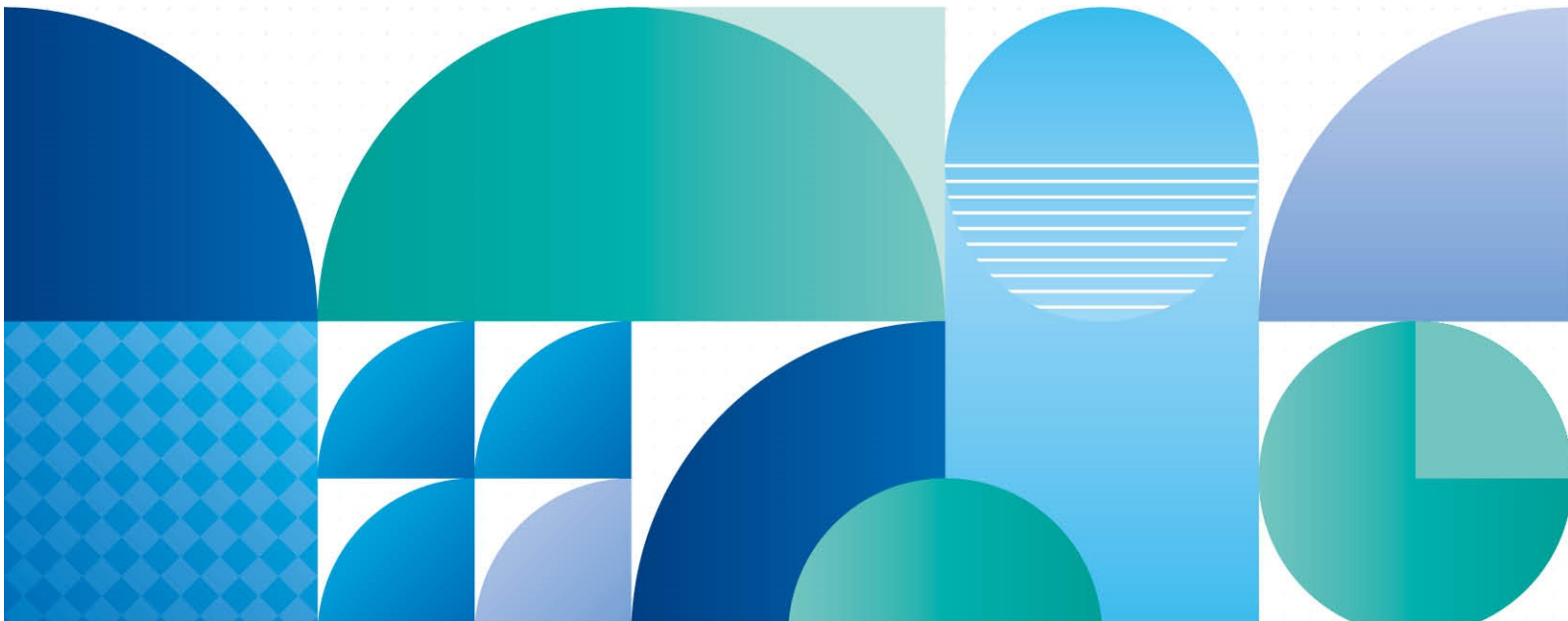
Health Economics Unit

NHS
Midlands and Lancashire
Commissioning Support Unit

Risk stratification

A ‘how to’ guide



Health Economics Unit

Table of contents

1 Background

Healthcare systems are facing major challenges in managing increased demand for care and the costs of new technologies with constrained resources. These are compounded by unplanned hospital admissions, which are generally more costly than planned admission and a proportion of which are preventable. The triple aim of healthcare, in which the goals of improving the individual experience of care, improving the health of populations, and reducing the per-person cost of care are linked,[1] has become a popular healthcare model. NHS England has adopted this strategy as part of its population health management (PHM) approach in its Long Term Plan.[2] PHM focuses on identifying key outcomes for specific groups that have a mix of shared and varying characteristics over and above diagnosis,[3] with the aim of identifying opportunities to improve their care. Additionally, it can inform planning and investment for a variety of interventions.[4] Lastly, it puts in place a framework against which to monitor and evaluate the effectiveness of interventions.

Risk stratification is a method of assessing the potential scale of future adverse events among patients at high, medium, and low risk. By identifying these groups, health planning may be adapted to meet their needs by providing interventions to avoid these adverse events happening. There are three main approaches to risk stratification (Table 1). As the first two methods of clinical judgement and threshold modelling, are known to have limited effectiveness, this guide focuses on the use of predictive modelling.

Table 1: Common types of health risk stratification methods

Approach	Description	Pros	Cons	Accuracy
Clinical judgement	Clinicians use medical knowledge and training combined with knowledge of patients to identify individuals at high risk; often adopted due to a belief (generally unfounded) that tools are difficult to use and require information not readily available	• Can help to identify patients most likely to benefit from treatment (impactibility)[5] • Widely used and accepted by clinicians	• Accuracy is lower than methods that can consider larger populations regularly and repeatedly[6] • Predictions cannot include patients clinicians do not see • Decisions are susceptible to cognitive bias	Overall, predictive value is very low beyond the individual patient

[1] Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood). 2008;27(3):759-69. doi: 10.1377/hlthaff.27.3.759.
[2] NHS England. The NHS Long Term Plan. Jan 7 2019. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf
[3] Buck D, Baylis A, Dougall D, Robertson R. A vision for population health. Towards a healthier future. November 2018. https://www.kingsfund.org.uk/sites/default/files/2018-11/A%20vision%20for%20pop%20health%20summary%20online%20version.pdf
[4] Integrating care for high-risk patients in England using the virtual ward model: lessons in the process of care integration from three case sites
[5] Paton F, Wilson P, Wright K. Predictive validity of tools used to assess the risk of unplanned admissions: A rapid review of the evidence. York: University of York, 2014.
[6] Allaudeen N, Schnipper JL, Orav EJ, Wachter RM, Vidyarthi AR. Inability of providers to predict unplanned readmissions, J Gen Intern Med 2011;26:771-76. doi: 10.1007/s11606-011-1663-3.

Threshold modelling	A rule-based 'catch-all' method that identifies any individual who meets a defined high-risk threshold	• Can use readily available data • Takes minimal time and data resources	• More accurate at identifying individuals at historic risk than at future risk • Likely to select patients after period of greatest need, meaning risk might reduce even without prevention interventions • Risk of regression towards the mean (extreme values move towards the average on second measurement)	Accuracy very limited, as even one extreme value can lead to predictions no better than chance[7]
Predictive modelling	Based on statistical analysis of multiple past characteristics of patients (e.g., clinical, pharmacy, costs, and sociodemographic)[8]	• Roughly twice as accurate as threshold modelling, although accuracy depends on data, statistical technique used, and event being predicted • Not restricted to using clinical data, improving relevance to triple aim • Multiple tools are available	• Tools might need licences, incurring costs • Might require analytical support (can be provided by Commissioning Support Units and other organisations*) • Limited prediction of impactibility	Most accurate method as risk of cognitive bias and regression to the mean are reduced but still does not predict who is most likely to benefit

*For example, Association of Professional Healthcare Analysts or private companies.

[7] Roland M, Abel G. Reducing emergency admissions: are we on the right track? BMJ 2012;345:e6017. doi: 10.1136/bmj.e6017.
[8] Curry N, Billings J, Darin B, Dixon J, Williams M, Wenngerg D. Predictive risk project. Literature review. https://www.kingsfund.org.uk/sites/default/files/field/field_document/predictive-risk-literature-review-june2005.pdf

Risk stratification predictive modelling tools are widely used, and many have been developed, often for specific diseases. For example, healthcare providers might be familiar with the Framingham risk prediction tool and QRisk2 for cardiovascular care, SIRS and qSOFA for sepsis, and the Combined Predictive Model and PARR-30 for predicting risk of readmission to hospital.

The most accurate risk stratification outcomes are achieved by using carefully selected specific inputs. Unfortunately, populations of patients and even individual diseases are far from specific. For example, accident and emergency visits and unplanned hospital admissions are associated with use a lot of resources and high costs that would be ideal for reducing, but as well as patients who are high risk through the need for acute care, many other factors, such as age, sex, chronic conditions, and so on, also put different groups of people at high risk. Calculating and understanding the probability of a particular outcome for an individual, however, may not be enough for healthcare professionals to intervene in the most effective way to delay or prevent the outcome or divert the course of a disease, and often needs to be supported by additional information to determine the most accurate or appropriate model.[9]

Many risk stratification predictive modelling tools have been developed, often for specific diseases. However, calculating and understanding the probability of a particular outcome for an individual may not be enough for healthcare professionals to intervene effectively to delay or prevent the outcome or divert the course of a disease. Often, additional information is needed to determine the most accurate or appropriate model to use.

If the high-risk stratum is highly heterogeneous, targeting of interventions is difficult. For instance, among people attending the accident and emergency department for the acute care, little reduction in risk may be possible because of the unpredictability of causes and complexity of types of injury.[10,11,12] Yet, among other patients it is hard to find clear targets or to know how to splice the population to maximise outcomes. In a randomised stepped-wedge trial to assess the use of the PRISM risk stratification tool in primary care, the effects on service usage, costs, mortality, quality of life, and satisfaction in relation to accident and emergency usage was measured in 32 general practices and involved 230,000 patients. Even with training and support for staff, increases, not decreases, were found in unplanned admissions, accident and emergency attendances, and overall healthcare costs.[13]

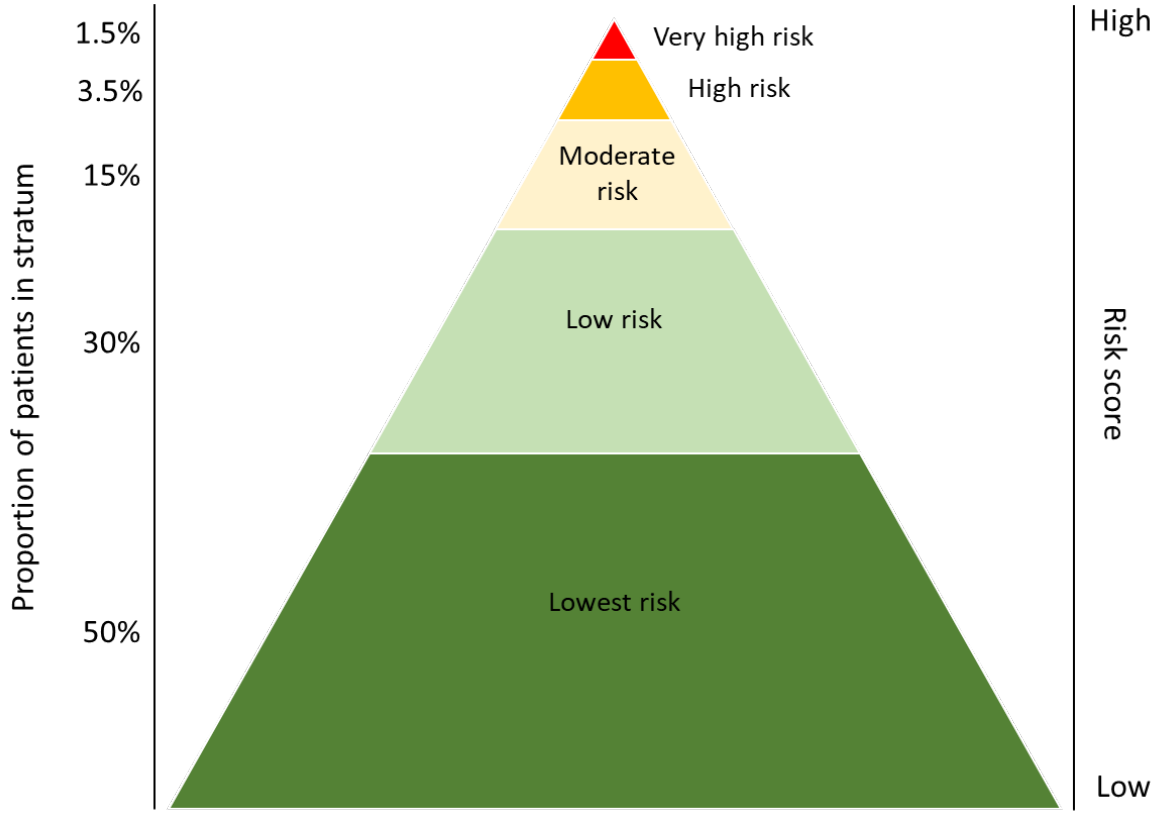
This how to guide covers the basics of what risk stratification is, how it is used, how to assess accuracy of risk assignment, and provide various examples highlighting how its potential may be maximised.

[9] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74.
[10] Woodhams V, de Lusignan S, Mughal S, et al. Triumph of hope over experience: learning from interventions to reduce avoidable hospital admissions identified through an Academic Health and Social Care Network. BMC Health Serv Res 2012;12:153. doi: 10.1186/1472-6963-12-153.
[11] 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;333:327. doi: 10.1136/bmj.38870.657917.AE.
[12] Lewis G, Kirkham H, Duncan I, Vaithianathan R. How health systems could avert 'Triple Fail' events that are harmful, are costly, and result in poor patient satisfaction. Health Aff (Millwood) 2013;32(4):669-76.
[13] Snooks H, Bailey-Jones K, Burge Jones D, et al. Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC). Southampton (UK): NIHR Journals Library; 2018 Jan.

2 What is risk stratification?

Risk stratification predictive modelling tools are widely commissioned and used in healthcare, partly because the concept is simple: by predicting the risk of future adverse events, efforts can be targeted to avoid or mitigate them. They use clinical thresholds to stratify a population into people at high, medium, and low of a given event. It is also possible to incorporate further factors, such as interventions, comorbidities, demographics, geography, and social factors, into calculations. This process can help to keep numbers within each cohort manageable and provides some leeway for clinical discretion (**Figure 1**). For example, when the focus is on risk of future adverse events, interventions could be targeted towards patients in the high-risk stratum as well as those in the highest-risk stratum.

Figure 1: Example risk stratification pyramid indicating stratum sizes



The thresholds set mean that stratum size decreases as predicted risk rises, making the cohort manageable even, for example, if the second highest stratum is targeted for intervention.

By predicting the risk of future adverse events, efforts can be targeted to avoid or mitigate them. Risk stratification predictive modelling tools use clinical thresholds to stratify a population into people at high, medium, and low of a given event.

Stratification of a population by risk can help with decision-making about the most appropriate proactive interventions for needs. Currently, much of the focus with risk stratification programmes in the NHS is on predicting events that could lead to “triple fail”. Such admissions are important for three reasons. First, they may be an indicator of suboptimal care; second, they are generally unpleasant and undesirable for patients and their families; and third, they are costly to the health service. Similar potential “triple fail” events are readmissions to hospital within 30 days of discharge and admission to a nursing home in the next 12 months. Risk stratification tools have already been developed using NHS data to predict such events so it will be important for local NHS organisations to consider the potential role of these tools as another way of improving the health of their local population.

Currently, much of the focus with risk stratification programmes in the NHS is on predicting events that could lead to “triple fail”. Risk stratification tools have already been developed using NHS data to predict such events so it will be important for local NHS organisations to consider the potential role of these tools as another way of improving the health of their local population.

Risk stratification is most beneficial if its use triggers actions (i.e., acts as a screening tool or intervention rather than merely a data analysis tool) and if those actions deliver benefits that outweigh the costs. In reality, though, effort tends to be mainly or completely concentrated in the highest-risk stratum, but this focus can lead to missed opportunities. Additionally, no risk stratification tool is 100% accurate. Consideration should be given to increasing the size of the highest-risk stratum to capture more unplanned admissions (although issues with this are discussed in Use and performance of risk stratification) or extending intervention plans to all strata (e.g., the highest-risk stratum can be given appropriate proactive interventions, but the plan would also include other routes to care, such as social prescribing, for patients at medium risk and continued usual care and education on self-care for lower-risk patients; see Impactibility).

Risk stratification is most beneficial if its use triggers actions (i.e., acts as a screening tool or intervention rather than merely a data analysis tool) and if those actions deliver benefits that outweigh the costs.

A summary of the pros and cons of risk stratification is provided in **Box 1**.

Box 1: summary of risk stratification pros and cons

Pros

- Case finding can ensure that individuals at risk of an adverse event can be offered an intervention designed to reduce that risk.
- Stratifying a population by risk can guide how to target appropriate proactive interventions
- It can be used as a PHM planning tool, enabling commissioners and providers to gain a detailed picture of the future risk profile of its population, allowing them to design care pathways and target funds and interventions appropriately.

Cons

- Inputs must be carefully selected to maximise accuracy
- No predictive model is 100% accurate
- A small high-risk stratum can lead to missed opportunities.

3 Preparing for risk stratification

3.1 Information governance

The NHS has published guidelines on complying with the legal framework for managing patients' data in the Health and Social Care Act 2012, the NHS Act 2006, the Data Protection Act, and the Human Rights Act.

In order for CCGs/GPs to undertake risk stratification they must provide assurance to NHS England that they or their risk stratification tool providers meet the Confidential Assurance Group (CAG) approval conditions, as set out in the Risk Stratification Assurance Statement. A documented Risk Assessment Assurance action plan should be completed to demonstrate evidence of implementation of the requests.

The NHS also has a list of approved risk stratification organisations, which is updated monthly.

3.2 First steps

3.2.1 Gap analysis

The first step for an NHS organisation interested in performing risk stratification is to do a gap or opportunity analysis to confirm the current situation compared with the goal. This process involves analysing population data to identify the incidence of low-quality, high-cost, poor-experience events, such as unplanned hospital admissions. This can include approaches such as literature searches, document/record analyses, and focus groups and surveys. The NHS provides the Change Model GAP analysis tool to help track and score information and findings.

3.2.2 Impact assessments

Consideration needs to be given to the ethics of predicting adverse events and offering interventions designed to prevent them. False-positive and false-negative assignments to the highest-risk stratum could lead to harm through unnecessary testing and/or treatment, potential anxiety to patients (i.e., through receiving an intervention for an event that would not have occurred) and might waste resources. As risk stratification is analogous to population screening, it is suggested that the ten principles of Wilson and Junger (**Box 2**) should be met.[14] The issue of within-stratum performance and how to check costs are not exceeded are discussed in Use and performance of risk stratification.

Box 2: Wilson & Jungner's principles of screening

1) The condition should be an important health problem
2) There should be an accepted treatment for patients with recognised disease
3) Facilities for diagnosis and treatment should be available
4) There should be a recognisable latent or early symptomatic phase
5) There should be a suitable test or examination
6) The test should be acceptable to the population
7) The natural history of the condition, including development from latent to declared disease, should be adequately understood
8) There should be an agreed policy on whom to treat as patients
9) The cost of case-finding (including a diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
10) Case-finding should be a continuous process and not a "once and for all" project

[14] Wilson J, Junger G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968. https://apps.who.int/iris/bitstream/handle/10665/37650/WHO_PHP_34.pdf?sequence=17.

3.2.3 Infrastructure assessment

Before starting risk analysis, a comprehensive healthcare infrastructure assessment should be performed to ensure that any needs identified can be met at local and/or broader levels, particularly if there might be an increased uptake of services with implications on resources, including capacity to meet workload. If a suitable infrastructure is not in place, this should be addressed before moving on to the next stage. Key interdependent services include:

- Community teams
- Primary care and clinics
- Clinical networks
- Secondary care

Health inequalities can arise due to a range of factors, such as geography and socioeconomic status, but also including ethnicity, religion, age, and sex. Collaboration with various local non-healthcare organisations, faith communities and voluntary groups should be considered to facilitate channels of communication and improve relationships with groups with high needs/risk.

3.2.4 Segmentation analysis

It can be useful to separate the overall population into smaller groups to increase the relevance of the intervention. This analysis can be facilitated by population segmentation. Useful information to use could be patient identifiers, geography, ethnicity, age, condition, procedure, and healthcare setting, usage, and costs. The main segment parameters should reflect the goal for change of care after risk stratification. The development of 'use cases' based on the current system can help to provide baselines against which change can be measured and to inform which interventions should be implemented, for whom, and by whom, and how outcomes will be measured. The results should be considered in the wider context of the system to reveal gaps, opportunities, and impact on care.

4 Use and performance of risk stratification

The ability of a risk prediction tool to discriminate between patients who do and do not experience the adverse outcome (e.g., death or an unplanned hospital admission) is a key indicator of performance. The performance of a tool to discriminate between strata is indicated by the *c*-statistic – an aggregate number that reflects the distribution of true positives and true negatives across all risk scores. Values range from 0 to 1. Perfect discrimination is indicated by a value of 1 (values ≥0.8 are generally judged to be good and of 0.7–0.8 reasonable), discrimination is equivalent to chance by a value of 0.5, and poor discrimination lower values.

However, while risk stratification models can accurately predict future adverse health outcomes, such as readmission risk or 1-year mortality risk,[10,11,12,15,16] their use has not consistently led to improvements in health outcomes across the population.[16]

[15] Bernstein RH. New arrows in the quiver for targeting care management: High-risk versus high-opportunity case identification. J Ambul Care Manage 2007;30:39-51.
[16] Bardsley M, Blunt I, Davies S, Dixon J. Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. BMJ Open 2013;3:e002007.

Furthermore, as many risk stratification models predict future adverse health outcomes through current or previous healthcare activity and use a limited number of variables,[17,18,19,20] they may miss out on valuable additional information that could better direct resources to patients amenable to benefit.[21,22] Additionally, it is key to the success of risk stratification to ensure that high-risk individuals are not conflated with those most likely to benefit as there is evidence indicating that these can be highly separated groups.[15,23,24]

While risk stratification models can accurately predict future adverse health outcomes, their use has not consistently led to improvements in health outcomes across the population. It is key to the success of risk stratification to ensure that high-risk individuals are not conflated with those most likely to benefit as there is evidence indicating that these can be highly separated groups.

Since the *c*-statistic assesses the tool as a whole, it can be more important to consider the performance within a given stratum. Positive predictive value (PPV) is very useful for this because it indicates the probability of the risk prediction tool correctly identifying from among people who might or might have the condition all those who do have it (i.e., true positives) while avoiding categorising some people as having the condition when they do not (i.e., false positives).[25] The higher the PPV, the better the risk prediction tool is at picking out people who will experience the adverse event.

4.1 Within-stratum accuracy by sensitivity and PPV

Assessment of the accuracy of risk-stratum assignment of patients needs to consider the adverse impact of false-positive and false-negative results as well as the benefits of true-positive and true-negative results (**Figure 2, Table 2**).[26] For example, if risk stratification of a population of 100,000, assigns 1% (1,000) to the highest-risk stratum, not all these people will experience the event of interest because prevalence is actually only 0.2%, so only 200 will be affected. However, because they are in the highest-risk stratum, all 1,000 patients are assigned the intervention of interest.

[17] 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.e001667.
[18] Bottle A, Aylin P, Majeed A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. J R Soc Med 2006;99:406-14.
[19] Donzé J, Aujesky D, Williams D, Schnipper JL. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. JAMA Intern Med 2013;173:632-38.
[20] van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010;182:551-57.
[21] Freund T, Wensing M, Geissler S, Peters-Klimm F, Mahler C, Boyd CM, et al. Primary care physicians' experiences with case finding for practice-based care management. Am J Managed Care 2012;18:e155-61.
[22] Steventon A, Billings J. Preventing hospital readmissions: the importance of considering 'impactibility,' not just predicted risk. BMJ Qual Saf 2017;26:782-85.
[23] Dubard CA, Jackson CT. Active redesign of a Medicaid Care management strategy for greater return on investment: predicting impactability. Popul Health Manage 2018;21:102-09.
[24] Flaks-Manov N, Srulovici E, Yahalom R, Perry-Mezre H, Balicer R, Shadmi E. Preventing hospital readmissions: healthcare providers' perspectives on "impactibility" beyond EHR 30-day readmission risk prediction. J Gen Intern Med 2020;35:1484-89.
[25] Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307. doi: 10.3389/fpubh.2017.00307.
[26] Lewis G. Next steps for Risk Stratification in the NHS. https://www.england.nhs.uk/wp-content/uploads/2015/01/nxt-steps-risk-strat-glewis.pdf

Figure 2: Measurement of within-stratum performance – sensitivity and PPV

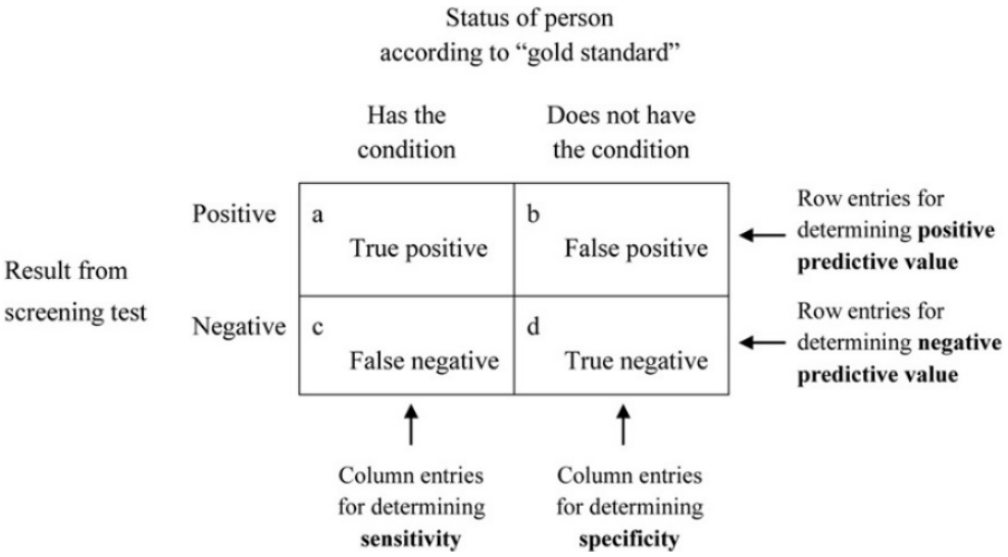


Table 2: Example of sensitivity and PPV calculations

	Meet criteria	Do not meet criteria	Total
Test positive	a) True positive (correctly identified as being at risk and receive intervention) = 178	b) False positive (wrongly identified as being at risk but receive intervention) = 145	323
Test negative	c) False negative (wrongly identified as not being at risk and did not receive intervention) = 22	d) True negative (correctly identified as not being at risk and did not receive intervention) 655	677
Total	200	800	1,000

True positive=person is correctly identified as being at risk

True negative=person is correctly identified as not being at risk

False positive=person is wrongly identified as being at risk

False negative=person is wrongly identified as not being at risk

Checking the within-stratum accuracy indicates how many would have needed the intervention (i.e., those who would have had the event but it was prevented by the intervention).

Sensitivity considers only people who will experience the event (i.e., true positives and false negatives). By contrast, PPV, considers only people at risk of the event, irrespective of whether or not they were correctly identified (i.e., true positives and false positives).[25]

- In the example shown in Table 2 sensitivity and PPV may be calculated as follows:
- The first column can be used to calculate sensitivity = [a/(a+c)] = **0.89**
- The first row can be used to calculate PPV = [a/(a+b)] **0.55**

Therefore, 89% of the people at risk of the event are correctly identified. However, the PPV indicates that only 55% of the 323 people at risk of the event would experience it.

The risk threshold may be altered to increase or decrease the number of true positives captured in the highest-risk stratum. However, there is a trade off in the opposite direction with PPV because the number of false positives will also change.[26] The risk threshold may be lowered to increase the number of people identified as being high risk. This causes the sensitivity of the model to increase (i.e., a higher absolute number of people at high risk will be identified) but leads to a decrease in the PPV (i.e., the proportion of those identified that are actually at high risk) because there will also be more false positives. In the example shown, with a stratum threshold of 1% there is high sensitivity (most of the people who are at risk in the stratum have been identified correctly) but there is low positive predictive value (a low number of people in the overall population who are truly positive have been assigned to the high-risk stratum).

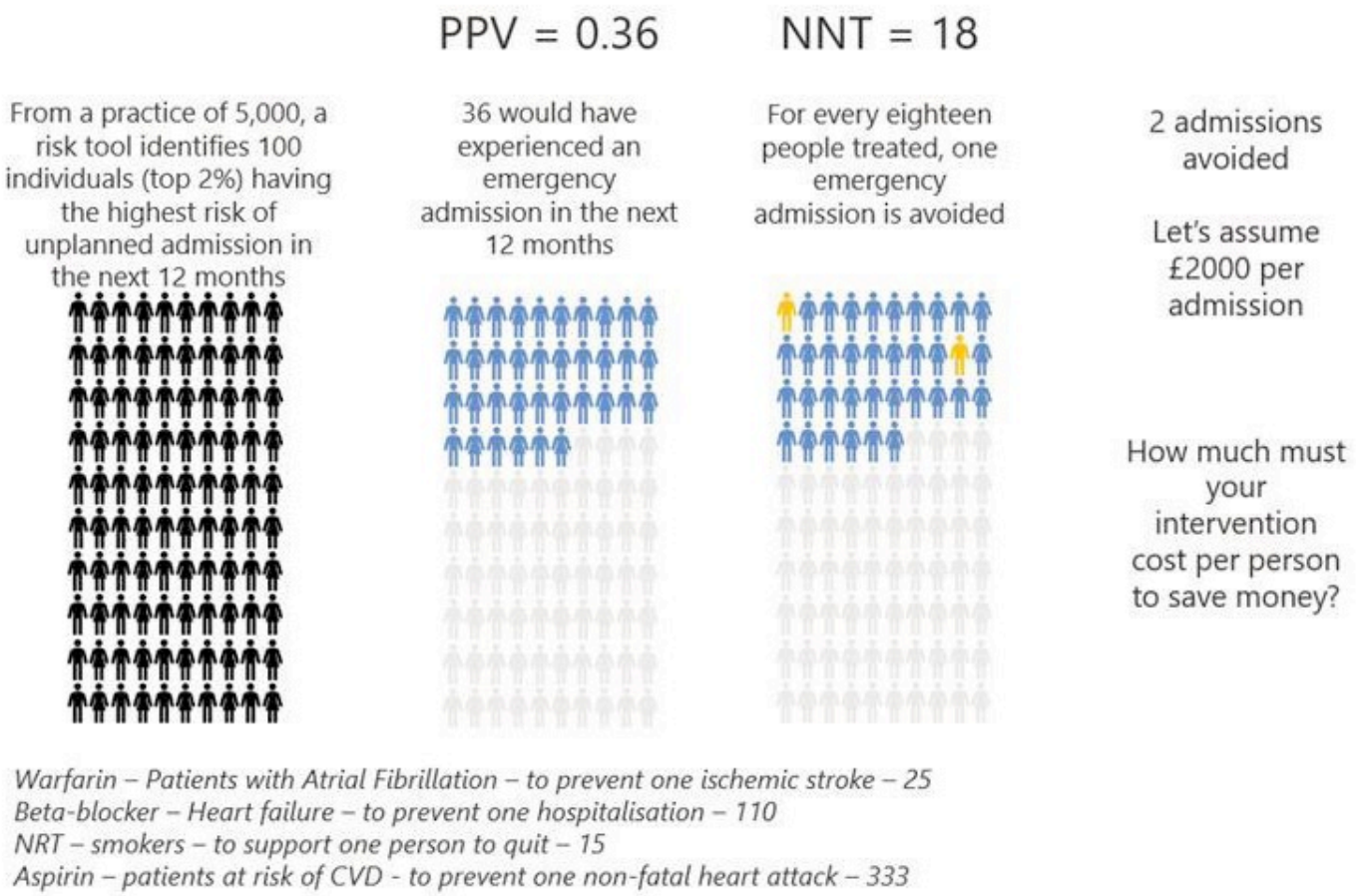
4.2 Within-stratum accuracy and by PPV and number needed to treat

Another method for assessing performance has been proposed by Wyatt et al[27] combines the PPV and number needed to treat (NNT). This approach 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, and is calculated as (I <A*PPV/NNT).

Figure 3 provides a worked example in which a hypothetical general practice, with 5,000 patients, uses a risk prediction tool to identify the top 2% (100) of patients at risk of an unplanned hospital admission in the next 12 months.

[27] Wyatt S, Mohammed MA, Rahim S, Spilsbury P. Is risk stratification likely to improve the use of NHS resources? https://www.strategyunitwm.nhs.uk/sites/default/files/2021-10/RiskStratification-StrategyUnitPaper.pdf

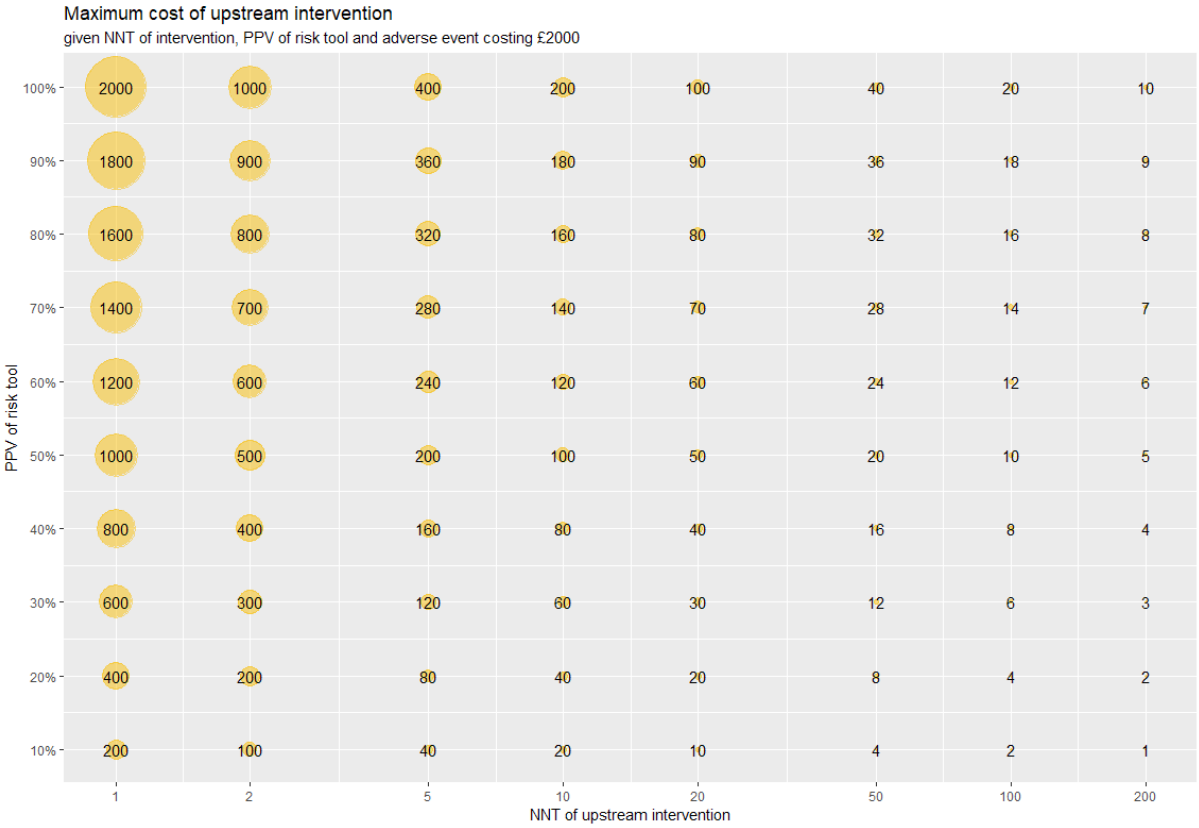
Figure 3: Measurement of within-stratum performance – PPV and number NNT



The PPV of the risk prediction tool in this top 2% is reported to be 36%, meaning that 36 of the 100 identified patients would be expected to experience an unplanned admission whereas 64 would not but all would receive an intervention. The example uses a NNT of 18 (i.e., for every 18 identified people treated who would otherwise have been admitted, one unplanned admission would be avoided). Thus, among the 36 patients who go on to experience the event, the intervention would avoid 2 unwanted outcomes. Therefore, an unplanned admission costs £2,000, to save money the intervention must cost less than £40 per patient (i.e., n = 2 x 2000/100). This assumes that there are no costs to developing and deploying the risk prediction tool, but such costs could be added into the equation if required.

This knowledge of the cost threshold provides important information to commissioners. As shown in Figure 4, it is recommended that commissioners are provided with a range of NNTs, so they can apply them to the model and compare outcomes. Lower NNTs (i.e., meaning more effective interventions), will translate to being able to afford more per identified patient for a given PPV because impacts become more pronounced.

Figure 4: Maximum cost of an upstream intervention (£) per identified patient for a range of PPVs and NNTs where the adverse event costs £2,000



5 After risk stratification

5.1 Impactibility

It must also be borne in mind that not all events will be preventable with the proposed intervention,[Error! Bookmark not defined.] either because current interventions are not effective for all people and/or patients do not accept the intervention. Additionally, as mentioned earlier, preventive interventions are generally offered only to patients in the highest-risk stratum, mainly because they incur disproportionately high healthcare costs compared with the rest of a population even though most events in the population will not occur among people in this stratum. For instance, if 10% of events occur among the top 1% of a population assigned to the highest-risk stratum, 90% will come from other strata. So, targeting only the highest-risk group is likely to have a small effect on costs overall and might even increase them along with health inequalities.

The concept of 'impactibility' considers which patients within strata are most likely to benefit from the intervention so that care may be targeted to these patients first. This aims to avoid risk of harm through increased anxiety, over-testing, and/or over-treatment. There are several approaches to assessing impactibility, as discussed by Orlowski et al.[28]

[28] Orlowski A, Snow S, Humphreys H, Smith W, Jones RS, Ashton R, Buck J, Bottle A. Bridging the impactibility gap in population health management: a systematic review. BMJ Open 2021;11:e052455. doi: 10.1136/bmjopen-2021-052455.

Not all events will be preventable with the proposed intervention either because current interventions are not effective for all people and/or patients do not accept the intervention. Targeting only the highest-risk group is likely to have a small effect on costs overall and might even increase them along with health inequalities. The concept of 'impactibility' considers which patients within strata are most likely to benefit from the intervention so that care may be targeted to these patients first.

5.2 Data review and effect monitoring

The risk stratification exercise should not be a one-off. Rather, the data generated should be used in a feedback loop to improve the performance of the programme. Patients within the highest stratum for unplanned admissions and other key "triple fail" events change rapidly. Therefore, it is useful to rerun the analysis regularly and to keep track of similarities and differences in the data.

Evaluation and improvement mechanisms should be determined to ensure that learning is captured and utilised, to drive improvements, and to understand the impacts and outcomes of the initiative.

Data generated should be used in a feedback loop to improve the performance of the programme. It is useful to rerun the analysis regularly and to keep track of similarities and differences in the data to ensure that learning is captured and utilised.

Consider working with other NHS organisations and pooling data. The numbers for local-level risk stratification exercises might be small, and assessment of larger datasets might help to reveal new findings.

6 Case study

Using local data to more accurately model patients' risk of emergency admission

A model commonly used in the UK for identifying people at risk of emergency admission is the Wales model, which is based on data from 2008. Recognising changes since 2008 in the way care is delivered, as well as an improvement in the quality of health and care data available, the data science team at Cheshire CCG set about developing a new model using their local datasets. Conscious of the opportunity to compare predictions from their model to the Wales model, they decided to ask the same question: 'which patients experienced an emergency admission in the following 12 months?'

Due to the richness of the linked data available, they began with more than 2,000 variables, compared to around 30 used in the Wales model. To refine the list of variables, they turned to recursive feature elimination (RFE), a form of machine learning that uses an algorithm to systematically remove the least important variable over and over until you have a more manageable range of key variables – in this case around 100.

With around 600,000 data points, they could see which patients in the area did experience an emergency admission. This was used to build and train a model to learn the features and experiences of someone who ended up having an emergency admission. They explored several types of models but landed on a generalised linear model which uses variables and coefficients to drive how important something is and delivers a risk score which, in this case, clinicians can use to better inform and target their interventions.

A bonus for the team was having the Wales model to use as a baseline to confirm the success of their new model. As an illustrative example, if you took an imaginary group of 1,000 people, 62 would on to experience an emergency admission. When asked to find the 62 people with the highest risk in this hypothetical village, the Wales model may correctly identify 14, meaning there are 48 people who are given a high risk score, who don't have an emergency admission. The Cheshire CCG model was slightly more accurate and would result in correctly identifying 17 people from the top 62. As well as making better predictions on emergency admissions, this local model was also slightly better at predicting people who would not have an emergency admission.

The end goal was to provide a tool for GPs and other people in primary care who need to understand why someone has been given a certain the score and the team developed a user-friendly platform where clinicians can view real-time data on risk among their patients.