPV—alongside the cost and effectiveness of the intervention and the cost of the adverse event. This framework, which aligns with approaches suggested by others [10], combines PPV and NNT to provide a practical assessment of whether a given risk prediction tool can deliver meaningful, cost-effective outcomes before it is implemented in a given population. We refer to this as a design-stage evaluation, summarized in Box 1.

Taking a population perspective, we focus on PPV because it indicates the proportion of high-risk individuals who are likely to experience the adverse event, fulfilling the primary purpose of risk models in practice. A higher PPV enhances the tool's accuracy in identifying individuals at risk of the adverse event. However, prediction alone does not alter outcomes—an effective intervention is necessary to achieve this. Intervention effectiveness is captured by the NNT, while cost considerations involve comparing the unit cost of the intervention (I) against the average cost of the adverse event (A). For an intervention to be cost-saving, its unit cost (I)

must be less than the average cost of the adverse event (A) multiplied by the PPV-to-NNT ratio, expressed I < A*PPV/NNT (see Box 1).

Box 1 Evaluation of risk prediction tools based on the PPV of the tool, the NNT and cost (I) of the intervention and the cost (A) of the adverse event. (IAPN)

- **PPV** is the positive predictive value of a risk prediction tool which aims to identify patients who will have an adverse event in a given period;
- **NNT** is the average number of people that need to receive the intervention in order to avoid one adverse event
- I is the unit cost of an intervention to prevent an adverse event which is delivered to those identified by the predictive risk tool then,
- A is the average cost of an adverse event, then
- I < A*PPV/NNT for the intervention to save money.

We use the acronym IAPN to refer this framework. We apply this approach to a hypothetical scenario below.

Hypothetical illustration

Consider a hypothetical general practice with 5000 patients, where a risk prediction tool is to be used to identify the top 2% (n=100) of patients at risk of an unplanned admission to hospital in the next 12 months. The scenario is not untypical of risk prediction tools to avoid unplanned hospital admissions [7, 11]. Figure 1 shows this worked example.

The PPV of the risk prediction tool in this top 2% is reported as 36%, in other words 36 of the 100 identified patients would be expected to experience an unplanned admission (64 of the 100 would not). These indicative values (top 2%, PPV = 36%) for the risk prediction tool are not dissimilar to what is reported in practice [12]. Nevertheless, in this scenario, all 100 identified patients would be subject to an intervention designed to reduce the risk of an unplanned admissions. Let us imagine an intervention with an NNT of 18 (ie for every 18 identified people treated who would otherwise have been admitted, 1 unplanned admission would be avoided). So, of the 36 patients who go on to experience the event, our intervention would avoid 2 such events. Let us assume that an unplanned admission costs on average £2000. To save money, our upstream intervention must cost less than £40 (= $2 \times 2000/100$) per patient.

Figure 2 illustrates how the cost per identified patient varies with PPV (10% to 100%) and NNT (1–200) values for our worked example. The general message is that the



Fig. 1 Worked example demonstrating the evaluation of a hypothetical risk prediction tool, using the PPV of the risk prediction tool, the NNT of an upstream intervention and the costs associated with the intervention and the adverse event

lower the NNT (ie more effective interventions) the more we can afford to pay per identified patient for a given PPV and that impact of improvements in PPV become more pronounced with more effective interventions (lower NNTs).

Discussion

We, along with others [10], offer this simple IAPN framework as a practical approach that enables decision makers to assess the potential of risk prediction tools to succeed in practice. High quality decision making requires access to relevant and reliable information. Our proposed framework does this by combining the PPV with the NNT and shows that to save money, the unit cost of an intervention (I) must be less than the average cost of the adverse event (A) multiplied by the ratio of the PPV/NNT. (I < A*PPV/NNT).

There are a few caveats to our illustrative example. We have not included the preliminary costs of developing and deploying the risk prediction tool in IT systems because these are generally considered to be much lower than the cost of using them to intervene to reduce adverse healthcare outcomes [9]. Where these preliminary costs are available and deemed material, they may be incorporated into the calculus. We used a single PPV but changing the risk threshold for defining low and high-risk patients by focusing on say the top 5% (or 1%) instead of the top 2% of cases would induce a lower (or higher) PPV. Furthermore, the recognition that not all high-risk patients are amenable to avoiding the adverse outcome, has led to approaches to identify subsets of atrisk patients for whom the intervention is expected to be more successful [13]. Such "impactibility" based models are also subject to the formula described in box 1.

The NNT was promoted in clinical decision making over 25 years ago [11, 14] and is now widely used but has attracted criticism [15–17]. For example, the NNT is heavily influenced by the baseline risk in the population being studied (e.g., how likely the event is to occur without treatment). In high-risk populations, the NNT will generally be lower, whereas it may be much higher in low-risk populations. Many of the criticism relate to the use of the NNT for clinical decision making in individual patients, whereas we use the NNT alongside the PPV for a given population (not individual clinical decision making).

Despite the above caveats, we strongly recommend that all risk prediction schemes, pending, current or future, should undergo evaluation using this IAPN framework. In our worked example the NNT is 18 with a cost of £40



Maximum cost of upstream intervention

given NNT of intervention, PPV of risk tool and adverse event costing £2000

Fig. 2 Maximum cost (in circles) of an upstream intervention (£) per identified patient for a range of PPVs and NNTs where the adverse event costs £2000

per patient. This is a critical issue to make transparent to decision makers (and other stakeholders) who then need to make explicit their degree of belief around such an intervention and its cost whilst noting the tendency for optimism bias. As shown in Fig. 1, we suggest that decision makers are supplied with NNTs, and associated costs for a range of comparable interventions, to help calibrate their judgements whilst recognising that further refinements could be made by incorporating statistical uncertainty around the terms in the IAPN equation.

Acknowledgements

Not applicable.

Author contributions

SW developed the approach. MAM drafted the paper. PS provided critical guidance and support. All authors contributed to the final manuscript.

Funding

None.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 September 2024 Accepted: 10 November 2024 Published online: 04 December 2024

References

- Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T, Smith SM. Risk prediction models to predict emergency hospital admission in communitydwelling adults: a systematic review. Med Care. 2014;52(8):751–65.
- Georghiou T, Steventon A, Billings J, Blunt I, Lewis G, Bardsley M. Predictive risk and health care: an overview. London: The Nuffield Trust; 2011.
- Oddy C, Zhang J, Morley J, et al. Promising algorithms to perilous applications: a systematic review of risk stratification tools for predicting healthcare utilisation. BMJ Health Care Inform. 2024;31:e101065. https:// doi.org/10.1136/bmjhci-2024-101065.
- 4. Lewis G, Curry N, Bardsley M. Choosing a predictive risk model: a guide for commissioners in England. London: Nuffield trust; 2011. p. 20.

- Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ. 2006;333(7563):327.
- Billings J, Blunt I, Steventon A, Georghiou T, Lewis G, Bardsley M. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ Open. 2012;2(4):e001667.
- Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. BMJ Open. 2013;3(8):e003482.
- Monaghan TF, Rahman SN, Agudelo CW, Wein AJ, Lazar JM, Everaert K, Dmochowski RR. Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. Medicina. 2021;57(5):503.
- Snooks H, Bailey-Jones K, Burge-Jones D, Dale J, Davies J, Evans B, et al. Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC). Health Serv Deliv Res. 2018. https://doi.org/10. 3310/hsdr06010.
- Liu VX, Bates DW, Wiens J, Shah NH. The number needed to benefit: estimating the value of predictive analytics in healthcare. J Am Med Inform Assoc. 2019;26(12):1655–9.
- Bottle A, Aylin P, Majeed A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. J R Soc Med. 2006;99(8):406–14.
- 12. Wennberg D, Dixon J, Billings J. Combined Predictive Model–Final Report. 2006 https://www.kingsfund.org.uk/sites/default/files/field/field_docum ent/PARR-combined-predictive-model-final-report-dec06.pdf.
- Lewis GH. "Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. Milbank Q. 2010;88(2):240–55.
- 14. Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! BMJ Evidence Based Med. 1996;1(6):164.
- 15. Dowie J. The 'number needed to treat'and the 'adjusted NNT'in health care decision-making. J Health Serv Res Policy. 1998;3(1):44–9.
- McAlister FA. The "number needed to treat" turns 20—and continues to be used and misused. CMAJ. 2008;179(6):549–53.
- 17. Hutton JL. Number needed to treat: properties and problems. J R Stat Soc A Stat Soc. 2000;163(3):381–402.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.