The Strategy Unit.

Is risk stratification likely to improve the use of NHS resources?

October 2021

NHS
Midlands and Lancashire
Commissioning Support Unit

Document control

Document Title	Is risk stratification likely to improve the use of NHS resources?
Written by	Steven Wyatt, Mohammed A Mohammed, Shammas Rahim, Peter Spilsbury
	The authors declare no conflicts of interest and no specific funding for this paper.
Date	13th October 2021

Contact:

Tel: 0121 612 1538

Email: strategy.unit@nhs.net

Web: www.strategyunitwm.nhs.uk

Twitter: @strategy_unit

Is using a risk prediction tool likely to be worth it?

We think there is a simple, insightful way for commissioners to determine if a risk prediction tool might work and save money before using it. But we suspect that decision makers don't currently think in these terms. This paper sets out our arguments, and we invite others to contribute.

Abstract

Risk prediction tools are ubiquitous in healthcare. These tools are often used to identify individuals at high risk of an adverse event and who may benefit from an upstream intervention. The assessment of risk prediction tools usually revolves around performance statistics such as sensitivity, specificity, discrimination and positive predictive value (PPV).

Published risk prediction tools are usually reported to have good performance statistics but with little guidance on how to commission such tools. Despite their reported performance, there is a paucity of evidence on the extent to which risk prediction tools work in practice and save money.

Whilst such studies may be challenging to undertake, we offer a simple framework for evaluating the plausibility of success at the design stage. This design stage evaluation of risk prediction tools combines the PPV with the number needed to treat (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).

Background

Healthcare systems are facing major challenges managing increased demand for care and the costs of new technologies within constrained resources. Pressure in the field of unplanned care is particularly marked with the increase in unplanned use of emergency departments (EDs) often being seen as a direct failure to address the needs of patients upstream in primary care[i,ii].

It is argued that if the correct intervention had been provided, then subsequent unplanned episodes of care and associated costs could have been avoided.

A key question then, is how can we identify patients at risk of an adverse binary outcome or event (e.g. death, unplanned admission), so that we might intervene earlier to mitigate the adverse outcome and its associated costs?

Risk prediction tools

This is the rationale for numerous risk prediction tools[ii,iii]. (Although use is so widespread that this rationale is often forgotten). Examples in primary and secondary care include cardiovascular risk

prediction (using the Framingham risk prediction tool[iv] and QRisk2[v]), sepsis risk prediction (using SIRS and qSOFA[vi]) and the risk of readmission to hospital (using the Combined Predictive Model[vii] and PARR-30[viii]).

The quality of risk prediction tools is usually assessed via a set of statistical performance metrics such as sensitivity, specificity and discrimination as described in Table 1.

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	**High-risk**	True Positive (TP)	False Positive (FP)
	Low-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 *c*-statistics or area under receiver operating characteristic):		The probability that a randomly selected person who was admitted to hospital will have a higher modelled probability of admission than a randomly selected person who was not admitted to hospital.	

For instance, the ability of a risk prediction tool to discriminate between patients who do and do not experience the adverse outcome (e.g. death or 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 and the 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.

The *c*-statistic is often used to compare different risk prediction models. However, the PPV is also very useful because it tells us the proportion of high-risk people who will experience the event – which is the primary purpose of the risk model in practice and therefore a crucial performance metric. The higher the PPV, the better the risk prediction tool is at picking out people who will experience the adverse event.

Whilst the statistical performance of risk prediction tools is well reported, there is little guidance on how to commission such tools and the extent to which they impact improved outcomes and reduced costs are infrequently reported[ix,x]. Although such studies may be challenging to undertake, they are nevertheless crucial in determining the extent to which the risk prediction tool has served its purpose.

One notable 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, 230,000 patients)[xi]. 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.

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 roll out.

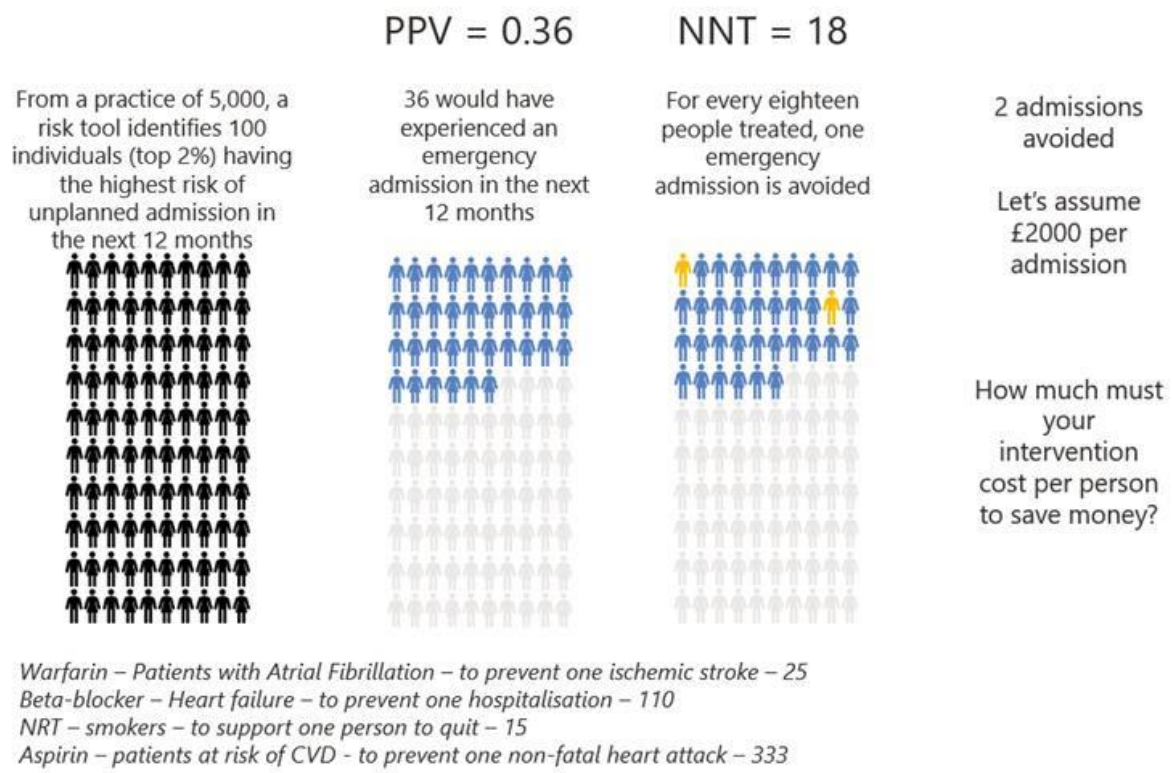
Design stage evaluation of risk prediction tools

We propose a simple design stage evaluation framework for assessing the impact of risk prediction tools which combines the PPV of risk prediction tools with the number needed to treat upstream interventions and the cost of the adverse event and upstream intervention.

The Figure below provides a worked example. It demonstrates the evaluation of a hypothetical risk prediction tool, using the PPV of the risk prediction tool, the Number Needed to Treat (NNT)[1] of an upstream intervention and the costs associated with the adverse event and intervention.

[1] The Number Needed to Treat is the average number of people that would need to receive the intervention to prevent one additional negative outcome.

A worked example



Warfarin – Patients with Atrial Fibrillation – to prevent one ischemic stroke – 25
Beta-blocker – Heart failure – to prevent one hospitalisation – 110
NRT – smokers – to support one person to quit – 15
Aspirin – patients at risk of CVD - to prevent one non-fatal heart attack – 333

Consider a hypothetical general practice with 5,000 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 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 for the risk prediction tool are not dissimilar to what is reported in practice[xii]. 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 now imagine an intervention with no downside risk and with an NNT of 18 (i.e. 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 further us assume that an unplanned admission costs on average £2,000. In order to save money, our upstream intervention must cost less than £40 (n = 2 x 2000/100) per patient. Decision makers may now factor this insight into their decision to proceed. (The box below shows the general equation which relates PPV, NNT, the costs of adverse events and upstream interventions).

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[ix]. Nevertheless, 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 at-risk patients for whom the intervention is expected to be more successful[xiii]. Such 'impactibility' based models are also subject to the formula described in box 2.

Box 2: The relationship between the cost of the upstream intervention, the cost of the adverse event, PPV and NNT

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

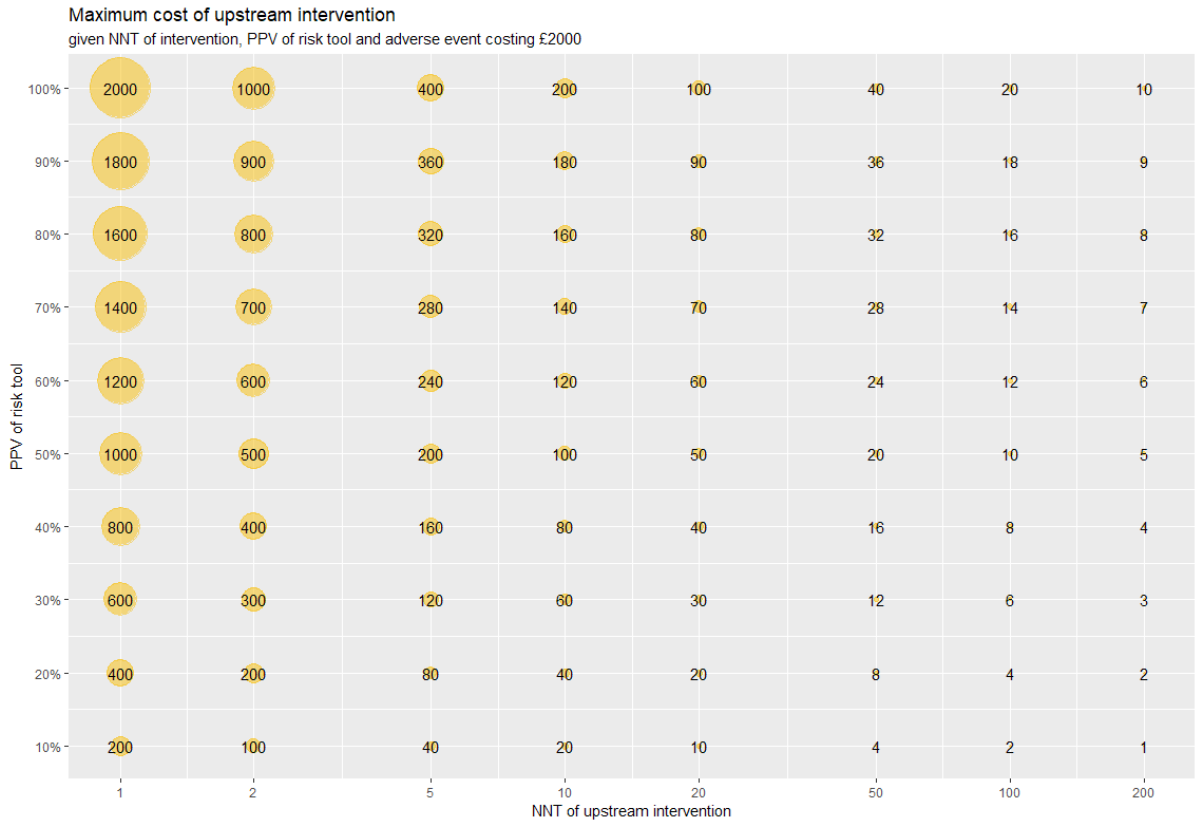
The plausibility of an upstream intervention with an NNT of 18 having a cost of £40 per patient is a critical issue to make transparent to decision makers before commissioning any risk prediction tool. Decision makers need to make explicit their degree of belief around such an intervention and its cost while noting the tendency for optimism bias.

As shown in Box 1, we suggest that decision makers are supplied with NNTs and associated costs for a range of comparable interventions to help calibrate their judgements while further refinements could be made by incorporating statistical uncertainty around these estimates.

Figure 1 illustrates how the cost per identified patient varies with PPV and NNT values for our worked example. The general message is that the lower the NNT (i.e. more effective interventions),

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

Figure 1 – Maximum cost of an upstream intervention (£) per identified patient for a range of PPVs and NNTs where the adverse event costs £2,000.



In summary, we offer this simple framework as a practical tool to enable decision makers to assess the potential of risk prediction tools at the design stage. We welcome feedback on its use and continual efforts to improve it.

[i] Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T, Smith SM. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. Medical care. 2014 Aug;52(8):751.
[ii] Nuffield Trust (2011a) *Predictive Risk and Health Care: An overview*. London: Nuffield Trust.
[iii] Billings J, Dixon J, Mijanovich T and Wennberg D (2006) 'Case finding for patients at risk of readmission to hospital:development of algorithm to identify high risk patients', *BMJ* 333, 327.
[iv] Lloyd-Jones DM, Wilson PW, Larson MG, Beiser A, Leip EP, D'Agostino RB, Levy D. Framingham risk score and prediction of lifetime risk for coronary heart disease. The American journal of cardiology. 2004 Jul 1;94(1):20-4.
[v] Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. Bmj. 2012 Jun 21;344:e4181.
[vi] Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, Pilcher DV. Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. Jama. 2017 Jan 17;317(3):290-300.
[vii] Wennberg D and others (2006) *Combined Predictive Model: Final report and technical documentation*. London: The King's Fund.
[viii] Nuffield Trust (2011b) *Predicting Risk of Hospital Readmission with PARR-30.* Available at www.nuffieldtrust.org.uk/ourwork/projects/predicting-risk-hospitalreadmission-parr-30
[ix] Lewis G, Curry N, Bardsley M. Choosing a predictive risk model: a guide for commissioners in England. London: Nuffield Trust. 2011 Nov 17;20.
https://imperialcollegehealthpartners.com/wp-content/uploads/2018/07/Choosing-a-predictive-model.pdf
[x] Wallace E, Stuart E, Vaughan N, Bennett K, Fahey T, Smith SM. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. Medical care. 2014 Aug;52(8):751.
[xi] 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;6(1).
[xii] Wennberg D, Dixon J, Billings J. Combined Predictive Model–Final Report. (2006) https://www.kingsfund.org.uk/sites/default/files/field/field_document/PARR-combined-predictive-model-final-report-dec06.pdf).
[xiii] Lewis G (2010) '"Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs', Milbank Q 88(2), 240–255.