BMJ Health & Care Informatics

Promising algorithms to perilous applications: a systematic review of risk stratification tools for predicting healthcare utilisation

Christopher Oddy ,[1] Joe Zhang,[2,3] Jessica Morley,[4] Hutan Ashrafian[2]

To cite: 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. doi:10.1136/bmjhci-2024-101065

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2024-101065).

Received 23 February 2024
Accepted 14 May 2024

Check for updates

© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

[1]Department of Anaesthesia, Critical Care and Pain, Kingston Hospital NHS Foundation Trust, London, UK
[2]Imperial College London Institute of Global Health Innovation, London, UK
[3]London AI Centre, Guy's and St. Thomas' Hospital, London, UK
[4]Digital Ethics Center, Yale University, New Haven, Connecticut, USA

Correspondence to
Dr Christopher Oddy;
christopher.oddy1@nhs.net

ABSTRACT

Objectives Risk stratification tools that predict healthcare utilisation are extensively integrated into primary care systems worldwide, forming a key component of anticipatory care pathways, where high-risk individuals are targeted by preventative interventions. Existing work broadly focuses on comparing model performance in retrospective cohorts with little attention paid to efficacy in reducing morbidity when deployed in different global contexts. We review the evidence supporting the use of such tools in real-world settings, from retrospective dataset performance to pathway evaluation.

Methods A systematic search was undertaken to identify studies reporting the development, validation and deployment of models that predict healthcare utilisation in unselected primary care cohorts, comparable to their current real-world application.

Results Among 3897 articles screened, 51 studies were identified evaluating 28 risk prediction models. Half underwent external validation yet only two were validated internationally. No association between validation context and model discrimination was observed. The majority of real-world evaluation studies reported no change, or indeed significant increases, in healthcare utilisation within targeted groups, with only one-third of reports demonstrating some benefit.

Discussion While model discrimination appears satisfactorily robust to application context there is little evidence to suggest that accurate identification of high-risk individuals can be reliably translated to improvements in service delivery or morbidity.

Conclusions The evidence does not support further integration of care pathways with costly population-level interventions based on risk prediction in unselected primary care cohorts. There is an urgent need to independently appraise the safety, efficacy and cost-effectiveness of risk prediction systems that are already widely deployed within primary care.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Risk prediction models that stratify primary care populations according to their likelihood of accessing healthcare resources are generally considered to perform well within similar contexts to those in which they were derived. It is unclear how they perform when deployed in wider global contexts and indeed if their application can be harnessed to reduce resource demands.

WHAT THIS STUDY ADDS

⇒ We find that most models have not been studied in a sufficient diversity of contexts to appraise the robustness of prediction, however, those that have appear to retain their discriminatory ability. The real-world application of these models to reduce healthcare resource use in unselected cohorts has produced disappointing results, with an equal weight of evidence suggesting a harmful effect as a beneficial one in this context.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our results call into question the common, and costly, practice of commissioning population health management strategies based on risk stratification of whole primary care populations without a concrete understanding of the associated risks.

INTRODUCTION

Risk stratification tools that predict healthcare resource use are widely used in primary care settings.[1–6] These tools are integral to population health management (PHM) strategies around the world, enabled by the availability of routinely collected data from sources such as electronic health records.[7] Risk stratification tools typically use predictive models that are developed through statistical or machine learning (ML) techniques, to generate an individual risk score for some measure of resource use. These scores form a key component of anticipatory care pathways, where those at the highest risk may be targeted for specific interventions aimed at reducing future morbidity.[8–11] The process by which these tools are ideally developed and deployed within healthcare systems is summarised in figure 1.

A growing body of literature describes the development and validation of risk

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

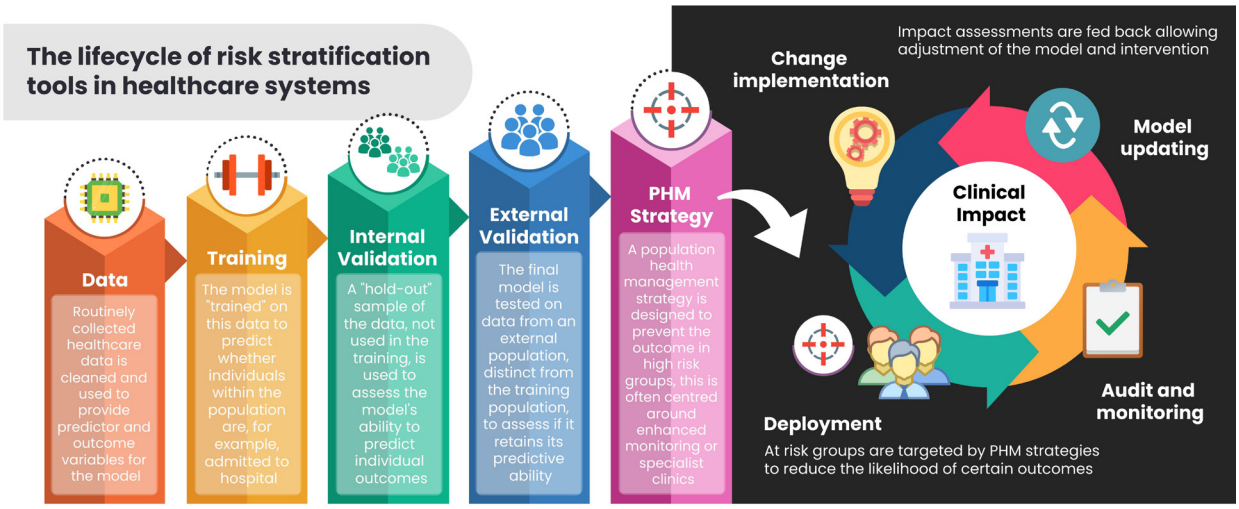

Figure 1 An infographic describing an idealised process for developing and deploying a risk prediction tool within a healthcare system. In black is the deployment cycle, linking risk prediction tools and their associated population health management measures to a lifecycle of evidence generation, impact evaluation and monitoring for negative consequences that are fed back into the model and intervention.

stratification tools in the primary care setting reporting an acceptable discriminatory power for the majority of models.[1 2 12 13] However, existing work broadly focuses on the assessment of model performance within retrospective datasets, with little attention paid to their efficacy in real-world settings, where the clinical impacts of deploying these algorithms within a population are assessed. Commercial literature asserts the efficacy of interventions based on algorithmic case selection in improving key outcomes, such as hospital admission rates, but suffers from a lack of transparency in data and methodology.[14 15]

Predictive models that appear accurate in development are increasingly found to be ineffective or unsafe when deployed in clinical pathways. Predictive performance may be diminished when translated to demographically and culturally distinct populations, or when deployed using electronic health data with differing characteristics. Differences in how healthcare resources are used in local settings, alongside inherent biases inlaid within such technologies, may result in varying clinical effectiveness from inconsistent intervention thresholds, variation in the physical clinical interventions that are deployed, to sociotechnical variation across end-users and processes.[16–20] Resultantly, where an algorithm is deployed into an untested context without real-world evidence for a comparable integrated pathway, there are risks to both patient safety and exacerbation of healthcare inequalities through a lack of fairness in prediction or intervention allocation.

With extensive integration of risk stratification into pathways within primary care systems worldwide it is of paramount importance to establish the current evidence base on which these care-defining interventions can be appraised. We therefore systematically review the available literature concerning risk stratification tools for predicting future healthcare utilisation in primary care populations. We present three aims: (1) to update existing evidence for algorithmic solutions with attention paid to predictive performance and risk of bias in dataset evaluation, as well as real-world clinical outcomes; (2) to describe the transfer of algorithms from initial development to testing and deployment in different global contexts and (3) to evaluate risks in cross-context transfer and application. Based on our findings, we provide recommendations for the responsible evaluation and deployment of predictive risk stratification tools.

METHODS

Search strategy

A systematic search of the MEDLINE, Embase and Global Health databases was carried out on 18 July 2023 via the Ovid platform. PRISMA guidelines were followed throughout the conduct and reporting of this review.[21] A combination of keywords and MeSH terms was used to curate relevant literature, details of which are available in online supplemental material.

Inclusion and exclusion criteria

We defined our inclusion criteria using the Population, Intervention, Control and Outcome method. The population of our analysis was selected to be comparable to the populations in which these models are currently in use. We therefore included only papers that applied algorithms to unselected primary care populations, where deployment was to the entire patient population for a given organisation without selection of particular groups. Prestratified populations, such as specific disease groups, or groups previously identified as high risk for healthcare utilisation, were excluded. Age-stratified populations were permitted as this is a pragmatic selection criterion adopted by the majority of predictive modelling work. Publications applying algorithms to historic research

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

study datasets or specifically designed questionnaires (ie, not routinely collected or 'real-world data'[22]) were also excluded.

Our intervention was defined as the application of a risk stratification model to an appropriate population in the process of derivation or validation, or to perform case selection as part of a PHM strategy. Models reliant on non-routinely collected data, such as questionnaire results, were excluded.

Outcomes included measures of predictive performance across five main categories: access to primary care services; emergency department attendance; healthcare costs; hospital admissions and readmission. Studies examining risk of readmission were included provided that the study population was not limited to patients with a recent admission. A group formed of those who had recently been admitted would, by definition, no longer be considered unselected and would thus violate our population criteria. Composite (eg, admissions and mortality as a single endpoint) and component (eg, respiratory admissions instead of total admissions) outcomes were excluded. We also considered clinical impact assessments related to a real-world evaluation.

Study selection and quality appraisal

Titles and abstracts were screened by two reviewers (CO/JZ) according to the criteria set out above, with all conflicts decided by a third (JM). Eligible publications were read in full and assessed for exclusions not apparent in the title or abstract, and for methodological quality.[23] Risk of bias was assessed using the Prediction model Risk Of Bias ASsessment Tool.[24]

Data extraction

We extracted information regarding model characteristics, study design and context, predictive performance, and measures of clinical impact from any associated intervention where evaluation took place in a real-world setting. Due to significant heterogeneity in study design and reporting a meta-analysis was not conducted. C-statistics were used as the primary outcome for model performance. A subset of papers did not report discrimination, but instead reported goodness of fit using coefficient of determination (R^2) which were extracted where available. Impact evaluations were described using the terminology and significance testing employed in the original paper, commonly expressed as the absolute difference (AD) between groups or odds ratios (OR).

Model appraisal

Models that appeared in multiple studies were qualitatively appraised by comparing their derivation methodology to subsequent external validation or clinical evaluation studies. For each model we report: the context of its original development; contexts in which the model's predictive performance has been tested; and contexts in which the model's real-world impacts have been assessed. Results were synthesised separately as the outcome of either internal or external validation. Internal validation was defined as any measure of predictive performance within the same population in which the model was derived, and external as any validation using data from a separate population.

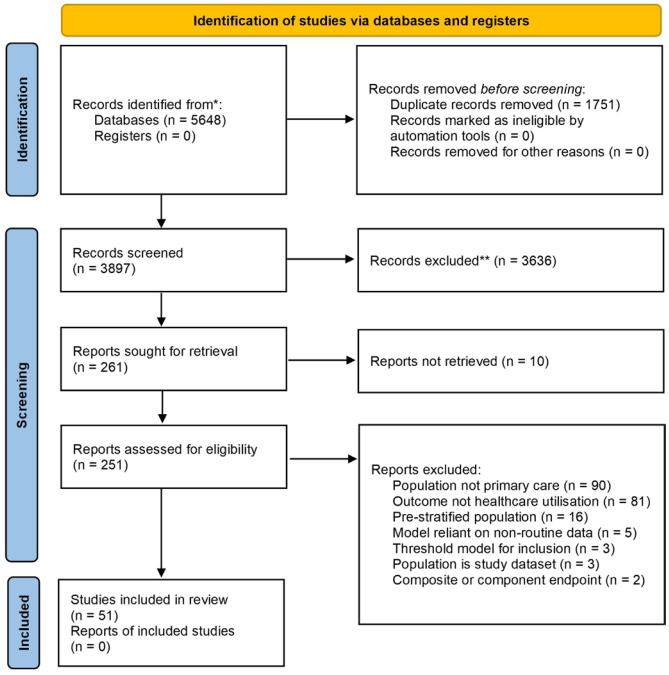

Figure 2 A PRISMA flow diagram showing the process of study selection for our analysis. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

RESULTS

Systematic review

Our review identified 3897 publications eligible for screening after duplicates were removed (figure 2). Of these, 3636 were excluded on the basis of their title or abstract alone leaving 261 that were sought for retrieval. Full texts could not be retrieved for 10 publications, thus 251 were reviewed in full. A total of 51 publications met our criteria and were included in our final analysis (online supplemental table 1).[25–75] Further detail about the identified models, along with our risk of bias analysis, can be found in online supplemental materials.

The majority of studies were based in the USA (23), with the remainder set in the UK (10), Spain (9), Canada (2), Italy (2), New Zealand (2), Australia (1), Ireland (1) and Israel (1). Population sizes ranged from 96 to 5.4 million with a median value of 94 264 (IQR 12 800–434 027). Hospital admission was the most commonly predicted outcome (34), followed by healthcare costs (14), emergency department attendance (9), access to primary care services (8), mortality (5) and readmission (2).

19 studies reported the derivation and internal validation of a risk stratification model with 32 describing validation of a model in a separate population dataset. 10 studies reported the results of implementing PHM measures based on case selection by a risk stratification

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

model in a real-world clinical pathway. These included five randomised control trials (RCTs), three prospective cohort studies and two retrospective cohort studies. PHM strategies used were case management (8), telemonitoring (4) and care coordination (3).

We identified 28 risk stratification tools across all studies. 42 studies examined a single model, whereas 9 studied the comparative efficacy of several models. Johns-Hopkins ACG was the most studied algorithm (20), followed by the Charlson Comorbidity Index (10), Hierarchical Condition Categories (8), the Chronic Illness and Disability Payment System (3), RxRisk (3), the Elder Risk Assessment Index (2), the Patients At Risk of Rehospitalisation algorithm (2) and QAdmissions (2). Of the remainder, four were proprietary ML algorithms.

Results of internal and external validation studies

A summary of the derivation characteristics of each of the 28 discovered models is compared with the results of subsequent validation studies in online supplemental table 2.[25–84] The results of internal validation studies echoed previous reviews with C-statistics for various outcomes ranging from 0.67 to 0.90. Notably, three of the highest C-statistics within internal validation samples were displayed by models derived using ML techniques—0.84,[67] 0.85[42] and 0.90.[55]

Half (14) of the discovered models underwent external validation. Of these, only the Charlson Comorbidity Index and the Johns Hopkins ACG System were validated internationally. Model performance in external validation studies generally resembled internal validation performance for each model, with C-statistics ranging from 0.53 to 0.88. Accounting for heterogeneity in study design and reporting, there was no evident association between validation context and model discrimination, with models broadly displaying consistent predictive performance when transported to external datasets.

Results of real-world evaluation studies

Two studies reported the implementation of risk stratification tools into care pathways within the same population used for development. The Nairn Case Finder[73] and the Predictive RIsk Stratification Model (PRISM)[25] algorithms were used to identify those that might benefit from case management, both in the hope of reducing hospital admissions. In a prospective stepped-wedge clinical trial conducted across more than 230 000 patients in 32 primary care practices, the practice resource allocation intervention linked to PRISM resulted in significantly increased hospital admissions (OR 1.44 (95% CI 1.39 to 1.50), p<0.001), as well as increased emergency presentations, time in hospital, and primary care workload. The intervention guided by the Nairn Case Finder significantly reduced hospital admissions (AD=42.5%, p=0.002) in a population of 96 high-risk patients from a single locality, when matched 1:1 on risk score to patients in a separate control population.

Eight of the discovered models were deployed as tools for case selection as part of a PHM strategy in a separate context from development. The Johns Hopkins ACG System was deployed in two separate studies, whereas each of the other models was deployed only once. Healthcare utilisation measures were not significantly influenced by interventions guided by the Hierarchical Condition Category[71] and PacifiCare's Medicare Risk Programme[37] models. Similarly equivocal evidence for the efficacy of interventions linked with the Johns Hopkins ACG System was observed, with one study showing no benefit[31] and the other demonstrating benefit in groups selected by the model (OR 0.91 (95% CI 0.86 to 0.96)) but reciprocal harm in non-prioritised groups (OR 1.19 (95% CI 1.09 to 1.30)).[32] Interventions linked with the Elder Risk Assessment Index[30] and QAdmissions[48] algorithms led to significant increases in mortality (AD 10.8%, p=0.008) and hospital admissions (difference in difference 79.8 (95% CI 21.2 to 138.4), p=0.01), respectively.

Significant reductions in hospital admissions were achieved through interventions guided by the combined predictive model (AD=−0.9, p<0.001),[39] Patients At Risk for Rehospitalisation algorithm (AD=−0.3, p<0.001)[39] and SCAN Health Plan Model (AD=11.5%, p=0.02).[51] Figure 3 summarises the main findings of this review, describing only the models that underwent external validation or real-world evaluation.

DISCUSSION

Main results

Our review identifies 28 risk stratification tools designed to predict healthcare utilisation in an unselected primary care population. The discriminatory ability of half of the discovered models was validated in an external cohort. However, only two, the Charlson Comorbidity Index and Johns Hopkins ACG System, were validated in a different country from their derivation dataset. No evident association between validation context and model discrimination was observed. Models derived using ML techniques displayed the best predictive performance, however, none of these models underwent external validation.

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.

Comparison with the literature

We corroborate the results of previous reviews by observing that the discriminatory power of a variety of risk stratification tools is robust to external validation.[1 2 12 13] We add that the context of model validation appears to have minimal impact on predictive performance and highlight a scarcity of literature appraising the impact of deploying these models to guide PHM strategies despite extensive

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

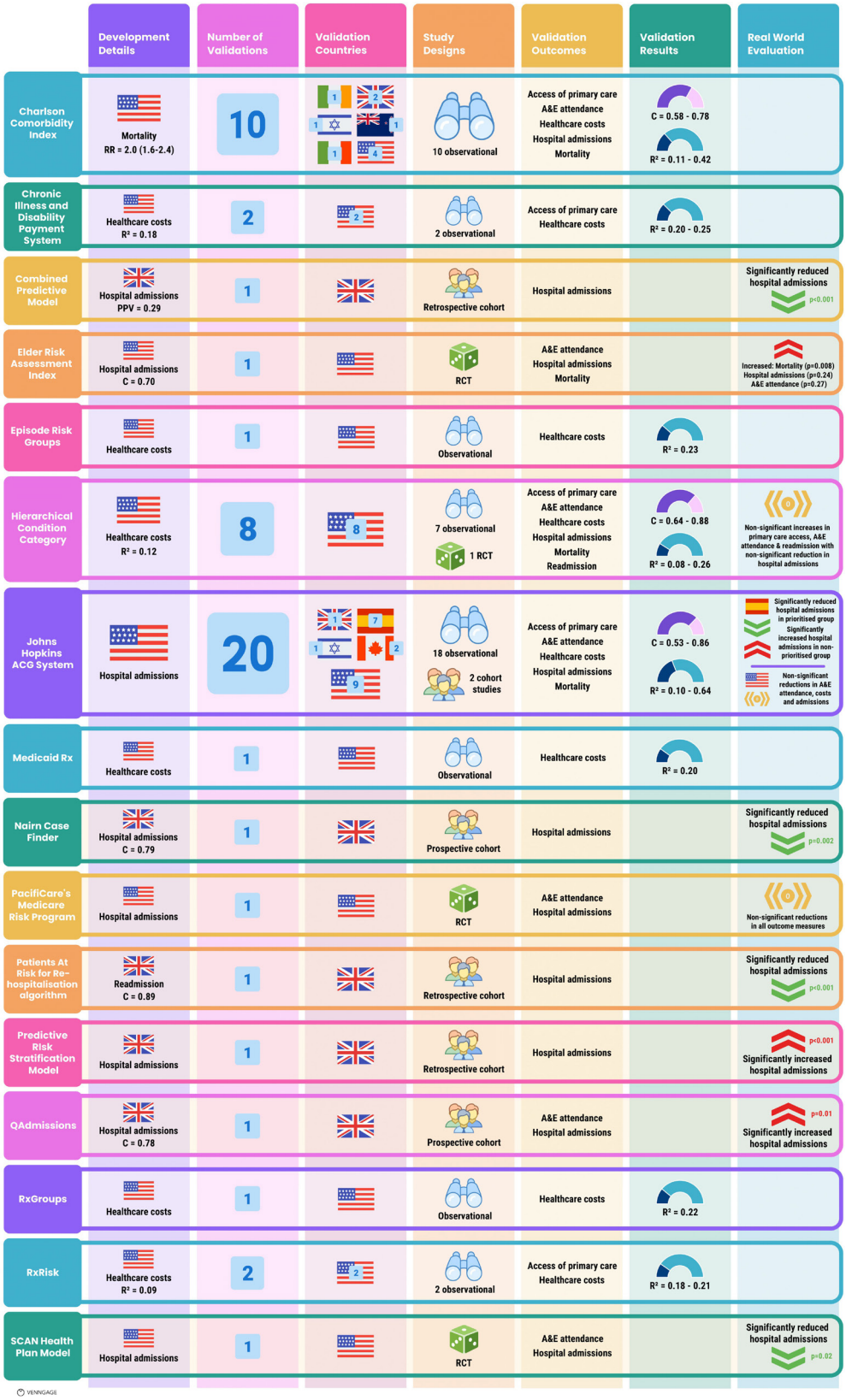

Figure 3 An infographic summarising the validation characteristics of the identified models that underwent external validation or real-world testing. Models that underwent more extensive validation processes are represented by larger boxes. Each box contains aggregated data for all of the external validation and real-world evaluation studies for each model. Validation countries are represented by flags with the number of studies based in each country overlying. R^2 and C-statistics are displayed as ranges for all of the outcome measures tested for each model for illustrative purposes only. A&E, accident and emergency department; PPV, positive predictive value; RCT, randomised controlled trial; RR, risk ratio.

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

integration of risk stratification into pathways within primary care systems worldwide.[3–6]

Our finding that deployment of these models is not consistently associated with reductions in healthcare utilisation is perhaps unsurprising. PHM strategies applied to unselected primary care cohorts, with case selection achieved through a variety of different means, have frequently been shown to increase costs without an associated reduction in morbidity.[9 85–87] A single 2014 meta-analysis, aggregating a heterogeneous group of strategies as a single intervention, demonstrated marginal reductions in resource use within a relevant cohort.[88] However, these findings were subject to substantial heterogeneity (I^2=58%–85%) and, while ostensibly the target population of this analysis was patients generally at high risk of healthcare resource use, the majority of included studies reported interventions targeted at specific disease cohorts. There is broad consensus that PHM strategies designed specifically for those with certain chronic conditions significantly reduce morbidity.[89–94] Taken with our findings, the available evidence indicates that the success of PHM strategies in specific disease groups may not be generalisable to unselected cohorts, and this remains the case when predictive modelling is employed to augment case selection.

The findings of our analysis of peer-reviewed literature stand in stark contrast to the impact statements of commercial suppliers of care systems that employ risk stratification. One such statement compared resource use statistics of product users to standardised national trends in an unadjusted analysis finding significant reductions in every parameter.[15] However, as is expressly the case for statements within product literature, a lack of transparency relating to the methods of data collection and analysis makes verifying these claims impossible.

Interpretation

We propose that the discouraging results of studies deploying risk stratification tools to guide PHM strategies primarily result from a mismatch between theoretical model development and complexities of real-world pathways. Risk stratifying patients by their likelihood of resource use alone almost invariably leads to the creation of a diverse intervention cohort, where individual clinical need is likely to be heterogeneous. This is likely the reason that population-level interventions have failed to replicate the results of successful programmes targeting specific chronic conditions. Presently, there is a paucity of evidence to guide best practice once high-risk users are identified, and no recommendations can be made about the efficacy of any single intervention over another. Results of real-world evaluation studies, therefore, present a cautionary tale of designing clinical pathways based on the principle of simply flagging high-risk patients without a concrete understanding of how this translates into practice.

We did not observe an effect of validation context on algorithmic performance. This is most likely due to the low number of comparable values obtained for each model, the heterogeneity of the study design, and a predictably small absolute effect size. Diminished performance when algorithms are deployed in new environments is a highly replicable finding, and our results should not be interpreted to contradict this established premise. However, this finding does imply that poor predictive performance is unlikely to be the primary reason for the failure of these algorithms to produce consistent results.

Limitations

It is important to put these findings within the context of our methodological constraints. Primarily, our analysis was limited by the heterogeneity of the included studies. Model performance was variably reported in terms of C-statistics and R^2 values which cannot be directly compared. Real-world evaluation studies suffered from a lack of uniformity of intervention as many reported the results of a bespoke system designed by the study authors. This prevented direct comparison of the efficacy of particular intervention categories within our study cohort as their results could not be appropriately aggregated. While our analysis identified several models with sufficient diversity of validation to demonstrate robust performance in a variety of contexts, this sample was small, and no strong conclusions can be drawn about the scale of algorithmic drift when such models are transported to new datasets. Finally, the majority of included publications were observational or cohort studies, with only a small number of RCTs identified.

Implications

The integration of risk stratification into pathways that define care decisions for millions of individuals around the world is already well established. Our findings suggest an absence of clinical impact, and indeed a signal of harm in a third of cases, raising several important considerations. First, this presents clear implications for patient safety, particularly in the absence of regular independent appraisal of the personal and system-wide effects. In addition to aggregate population health impacts, this includes the impact on individuals of incorrect stratification, and of negative biases through poorly calibrated algorithms. Second, the effects on provider workload of instituting and enacting these often time-consuming PHM interventions must be considered in the calculation of risk versus benefit. Finally, the absence of established benefits calls into question the cost-effectiveness of these programmes, particularly when used in healthcare systems where resources are constrained.

We therefore propose the following recommendations:

1. Deployment of individual-level risk prediction, with impact on clinical care pathways, must be subject to the same controls as other medical technologies. This would require matching their use to a responsible lifecycle of evidence generation, impact evaluation and monitoring for negative consequences. Such a lifecycle should include pre hoc evaluation, in the form of

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

local testing, and controlled trials for integrated pathways, as well as post hoc analyses of economic impact and healthcare outcomes in targeted and non-targeted groups. The first step in this process may be agreement on an auditable validation framework, such as BS 30440 developed by the British Standards Institution, to permit a more systematic approach to evaluation of such products.

2. National bodies involved in the procurement of commercial risk stratification services must review the cost-effectiveness and systemic implications of adjusting the likelihood of individuals within the population they serve accessing care based on personal predicted risk.
3. Regulatory bodies, including the Medicines and Healthcare products Regulatory Agency and the US Food and Drug Administration, must either confirm that risk stratification algorithms fall within their purview and are thus subject to the same regulation as other technologies defined as a 'Software as a Medical Device', or clarify why these algorithms do not fall into this category.

CONCLUSION

While model performance appears to generalise in most evaluations, 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 and cost-effectiveness of risk prediction systems that are already widely deployed within primary care.

X Christopher Oddy @_chrisoddy_

Acknowledgements JZ acknowledges funding from the Wellcome Trust (203928/Z/16/Z) and support from the National Institute for Health Research Biomedical Research Centre based at Imperial College NHS Trust and Imperial College London. JM is a Wellcome Trust Doctoral Fellow.

Contributors CO and JZ were jointly responsible for study conception and design. CO and JZ performed the process of abstract screening and study selection, with all conflicts resolved by JM. CO was responsible for the curation and reporting of the data, the creation of figures and tables and primary manuscript drafting. All authors contributed to manuscript drafting and revision.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests HA is the chief scientific officer of Preemptive Health and Medicine and Flagship Pioneering. JM was paid directly for giving a lecture at Health Education England on the topic of artificial intelligence in the NHS. JM has been a member of the INSIGHT DataTAB for HDR UK. JZ is employed as a Senior Informatician by Arcturis Data, a Real-World Data company.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Christopher Oddy http://orcid.org/0000-0002-2311-9261

REFERENCES

1 Alonso-Morán E, Nuño-Solinis R, Onder G, *et al*. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. *Eur J Intern Med* 2015;26:182–9.
2 Girwar S-AM, Jabroer R, Fiocco M, *et al*. A systematic review of risk stratification tools internationally used in primary care settings. *Health Sci Rep* 2021;4:e329.
3 Kingston M, Griffiths R, Hutchings H, *et al*. Emergency admission risk stratification tools in UK primary care: a cross-sectional survey of availability and use. *Br J Gen Pract* 2020;70:e740–8.
4 NHS England. Risk stratification. 2023. Available: https://www.england.nhs.uk/ig/risk-stratification/ [Accessed 28 Nov 2023].
5 Mora J, Iturralde MD, Prieto L, *et al*. Key aspects related to implementation of risk stratification in health care systems-the ASSEHS study. *BMC Health Serv Res* 2017;17:331.
6 Reddy A, Sessums L, Gupta R, *et al*. Risk stratification methods and provision of care management services in comprehensive primary care initiative practices. *Ann Fam Med* 2017;15:451–4.
7 Swarthout M, Bishop MA. Population health management: review of concepts and definitions. *Am J Health Syst Pharm* 2017;74:1405–11.
8 Hudon C, Chouinard M-C, Lambert M, *et al*. Effectiveness of case management interventions for frequent users of healthcare services: a scoping review. *BMJ Open* 2016;6:e012353.
9 Stokes J, Panagioti M, Alam R, *et al*. Effectiveness of case management for 'at risk' patients in primary care: a systematic review and meta-analysis. *PLoS ONE* 2015;10:e0132340.
10 Albertson EM, Chuang E, O'Masta B, *et al*. Systematic review of care coordination interventions linking health and social services for high-utilizing patient populations. *Popul Health Manag* 2022;25:73–85.
11 Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. *J Am Med Inform Assoc* 2007;14:269–77.
12 O'Caoimh R, Cornally N, Weathers E, *et al*. Risk prediction in the community: a systematic review of case-finding instruments that predict adverse Healthcare outcomes in community-dwelling older adults. *Maturitas* 2015;82:3–21.
13 Wallace E, Stuart E, Vaughan N, *et al*. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. *Med Care* 2014;52:751–65.
14 Graphnet Health. Carecentric: shared care record solution. 2023. Available: https://www.graphnethealth.com/solutions/integrated-digital-care-record/carecentric/ [Accessed 28 Nov 2023].
15 Prescribing Serives Ltd. Impact assessment 2015. 2015. Available: https://www.eclipsesolutions.org/impactassessment/ [Accessed 28 Nov 2023].
16 Tal E. Target specification bias, counterfactual prediction, and Algorithmic fairness in Healthcare. AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society; Montreal QC Canada, 2023
17 DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc* 2020;27:2020–3.
18 Flores L, Kim S, Young SD. Addressing bias in artificial intelligence for public health surveillance. *J Med Ethics* 2024;50:190–4.
19 Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health* 2019;9:020318.
20 Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. *Int J Med Inform* 2014;83:368–75.
21 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate Healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
22 Zhang J, Symons J, Agapow P, *et al*. Best practices in the real-world data life cycle. *PLOS Digit Health* 2022;1:e0000003.

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

23 McGinn TG, Guyatt GH, Wyer PC, *et al*. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. *JAMA* 2000;284:79–84.
24 Wolff RF, Moons KGM, Riley RD, *et al*. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.
25 Snooks H, Bailey-Jones K, Burge-Jones D, *et al*. Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC). *Health Serv Deliv Res* 2018;6:1–164.
26 Wallace E, McDowell R, Bennett K, *et al*. Comparison of count-based multimorbidity measures in predicting emergency admission and functional decline in older community-dwelling adults: a prospective cohort study. *BMJ Open* 2016;6:e013089.
27 Wahls TL, Barnett MJ, Rosenthal GE. Predicting resource utilization in a veterans health administration primary care population comparison of methods based on diagnoses and medications. *Med Care* 2004;42:123–8.
28 Van Houtte C, Gellen C, Ranchhod D. Acute admission risk stratification of New Zealand primary care patients using demographic, Multimorbidity, service usage and modifiable variables. *J Prim Health Care* 2022;14:116–23.
29 Tomlin AM, Lloyd HS, Tilyard MW. Risk stratification of New Zealand general practice patients for emergency admissions in the next year: adapting the Peony model for use in New Zealand. *J Prim Health Care* 2016;8:227–37.
30 Takahashi PY, Pecina JL, Upatising B, *et al*. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. *Arch Intern Med* 2012;172:773–9.
31 Sylvia ML, Griswold M, Dunbar L, *et al*. Guided care: cost and utilization outcomes in a pilot study. *Dis Manag* 2008;11:29–36.
32 Soto-Gordoa M, de Manuel E, Fullaondo A, *et al*. Impact of stratification on the effectiveness of a comprehensive patient-centered strategy for multimorbid patients. *Health Serv Res* 2019;54:466–73.
33 Sicras-Mainar A, Velasco-Velasco S, Navarro-Artieda R, *et al*. Obtaining the mean relative weights of the cost of care in Catalonia (Spain): retrospective application of the adjusted clinical groups case-mix system in primary health care. *J Eval Clin Pract* 2013;19:267–76.
34 Sicras-Mainar A, Navarro-Artieda R, ACG-BSA RGE. Validating the adjusted clinical groups ACG case-mix system in a Spanish population setting: a multicenter study. *Gac Sanit* 2009;23:228–31.
35 Sicras-Mainar A. Retrospective application of adjusted clinical groups (ACGs) at a primary care centre. *Aten Primaria* 2006;37:439–45.
36 Sibley LM, Moineddin R, Agha MM, *et al*. Risk adjustment using administrative data-based and survey-derived methods for explaining physician utilization. *Med Care* 2010;48:175–82.
37 Shannon GR, Wilber KH, Allen D. Reductions in costly healthcare service utilization: findings from the care advocate program. *J Am Geriatr Soc* 2006;54:1102–7.
38 Shadmi E, Balicer RD, Kinder K, *et al*. Assessing socioeconomic health care utilization inequity in Israel: impact of alternative approaches to morbidity adjustment. *BMC Public Health* 2011;11:609.
39 Reilly S, Abell J, Brand C, *et al*. Case management for people with long-term conditions: impact upon emergency admissions and associated length of stay. *Prim Health Care Res Dev* 2011;12:223–36.
40 Reid RJ, MacWilliam L, Verhulst L, *et al*. Performance of the ACG case-mix system in two Canadian provinces. *Med Care* 2001;39:86–99.
41 Rea F, Corrao G, Ludergnani M, *et al*. A new population-based risk stratification tool was developed and validated for predicting mortality, hospital admissions, and health care costs. *J Clin Epidemiol* 2019;116:62–71.
42 Rahimian F, Salimi-Khorshidi G, Payberah AH, *et al*. Predicting the risk of emergency admission with machine learning: development and validation using linked electronic health records. *PLoS Med* 2018;15:e1002695.
43 Petersen LA, Pietz K, Woodard LD, *et al*. Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. *Med Care* 2005;43:61–7.
44 Orueta JF, García-Alvarez A, Aurrekoetxea JJ, *et al*. FINGER (forming and identifying new groups of expected risks): developing and validating a new predictive model to identify patients with high Healthcare cost and at risk of admission. *BMJ Open* 2018;8:e019830.
45 Orueta J-F, Urraca J, Berraondo I, *et al*. Adjusted clinical groups (Acgs) explain the utilization of primary care in Spain based on information registered in the medical records: a cross-sectional study. *Health Policy* 2006;76:38–48.
46 Mosley DG, Peterson E, Martin DC. Do Hierarchical condition category model scores predict hospitalization risk in newly enrolled Medicare advantage participants as well as probability of repeated admission scores *J Am Geriatr Soc* 2009;57:2306–10.
47 Maltenfort MG, Chen Y, Forrest CB. Prediction of 30-day pediatric unplanned hospitalizations using the Johns Hopkins adjusted clinical groups risk adjustment system. *PLoS One* 2019;14:e0221233.
48 Lugo-Palacios DG, Hammond J, Allen T, *et al*. The impact of a combinatorial digital and organisational intervention on the management of long-term conditions in UK primary care: a non-randomised evaluation. *BMC Health Serv Res* 2019;19:159.
49 López-Aguilà S, Contel JC, Farré J, *et al*. Predictive model for emergency hospital admission and 6-month readmission. *Am J Manag Care* 2011;17:e348–57.
50 Liu C-F, Sales AE, Sharp ND, *et al*. Case-mix adjusting performance measures in a veteran population: pharmacy-and diagnosis-based approaches. *Health Serv Res* 2003;38:1319–37.
51 Levine S, Steinman BA, Attaway K, *et al*. Home care program for patients at high risk of hospitalization. *Am J Manag Care* 2012;18:e269–76.
52 Lemke KW, Weiner JP, Clark JM. Development and validation of a model for predicting inpatient hospitalization. *Med Care* 2012;50:131–9.
53 Kronick R, Gilmer T, Dreyfus T, *et al*. Improving health-based payment for medicaid beneficiaries: CDPS. *Health Care Financ Rev* 2000;21:29–64.
54 Khanna S, Rolls DA, Boyle J, *et al*. A risk stratification tool for Hospitalisation in Australia using primary care data. *Sci Rep* 2019;9:5011.
55 Jung D, Pollack HA, Konetzka RT. Predicting hospitalization among medicaid Home- and community-based services users using machine learning methods. *J Appl Gerontol* 2023;42:241–51.
56 Juncosa S, Bolíbar B, Roset M. Performance of an ambulatory Casemix measurement system in primary care in Spain: ambulatory care groups (ACGs). *The European Journal of Public Health* 1999;9:27–35.
57 Inouye SK, Zhang Y, Jones RN, *et al*. Risk factors for hospitalization among community-dwelling primary care older patients: development and validation of a predictive model. *Med Care* 2008;46:726–31.
58 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:e003482.
59 Haas LR, Takahashi PY, Shah ND, *et al*. Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care* 2013;19:725–32.
60 Greenwald S, Chamoun GF, Chamoun NG, *et al*. Risk stratification index 3.0, a broad set of models for predicting adverse events during and after hospital admission. *Anesthesiology* 2022;137:673–86.
61 Fishman PA, Goodman MJ, Hornbrook MC, *et al*. Risk adjustment using automated ambulatory Pharmacy data the RxRisk model. *Med Care* 2003;41:84–99.
62 Falasca P, Berardo A, Di Tommaso F. Development and validation of predictive Mosaico (Modello Statistico Combinato) on emergency admissions: can it also identify patients at high risk of frailty? *Ann Ist Super Sanita* 2011;47:220–8.
63 Dudley RA, Medlin CA, Hammann LB, *et al*. The best of both worlds?: potential of hybrid prospective/concurrent risk adjustment. *Med Care* 2003;41:56–69.
64 Donnan PT, Dorward DWT, Mutch B, *et al*. Development and validation of a model for predicting emergency admissions over the next year (PEONY): A UK historical cohort study. *Arch Intern Med* 2008;168:1416–22.
65 Cumming RB, Knutson DJ, Cameron BA, *et al*. A comparative analysis of claims-based methods of health risk assessment for commercial populations. *Society of Actuaries* 2002. Available: https://web.archive.org/web/20201107012013id_/https://www.soa.org/globalassets/assets/Files/Research/Projects/2005-comp-analysis-methods-commercial-populations.pdf
66 Crane SJ, Tung EE, Hanson GJ, *et al*. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. *BMC Health Serv Res* 2010;10:338.
67 Chen S, Bergman D, Miller K, *et al*. Using applied machine learning to predict healthcare utilization based on socioeconomic determinants of care. *Am J Manag Care* 2020;26:26–31.
68 Charlson ME, Charlson RE, Peterson JC, *et al*. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *J Clin Epidemiol* 2008;61:1234–40.

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.

69 Brilleman SL, Gravelle H, Hollinghurst S, *et al*. Keep it simple? Predicting primary health care costs with clinical morbidity measures. *J Health Econ* 2014;35:109–22.
70 Brilleman SL, Salisbury C. Comparing measures of Multimorbidity to predict outcomes in primary care: a cross sectional study. *Fam Pract* 2013;30:172–8.
71 Boult C, Reider L, Leff B, *et al*. The effect of guided care teams on the use of health services: results from a cluster-randomized controlled trial. *Arch Intern Med* 2011;171:460–6.
72 Billings J, Dixon J, Mijanovich T, *et al*. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ* 2006;333:327.
73 Baker A, Leak P, Ritchie LD, *et al*. Anticipatory care planning and integration: a primary care pilot study aimed at reducing unplanned Hospitalisation. *Br J Gen Pract* 2012;62:e113–20.
74 Ash AS, Ellis RP. Risk-adjusted payment and performance assessment for primary care. *Med Care* 2012;50:643–53.
75 Aguado A, Guinó E, Mukherjee B, *et al*. Variability in prescription drug expenditures explained by adjusted clinical groups (ACG) case-mix: a cross-sectional study of patient electronic records in primary care. *BMC Health Serv Res* 2008;8.
76 Levine S, Adams J, Attaway K, *et al*. Predicting the financial risks of seriously ill patients. 2011. Available: https://www.hmprg.org/wp-content/themes/HMPRG/backup/GOHIT/Work%20Groups/IDSR/August%2027%20-%20Predicting%20the%20Financial%20Risks%20of%20Seriously%20Ill%20Patients.pdf [Accessed 28 Nov 2023].
77 Hutchings HA, Evans BA, Fitzsimmons D, *et al*. Predictive risk stratification model: a progressive cluster-randomised trial in chronic conditions management (PRISMATIC) research protocol. *Trials* 2013;14:301.
78 Billings J, Mijanovich T, Dixon J, *et al*. Case finding Algorithms for patients at risk of re-Hospitalisation PARR1 and PARR2. 2006. Available: https://www.kingsfund.org.uk/sites/default/files/field/field_document/PARR-case-finding-algorithms-feb06.pdf [Accessed 28 Nov 2023].
79 Wilber KH, Allen D, Shannon GR, *et al*. Partnering managed care and community-based services for frail elders: the care advocate program. *J Am Geriatr Soc* 2003;51:807–12.
80 Johns Hopkins Medicine. Johns Hopkins ACG system. 2023. Available: https://www.hopkinsacg.org/ [Accessed 28 Nov 2023].
81 Pope G, Ellis R, Ash A, *et al*. Diagnostic cost group hierarchical condition category models for Medicare risk adjustment. Health Care Financing Administration. 2000. Available: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/reports/downloads/pope_2000_2.pdf [Accessed 28 Nov 2023].
82 Optum. Symmetry episode risk groups. 2022.
83 Wennberg D, Dixon J, Billings J, *et al*. Combined predictive model final report. 2006. Available: https://www.kingsfund.org.uk/sites/default/files/field/field_document/PARR-combined-predictive-model-final-report-dec06.pdf [Accessed 28 Nov 2023].
84 Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
85 Exley J, Abel GA, Fernandez J-L, *et al*. Impact of the Southwark and Lambeth integrated care older people's programme on hospital utilisation and costs: controlled time series and cost-consequence analysis. *BMJ Open* 2019;9:e024220.
86 Gravelle H, Dusheiko M, Sheaff R, *et al*. Impact of case management (Evercare) on frail elderly patients: controlled before and after analysis of quantitative outcome data. *BMJ* 2007;334:31.
87 Stokes J, Kristensen SR, Checkland K, *et al*. BMJ open effectiveness of multidisciplinary team case management: difference-Indifferences analysis. *BMJ Open* 2016;6:e010468.
88 Tricco AC, Antony J, Ivers NM, *et al*. Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services: a systematic review and meta-analysis. *CMAJ* 2014;186:E568–78.
89 Vanderplasschen W, Rapp RC, De Maeyer J, *et al*. A meta-analysis of the efficacy of case management for substance use disorders: a recovery perspective. *Front Psychiatry* 2019;10:186.
90 Backhouse A, Ukoumunne OC, Richards DA, *et al*. The effectiveness of community-based coordinating interventions in dementia care: a meta-analysis and subgroup analysis of intervention components. *BMC Health Serv Res* 2017;17:717.
91 Saragih ID, Saragih IS, Tarihoran DETAU, *et al*. A meta-analysis of studies of the effects of case management intervention for stroke survivors across three countries. *J Nurs Scholarsh* 2023;55:345–55.
92 Gorin SS, Haggstrom D, Han PKJ, *et al*. Cancer care coordination: a systematic review and meta-analysis of over 30 years of empirical studies. *Ann Behav Med* 2017;51:532–46.
93 Scholte NTB, Gürgöze MT, Aydin D, *et al*. Telemonitoring for heart failure: a meta-analysis. *Eur Heart J* 2023;44:2911–26.
94 Umeh CA, Torbela A, Saigal S, *et al*. Telemonitoring in heart failure patients: systematic review and meta-analysis of randomized controlled trials. *World J Cardiol* 2022;14:640–56.

BMJ Health Care Inform: first published as 10.1136/bmjhci-2024-101065 on 19 June 2024. Downloaded from http://informatics.bmj.com/ on July 9, 2024 at Imperial College London Library. Protected by copyright.