Journal of Clinical Epidemiology

Inconsistencies in handling missing data across stages of prediction modelling: a review of methods used

--Manuscript Draft--

Manuscript Number:	
Article Type:	Review Article
Keywords:	Statistical models; Prognosis; Predictive medicine; Missing data
Corresponding Author:	Antonia Dimitrova Tsvetanova UNITED KINGDOM
First Author:	Antonia Dimitrova Tsvetanova
Order of Authors:	Antonia Dimitrova Tsvetanova
	Matthew Sperrin
	Niels Peek
	Iain Buchan
	Stephanie Hyland
	Glen Martin
Manuscript Region of Origin:	Europe
Abstract:	Objective: No clear guidance exists on handling missing data at each stage of developing, validating and implementing a clinical prediction model (CPM). We aimed to review the approaches to handling missing data that underly the CPMs currently recommended for use in UK healthcare. Study design and Setting: A methods review to identify CPMs recommended by the National Institute for Health and Care Excellence (NICE), which summarized how missing data is handled across their pipelines. Results: 23 CPMs were included. Six missing data strategies were identified: complete case analysis (CCA), multiple imputation, imputation of mean values, k-nearest neighbours imputation, using an additional category for missingness, considering missing values as risk-factor-absent. 52% of the development articles and 48% of the validation articles did not report how missing data were handled. CCA was the most common approach used for development (40%) and validation (44%). At implementation, 57% of the CPMs required complete data entry, whilst 43% allowed missing values. 3 CPMs had consistent paths in their pipelines. Conclusion: A broad variety of methods for handling missing data underly the CPMs currently recommended for use in UK healthcare. Missing data handling strategies were generally inconsistent. Better quality assurance of CPMs needs greater clarity and consistency in handling of missing data.
Suggested Reviewers:	Angela Wood University Lecturer in Biostatistics, University of Cambridge amw79@medschl.cam.ac.uk
	Maarten van Smeden Assistant Professor, UMC Utrecht M.vanSmeden@umcutrecht.nl
	Rolf Groenwold Professor of Clinical Epidemiology, Leiden University: Universiteit Leiden r.h.h.groenwold@lumc.nl
Opposed Reviewers:	

Centre for Health Informatics
University of Manchester
Manchester
M13 9PL
antonia.tsvetanova@manchester.ac.uk

Editor, Journal of Clinical Epidemiology

XX February 2021

Inconsistencies in handling missing data across stages of prediction modelling: a review of methods used

Authors: Tsvetanova et al.

Dear Editor,

Thank you for the opportunity to submit the above titled manuscript for consideration of publication in the *Journal of Clinical Epidemiology*.

Our paper is a review of methods used to handle missing data across the pipeline of a clinical prediction model (CPM). Commonly, CPMs are increasingly developed and validated using routinely collected data, such as electronic health records, where missing data are present and need careful handling. Whilst methods for handling missing data have received considerable attention at the development stage of CPMs, we feel that this has been under-explored in the external validation and implementation stages. In turn, this lack of guidance has potentially resulted in inconsistency between imputation methods used across the stages of development, validation, and implementation of a CPM. Notably, the extent to which these methods are inconsistent is currently unclear.

In the paper we review twenty three CPMs that are currently recommended for use in the UK healthcare and summarise the methods used to handle missing data at the development, external validation and implementation stages. This paper will be of a value to those, facing the issue of missing data in prediction modelling.

We have uploaded our manuscript and supplemental material.

There are no conflicts of interest.

This manuscript arises from original research; has not been published elsewhere; and is not currently under submission elsewhere.

As corresponding author, I have had full access to all the data and have had final responsibility for the decision to submit the manuscript for publication. On behalf of my co-authors, I hope that you will find our paper of sufficient interest and quality for publication in the *Journal of Clinical Epidemiology.*

Yours sincerely,

Antonia Tsvetanova, BSc (Hons) MRes
PhD candidate
University of Manchester

What is New?

Key findings

- To date, missing data has been well-addressed at the development stage of a CPM, whilst it has been under-explored in the stages of external validation and implementation.
- Missing data has been generally handled inconsistently across the pipeline of CPMs used in the UK healthcare.
- Missing data and their handling has been poorly reported and accounted for.

What this adds to what was known

- No examples in which missing data were allowed in practice and where missing data were handled consistently between validation and implementation stages was found.

What is the implication and what should change now?

- A framework for handling missing data is needed to quality assure CPM pipelines.

Inconsistencies in handling missing data across stages of prediction modelling: a review of methods used

Antonia Tsvetanova[1], Matthew Sperrin[1], Niels Peek[1,2], Iain Buchan[1,4], Stephanie Hyland[3], Glen P. Martin[1]

[1]Centre for Health Informatics, Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[2]NIHR Manchester Biomedical Research Centre, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

[3]Microsoft Research Cambridge, UK

[4]Institute of Population Health, The University of Liverpool

Funding: This work was supported by EPSRC iCASE and Microsoft Research through its PhD Scholarship Programme.

Competing Interests: None to declare.

Corresponding Author:
Antonia Tsvetanova
PhD candidate
Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester,
Vaughan House, Portsmouth Street, M13 9GB Manchester, UK
Email: antonia.tsvetanova@manchester.ac.uk

Abstract

Objective: No clear guidance exists on handling missing data at each stage of developing, validating and implementing a clinical prediction model (CPM). We aimed to review the approaches to handling missing data that underly the CPMs currently recommended for use in UK healthcare.

Study design and Setting: A methods review to identify CPMs recommended by the National Institute for Health and Care Excellence (NICE), which summarized how missing data is handled across their pipelines.

Results: 23 CPMs were included. Six missing data strategies were identified: complete case analysis (CCA), multiple imputation, imputation of mean values, k-nearest neighbours imputation, using an additional category for missingness, considering missing values as risk-factor-absent. 52% of the development articles and 48% of the validation articles did not report how missing data were handled. CCA was the most common approach used for development (40%) and validation (44%). At implementation, 57% of the CPMs required complete data entry, whilst 43% allowed missing values. 3 CPMs had consistent paths in their pipelines.

Conclusion: A broad variety of methods for handling missing data underly the CPMs currently recommended for use in UK healthcare. Missing data handling strategies were generally inconsistent. Better quality assurance of CPMs needs greater clarity and consistency in handling of missing data.

Keywords:

Statistical models; Prognosis; Predictive medicine; Missing data; Imputation.

Running title: Inconsistencies in handling missing data across stages of prediction modelling: a review of methods used

Word count: 3055

What is New?

Key findings

- To date, missing data has been well-addressed at the development stage of a CPM, whilst it has been under-explored in the stages of external validation and implementation.
- Missing data has been generally handled inconsistently across the pipeline of CPMs used in the UK healthcare.
- Missing data and their handling has been poorly reported and accounted for.

What this adds to what was known

- No examples in which missing data were allowed in practice and where missing data were handled consistently between validation and implementation stages was found.

What is the implication and what should change now?

- A framework for handling missing data is needed to quality assure CPM pipelines.

1. Introduction

Clinical prediction models (CPMs) are statistical models or algorithms that use a set of predictor variables to calculate an individual's chance of developing or having a certain condition, and thus aid clinicians with the associated clinical reasoning and decision-making[1]. Three major phases can be identified in the CPM pipeline: (i) developing and internally validating a CPM; (ii) validating the model on new independent cohorts of patients (external validation), potentially adjusting or updating the model as needed; and (iii) implementing the model in clinical practice while monitoring its impacts[2].

CPMs are increasingly developed and validated using routinely collected data, such as electronic health records, where missing data are common and need careful handling in the CPM pipeline[3,4]. A simple, but inefficient (and potentially biased), approach is complete case analysis (CCA), where all patients with missing values are excluded. This method is only valid in situations where the data are missing completely at random and where there is sufficient sample size after missing case deletion to enable robust inference[5–7]. On the other hand, multiple imputation (MI) is often seen as the gold standard as it allows all data to be used, makes the *weaker* missing (*not completely*) at random assumption, and appropriately accounts for uncertainty in the missing data[8,9]. MI, however, is not easily used in clinical contexts, because we are dealing with one patient at the time and it often requires knowledge about the outcome that is not yet available[10–12].

Whilst methods for handling missing data have received considerable attention at the development stage of CPMs, there is a lack of research exploring handling missing data in external validation and implementation stages. In turn, this lack of guidance has potentially resulted in inconsistency between imputation methods used across the stages of development, validation, and implementation of a CPM. However, the extent to which missing data handling methods are consistent, or otherwise, across the CPM pipeline is currently unclear. Hoogland *et al* claimed that they have not been able to find an example, in which missing data were allowed in practice and where missing data were handled in a consistent way across validation and implementation[13]. Notably, any lack of consistency might lead to overly optimistic (or pessimistic) assessments of model performance estimated at development and validation stages of a CPM, compared with performance at time of implementation[14,15]. For example, suppose a CPM is developed and validated using CCA, then at implementation it is applied to all patients with missing data handled using mean imputation – that would be inconsistent because the natural variability of the data will be compromised if all the missing values are imputed by the mean of those, which are available.

Consistency of imputation methods across the stages of the CPM pipeline would help ensure that the predictive performance reported from external validation studies is based upon consistent methods (in terms of handling missing data) with those to be applied when the model is implemented.

The aim of this study was to review CPMs recommended for use in UK healthcare and to summarise the methods used to address missing data across the models' (i) development, (ii) external validation and (iii) implementation stages – the CPM pipeline.

2. Methods

We sought to identify and describe existing clinical prediction models used in UK clinical practice, with respect to missing data handling across their CPM pipeline. We only included models that are recommended by the National Institute for Health and Care Intelligence (NICE), which has a role in weighing the evidence around CPMs. We chose this approach, to ensure coverage of the whole CPM pipeline, noting that many models are developed but not validated or implemented[16]. We asked NICE for a list of CPMs they recommend. To widen the search, we added further CPMs identified in our earlier research and, we reached-out to the scientific community on Twitter. The tweet by AT on March 25th 2020 was seen 11,286 times, receiving 8 replies and 13 re-tweets – of these, we only included any CPMs that NICE recommends but had not mention in their list. A final list of CPMs was established and data were summarized in an information extraction table.

2.1. Search Strategy and selection criteria

We used Google Scholar to search for the original development papers, and papers that aimed to externally validate the CPMs. The search for the original development papers included synonyms for [development] combined with [prognostic/predictive/prediction model], and [developer's name], with the latter being identified by using MDCalc free online medical reference source[17]. The search for external validation papers was performed using forward citations from each of the original development papers, followed by *search within citing articles* option. This option assures that all the validation articles for the specific CPM, have cited the original derivation article of the CPM. The search terms for validation papers included [name of the CPM] – both abbreviated and extended. As a pragmatic approach to maintain a viable number of papers to screen, we selected the top ten most cited (according to Google scholar as-of 18/09/20) validation papers for each CPM. For validation papers, articles were included if they: (i) performed independent external validation of the CPMs; or (ii) reported a comparison between the CPM of interest and another CPM(s) within independent data. Articles that had a primary purpose of developing a new prediction model and comparing it with the CPM of

interest were excluded. We preferred to get information on the implementation stage for each CPM from documentation provided by the CPMs' developers. If this was not available, we obtained the information from the online tools of the model (e.g., stand-alone online calculator website or part of MDCalc). Each tool was tested to assess 1) if it is possible to obtain a prediction with missing information or not, and 2) if so, how the missing data were being handled by the CPM. The information extraction for all stages was completed on 22/09/20.

2.2. Data extraction and synthesis

Data extraction of included articles was completed by one author (AT). Additionally, 10% of the data extraction was independently undertaken by a second author (GPM). We categorised the eligible articles into two groups: development articles and external validation articles. The following details for each paper were extracted:

(i) For both development and validation studies, general information such as: author, year of publication and paper title.

(ii) For both development and validation studies, data surrounding the source of data used (e.g., cohort, case-control, registry), the sample size and the outcome of interest.

(iii) For development studies: the modelling method used (e.g., logistic, survival), and software used for analysis.

(iv) For both development and validation studies, missing data handling approach (e.g., complete case analysis, imputation methods, other or none reported).

(v) For both development and validation studies, model performance, reported strengths and limitations of the studies and any stated assumptions made.

Information on the implementation stage of each CPM was extracted from the original online calculator of the CPM, or from MDCalc.

The Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement has been applied to all the articles included in this review – items 9, 13a and 13b[18].

When extracting information on missing data, studies that have reported that missing data is present, but have not described how it was handled, were put in the category 'unclear'. This was used for item 13b. Studies where no information on the presence of missing data or any handling method was reported, were categorised as 'not reported'.

2.3. Definition of consistency across CPM pipeline

We define 'consistency' as where the missing data handling method used at *validation* is compatible with the method to be used *at* the *implementation* stage. (Note that no reference made to development approach). *Compatibility* means that the validation approach will accurately reflect the performance at implementation under identical missingness mechanisms (e.g. in our case, only permitting MAR). The following cases can be considered consistent:

(i) Where the same method is used at both validation and implementation stages (noting this excludes *CCA – all data required*, as this requires MCAR, but includes *MI – MI*, although this is never observed in practice.)

(ii) *MI at validation – all data required* at implementation, since MI is designed to reflect this appropriately under MAR.

For a CPM pipeline to be considered as having compatible missing data handling methods across its development, validation and implementation stages, all combinations of methods must comply with the above definition.

3. Results

The search strategy identified 23 clinical prediction models that met the eligibility criteria, as shown on Figure 1. (The original list of all CPMs could be found in Appendix A) Description of each CPM and its corresponding external validation papers are summarised in Table 1.

In total, information from 233 articles was extracted. Development articles were available for 23 out of 24 CPMs. There was one development article (Waterlow Score), for which access to the published paper could not be obtained, therefore, we did not consider this model further. A total of 210 external validation articles were included in this study. For six out of 23 CPMs, there were less than 10 validation papers available, when the search criteria were applied (QRISK, Thoracoscore, The Leicester practice risk score, FRAX, BOADICEA and NEWS2).

3.1. Missing Data

Six missing data approaches were identified through the literature search within the development and validation papers, summarised in table 2.

3.1.1. Missing data handling at Development stage of a CPM

From the 23 development papers, 12 (52%) did not report how the analysis handled missing data. The most common method for missing data handling was CCA, used in ten (44%) articled. MI was used in only one development article (4%).

3.1.2. Missing data handling at Validation stage of a CPM

From the 210 external validation articles, the approach to handling missing data was not reported in 100 studies (48%). As with the development studies, CCA was the most common method used in the validation articles (n=85, 40%). Multiple imputation was used in thirteen (6%) of the 210 studies. Twelve studies (6%) used 'other methods' such as single imputation, k-nearest neighbour (KNN) imputation, additional category for missing values (missing indicator method) or missing values considered as normal (e.g., if a comorbidity is not recorded, it is assumed to be absent).

3.1.3. Missing data handling at the Implementation stage of a CPM

When applied to an individual patient during the CPM's (iii) implementation stage, only one (4%) of the 23 CPMs (QRISK[254]) used mean imputation for a measure of deprivation when geographical region is unknown, conditional mean imputation based on ethnicity, age and sex, if there are missing values of Cholesterol/HDL ratio, blood pressure and BMI, and it uses zero imputation when the SD of the last two blood pressure readings is missing[254]. Eight CPMs (35%) make the assumption for missing values to be the lowest/normal – ABCD2[255], TIMI[256], CRB65[257], EuroScore[258], FRAX[259], CHADVASC[260], DG-ROMA[261], NEWS2[262]. Overall, thirteen CPMs (57%) require that all data is present when making a prediction at implementation stage; these models were Thoracoscore[263], NPI[264], Leicester Risk Score[265], PREDICT[266], Blatchford[267], HAS-BLED[268], GRACE[269], Framingham[270], Gleason[271], Braden Scale[272], APACHE[273], APGAR[274], MTS[76]. The remaining CPM – BOADICEA[275], (4%) has a category 'unknown' for non-continuous variables, however it is unclear what assumptions have been made in relation to missing data handling.

3.2. Consistency of missing data handling across the pipeline of a CPM

Overall, results showed that missing data were generally handled in an inconsistent way across the pipeline of a CPM, according to our definitions of 'consistency' as pictured on Figure 2. Consistent 'paths' (MI at validation – All data required at implementation) were observed for only three CPMs: Thoracoscore, Manchester Triage system, APGAR score (see Supplementary material).

3.3. TRIPOD statement

Across all the development and external validation articles included, 27 studies (12%) stated that potential bias and inconsistency in the results might have occurred owing to missing data. Of the development studies, 11 (48%) have followed checklist item 9 from the TRIPOD statement, which is to describe how missing data are handled. Similarly, from the 210 external validation studies, 112 (53%) have followed checklist item 9. All the development and

validation articles have described the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time (checklist item 13a from the TRIPOD statement). Of the development studies 12 (52%) have described the characteristics of the participants, including the number of participants with missing data (checklist item 13b). 126 (60%) of the validation studies have followed this point.

4. Discussion

In this review of CPMs recommended for use by NICE, we showed that there are inconsistencies across the CPMs pipeline with regards to missing data handling approaches. We found that only three CPMs met one of the two definitions for consistency in handling missing data that we have proposed. Indeed, Thoracoscore, Manchester Triage System and APGAR score had consistent paths (MI at validation – All data required at implementation). We considered this consistent, since MI is designed to reflect this appropriately under the weaker MAR assumption for missingness. We did not find any consistency paths where the same approach of handling missing data, is used in validation and implementation stages (MI – MI). This case has not been observed in practice, perhaps since it is challenging to use MI at implementation stage, where extra information and potentially other data is needed. Although, CCA – All data required might be also considered as 'consistent' by some, since it uses the same approach throughout the pipeline, we have excluded it has a possible 'consistent' path option, because it requires the missingness mechanism to be MCAR, which is rarely the case. Finally, we did not find any cases where missing data were allowed in practice, which is in line with the statement made by Hoogland *et al*[13]. It is less clear what the prediction made by the CPM will be if we allow missing data to occur at implementation, because the missing mechanism is likely to be different from the one at development and external validation stages[276].

CCA was the commonest approach for handling missing data at derivation and external validation of a CPM. Almost half of the studies did not report how missing data were handled. Furthermore, with exception of a few studies, no assumptions were made *explicit* in regard to the mechanisms of missingness. Overall, only half of the studies have adhered to the items from the TRIPOD statement that describe how missing data should be reported[277]. This could be because most of the articles were published before the TRIPOD was published in 2015.

When implementing CPMs in practice, methods for dealing with missing values are often driven by practical constraints. Issues with the applicability of multiple imputation (MI) methods

arise when the researcher only has access to published parameter estimates (for the CPM). For example, during the external validation and implementation stages, to make predictions for a new individual with missing data, one would need extra information, such as the imputation models and potentially other data. In most cases, this is impractical (for prediction models) in real world settings, although a recent work from Nijman et al suggest how this can be avoided[278]. Furthermore, there are alternative to MI methods emerging, such as the pattern sub-model, where one fits a pattern mixture model for every missing data pattern, using only data from that pattern[4].

The issue not well addressed in the literature, however, remains the extent to which these methods affect CPM performance. We currently do not know whether there is an effect in using different missing data approaches are used across the pipeline, although some of the articles suggested that potential bias in the results might have occurred due to missing data. One study has stated that ROC, D and R^2 statistics were not similar between patients with complete data when compared to the results obtained using multiply imputed data set[28]. Another study has stated that the presence of a specific variable could have changed the coefficients in the remaining variables[32]. Furthermore, it has been also pointed out that missing data impeded the categorization of some of the patients, which, in turn impaired the ability to validate the CPMs more definitively[121,251]. Many studies have stated that missing data could have affected their findings[51,54,77,95,238]. We propose that the way missing data are handled during validation should be compatible to that which will be used when the CPM is implemented. Future work should explore the effect of incompatibility in terms of reported/estimated predictive performance.

4.1. Strengths and limitations

To our knowledge, this is the first review of how missing data have been handled across the development, external validation and implementation pipeline of CPMs recommended for use in UK clinical practice. Although our search for CPMs did not involve a systematic review of literature, we used standard reporting guidelines, such as TRIPOD, to evaluate each included model.

Our review has certain limitations that are worth mentioning. First, we did not have a systematic way to search for CPMs within the NICE guidelines. Therefore, we asked NICE directly to provide us with a list of CPM they recommend, the scientific community on Twitter, and our research group – of the latter two, we filtered only those CPMs that are recommended in the NICE guidelines, in case they were not mentioned by NICE themselves. Therefore, our study might suffer lack of generalizability, since other CPMs might be recommended for use

in other healthcare settings and our review clearly could not cover all the existing CPMs. However, the included models cover a broad spectrum of clinical areas. Second, the external validation studies criteria included only the top ten most cited articles for each CPM as a pragmatic approach to maintain a viable number of papers to screen. Although there are some models, for which the total number of validation studies was less than or equal to 10 (n=7 CPMs), for the majority of the CPMs, especially those developed before 2010, this inclusion criteria would have excluded some validations. Thus, other validation studies might have applied alternative imputation strategies to those covered here.

5. Conclusion

We found considerable diversity, inconsistency and lack of reported detail in how missing data are handled across the development, external validation, and implementation stages of 23 CPMs currently recommended for use in UK healthcare. A framework for handling missing data is needed to quality assure CPM pipelines.

Acknowledgements: We thank the National Institute for Health and Care Excellence (NICE) for providing us with a list of clinical prediction models that they recommend. In particular, Judith Richardson, Kay Nolan and Pall Jonsson for conducting a search within their guidelines and for the useful conversations.

References

1. Steyerberg, E. W. Clinical Prediction Models: Application of Prediction Models. *Springer Sci. + Bus. Media* 101–111 (2009) doi:10.1007/978-0-387-77244-8.
2. Riley, R. D., Windt, D. van der, Croft, P. & Moons, K. G. M. *Prognosis research in healthcare : concepts, methods, and impact*. (2019).
3. Goldstein, B. A., Navar, A. M., Pencina, M. J., Ioannidis, J. P. A. & Goldstein, B. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review Correspondence to. doi:10.1093/jamia/ocw042.
4. Mercaldo, S. F. & Blume, J. D. Missing data and prediction: the pattern submodel. *Biostatistics* **21**, 236–252 (2020).
5. Little, R. J. A. & Rubin, D. B. *Statistical analysis with missing data*.
6. Riley, R. D. *et al.* Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat. Med.* **38**, 1276–1296 (2019).
7. Janssen, K. J. M. *et al.* Missing covariate data in medical research: To impute is better than to ignore. *J. Clin. Epidemiol.* **63**, 721–727 (2010).

8. Wood, A. M., Royston, P. & White, I. R. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometrical J.* **57**, 614–632 (2015).
9. Mertens, B. J. A., Banzato, E. & Wreede, L. C. Construction and assessment of prediction rules for binary outcome in the presence of missing predictor data using multiple imputation and cross-validation: Methodological approach and data-based evaluation. *Biometrical J.* **62**, 724–741 (2020).
10. Enders, C. K. Multiple imputation as a flexible tool for missing data handling in clinical research. *Behav. Res. Ther.* **98**, 4–18 (2017).
11. Sterne, J. A. C. *et al.* Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ (Online)* vol. 339 157–160 (2009).
12. Wood, A. M., Royston, P. & White, I. R. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometrical J.* **57**, 614–632 (2015).
13. Hoogland, J. *et al.* Handling missing predictor values when validating and applying a prediction model to new patients. *Stat. Med.* **39**, 3591–3607 (2020).
14. Sperrin, M., Martin, G. P., Sisk, R. & Peek, N. Missing data should be handled differently for prediction than for description or causal explanation. *J. Clin. Epidemiol.* (2020) doi:10.1016/j.jclinepi.2020.03.028.
15. Smeden, M. van, Groenwold, R. H. H. & Moons, K. G. A cautionary note on the use of the missing indicator method for handling missing data in prediction research. *J. Clin. Epidemiol.* **0**, (2020).
16. Collins, G. S. *et al.* External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Medical Research Methodology* vol. 14 40 (2014).
17. MDCalc - Medical calculators, equations, scores, and guidelines. https://www.mdcalc.com/.
18. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Ann. Intern. Med.* **162**, 55–63 (2015).
19. Hippisley-Cox, J. *et al.* Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *Br. Med. J.* **335**, 136–141 (2007).
20. Hippisley-Cox, J. *et al.* Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2. *BMJ* **336**, 1475–1482 (2008).
21. Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease:

Prospective cohort study. *BMJ* **357**, (2017).

22. Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J. & Brindle, P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: A validation study. *Heart* **94**, 34–39 (2008).
23. Collins, G. S. & Altman, D. G. An independent and external validation of QRISK2 cardiovascular disease risk score: A prospective open cohort study. *BMJ* **340**, 1231 (2010).
24. Collins, G. S. & Altman, D. G. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: Independent and external validation of an updated version of QRISK2. *BMJ* **345**, (2012).
25. An independent external validation and evaluation of QRISK CVD risk prediction: a prospective open cohort study - Google Search. https://www.google.com/search?q=An+independent+external+validation+and+evaluation+of+QRISK+CVD+risk+prediction%3A+a+prospective+open+cohort+study&rlz=1C1GCEA_en-GBGB876GB876&oq=An+independent+external+validation+and+evaluation+of+QRISK+CVD+risk+prediction%3A+a+prospective+open+cohort+study&aqs=chrome.0.69i59.321j0j4&sourceid=chrome&ie=UTF-8.
26. Tillin, T. *et al.* Ethnicity and prediction of cardiovascular disease: Performance of QRISK2 and Framingham scores in a UK tri-ethnic prospective cohort study (SABRE-Southall and Brent REvisited). *Heart* **100**, 60–67 (2014).
27. van de Putte, D. E. F. *et al.* Unfavourable cardiovascular disease risk profiles in a cohort of Dutch and British haemophilia patients. *Thromb. Haemost.* **109**, 16–23 (2013).
28. Hippisley-Cox, J., Coupland, C. & Brindle, P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: A validation study. *BMJ Open* **4**, 5809 (2014).
29. Hosein, A., Stoute, V., Chadee, S. & Singh, N. R. Evaluating cardiovascular disease (CVD) risk scores for participants with known CVD and non-CVD in a multiracial/ethnic Caribbean sample. *PeerJ* **2020**, (2020).
30. Bazyani, A. *et al.* QRISK2 score in CABG patients correlated with risk factors. *Rev. Chim.* **70**, 1676–1680 (2019).
31. Falcoz, P. E. *et al.* The Thoracic Surgery Scoring System (Thoracoscore): Risk model for in-hospital death in 15,183 patients requiring thoracic surgery. *J. Thorac. Cardiovasc. Surg.* **133**, 325-332.e1 (2007).
32. Chamogeorgakis, T. P., Connery, C. P., Bhora, F., Nabong, A. & Toumpoulis, I. K.

Thoracoscore predicts midterm mortality in patients undergoing thoracic surgery. *J. Thorac. Cardiovasc. Surg.* **134**, 883–887 (2007).

33. Bradley, A. *et al.* Thoracoscore fails to predict complications following elective lung resection. *Eur. Respir. J.* **40**, 1496–1501 (2012).
34. Chamogeorgakis, T. *et al.* External validation of the modified Thoracoscore in a new thoracic surgery program: Prediction of in-hospital mortality. *Interact. Cardiovasc. Thorac. Surg.* **9**, 463–466 (2009).
35. Barua, A. *et al.* Accuracy of two scoring systems for risk stratification in thoracic surgery. *Interact. Cardiovasc. Thorac. Surg.* **14**, 556–559 (2012).
36. Qadri, S. S. A. *et al.* Could thoracoscore predict postoperative mortality in patients undergoing pneumonectomy? *Eur. J. Cardio-thoracic Surg.* **45**, 864–869 (2014).
37. O'Dowd, E. L. *et al.* Predicting death from surgery for lung cancer: A comparison of two scoring systems in two European countries. *Lung Cancer* **95**, 88–93 (2016).
38. Chesterfield-Thomas, G. & Goldsmith, I. Impact of preoperative pulmonary rehabilitation on the Thoracoscore of patients undergoing lung resection. *Interact. Cardiovasc. Thorac. Surg.* **23**, 729–732 (2016).
39. Sharkey, A. J. *et al.* Thoracoscore and European Society Objective Score fail to predict mortality in the UK. *CJAM Can. J. Addict. Med.* **6**, 270–275 (2015).
40. Die Loucou, J. *et al.* Validation and update of the thoracic surgery scoring system (Thoracoscore) risk model. *Eur. J. Cardiothorac. Surg.* **58**, 350–356 (2020).
41. Haybittle, J. L. *et al.* A prognostic index in primary breast cancer. *Br. J. Cancer* **45**, 361–366 (1982).
42. Galea, M. H., Blamey, R. W., Elston, C. E. & Ellis, I. O. The Nottingham prognostic index in primary breast cancer. *Breast Cancer Res. Treat.* **22**, 207–219 (1992).
43. Todd, J. H. *et al.* Confirmation of a prognostic index in primary breast cancer. *Br. J. Cancer* **56**, 489–492 (1987).
44. Blamey, R. W. *et al.* Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. *Eur. J. Cancer* **43**, 1548–1555 (2007).
45. Balslev, I. *et al.* The Nottingham Prognostic Index applied to 9,149 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG). *Breast Cancer Res. Treat.* **32**, 281–290 (1994).
46. D'Eredita', G., Giardina, C., Martellotta, M., Natale, T. & Ferrarese, F. Prognostic factors in breast cancer: The predictive value of the Nottingham Prognostic Index in patients with a long-term follow-up that were treated in a single institution. *Eur. J. Cancer* **37**, 591–596 (2001).
47. Sundquist, M. *et al.* Applying the Nottingham Prognostic Index to a Swedish breast

cancer population. *Breast Cancer Res. Treat.* **53**, 1–8 (1999).

48. Albergaria, A. *et al.* Nottingham Prognostic Index in Triple-Negative Breast Cancer: A reliable prognostic tool? *BMC Cancer* **11**, (2011).
49. Van Belle, V. *et al.* Qualitative assessment of the progesterone receptor and HER2 improves the Nottingham prognostic index up to 5 years after breast cancer diagnosis. *J. Clin. Oncol.* **28**, 4129–4134 (2010).
50. Green, A. R. *et al.* Nottingham prognostic index plus: Validation of a clinical decision making tool in breast cancer in an independent series. *J. Pathol. Clin. Res.* **2**, 32–40 (2016).
51. Lambertini, M. *et al.* The prognostic performance of Adjuvant! Online and Nottingham Prognostic Index in young breast cancer patients. *Br. J. Cancer* **115**, 1471–1478 (2016).
52. Gray, L. J. *et al.* The Leicester Risk Assessment score for detecting undiagnosed Type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. *Diabet. Med.* **27**, 887–895 (2010).
53. Gray, L. J., Khunti, K., Wilmot, E. G., Yates, T. & Davies, M. J. External validation of two diabetes risk scores in a young UK South Asian population. *Diabetes Res. Clin. Pract.* **104**, 451–458 (2014).
54. Barber, S. R., Dhalwani, N. N., Davies, M. J., Khunti, K. & Gray, L. J. External national validation of the Leicester Self-Assessment score for Type 2 diabetes using data from the English Longitudinal Study of Ageing. *Diabet. Med.* **34**, 1575–1583 (2017).
55. Wishart, G. C. *et al.* PREDICT: A new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res.* **12**, (2010).
56. Wishart, G. C. *et al.* A population-based validation of the prognostic model PREDICT for early breast cancer. *Eur. J. Surg. Oncol.* **37**, 411–417 (2011).
57. De Glas, N. A. *et al.* Validity of the online PREDICT tool in older patients with breast cancer: A population-based study. *Br. J. Cancer* **114**, 395–400 (2016).
58. Wishart, G. C. *et al.* Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. *BMC Cancer* **14**, (2014).
59. Wong, H. S. *et al.* The predictive accuracy of PREDICT: A personalized decision-making tool for southeast Asian women with breast cancer. *Med. (United States)* **94**, e593 (2015).
60. Maishman, T. *et al.* An evaluation of the prognostic model PREDICT using the POSH cohort of women aged ≤40 years at breast cancer diagnosis. *Br. J. Cancer* **112**, 983–991 (2015).
61. Engelhardt, E. G. *et al.* Accuracy of the online prognostication tools PREDICT and

Adjuvant! for early-stage breast cancer patients younger than 50 years. *Eur. J. Cancer* **78**, 37–44 (2017).

62. van Maaren, M. C. *et al.* Validation of the online prediction tool PREDICT v. 2.0 in the Dutch breast cancer population. *Eur. J. Cancer* **86**, 364–372 (2017).
63. Plakhins, G. *et al.* Underestimated survival predictions of the prognostic tools Adjuvant! Online and PREDICT in BRCA1-associated breast cancer patients. *Fam. Cancer* **12**, 683–689 (2013).
64. Gray, E., Marti, J., Brewster, D. H., Wyatt, J. C. & Hall, P. S. Independent validation of the PREDICT breast cancer prognosis prediction tool in 45,789 patients using Scottish Cancer Registry data. *Br. J. Cancer* **119**, 808–814 (2018).
65. Aguirre, U. *et al.* External validation of the PREDICT tool in Spanish women with breast cancer participating in population-based screening programmes. *J. Eval. Clin. Pract.* **25**, 873–880 (2019).
66. Kanis, J. A., Johnell, O., Oden, A., Johansson, H. & McCloskey, E. FRAX™ and the assessment of fracture probability in men and women from the UK. *Osteoporos. Int.* **19**, 385–397 (2008).
67. Leslie, W. D. *et al.* Independent clinical validation of a Canadian FRAX tool: Fracture prediction and model calibration. *J. Bone Miner. Res.* **25**, 2350–2358 (2010).
68. Trémollieres, F. A. *et al.* Fracture risk prediction using BMD and clinical risk factors in early postmenopausal women: Sensitivity of the WHO FRAX tool. *J. Bone Miner. Res.* **25**, 1002–1009 (2010).
69. Fraser, L. A. *et al.* Fracture prediction and calibration of a Canadian FRAX® tool: A population-based report from CaMos. *Osteoporos. Int.* **22**, 829–837 (2011).
70. Sandhu, S. K. *et al.* Prognosis of fracture: Evaluation of predictive accuracy of the FRAX™ algorithm and Garvan nomogram. *Osteoporos. Int.* **21**, 863–871 (2010).
71. Lippuner, K., Johansson, H., Kanis, J. A. & Rizzoli, R. FRAX® assessment of osteoporotic fracture probability in Switzerland. *Osteoporos. Int.* **21**, 381–389 (2010).
72. González-Macías, J., Marin, F., Vila, J. & Díez-Pérez, A. Probability of fractures predicted by FRAX® and observed incidence in the Spanish ECOSAP Study cohort. *Bone* **50**, 373–377 (2012).
73. Leslie, W. D. *et al.* High fracture probability with FRAX® usually indicates densitometric osteoporosis: Implications for clinical practice. *Osteoporos. Int.* **23**, 391–397 (2012).
74. Sornay-Rendu, E., Munoz, F., Delmas, P. D. & Chapurlat, R. D. The FRAX tool in French women: How well does it describe the real incidence of fracture in the OFELY cohort. *J. Bone Miner. Res.* **25**, 2101–2107 (2010).
75. Rubin, K. H. *et al.* Comparison of different screening tools (FRAX®, OST, ORAI,

OSIRIS, SCORE and age alone) to identify women with increased risk of fracture. A population-based prospective study. *Bone* **56**, 16–22 (2013).

76. *Emergency Triage*. *Emergency Triage* (John Wiley & Sons Ltd, 2013). doi:10.1002/9781118299029.
77. Zachariasse, J. M. *et al.* Validity of the Manchester triage system in emergency care: A prospective observational study. *PLoS One* **12**, (2017).
78. Seiger, N. *et al.* Improving the manchester triage system for pediatric emergency care: An international multicenter study. *PLoS One* **9**, (2014).
79. Seiger, N., Veen, M. Van, Steyerberg, E. W., Lei, J. Van Der & Moll, H. A. Accuracy of triage for children with chronic illness and infectious symptoms. *Pediatrics* vol. 132 e1602–e1608 (2013).
80. Aeimchanbanjong, K. Validation of different pediatric triage systems in the emergency department. *World J. Emerg. Med.* **8**, 223 (2017).
81. Van Veen, M. *et al.* Safety of the manchester triage system to identify less urgent patients in paediatric emergence care: A prospective observational study. *Arch. Dis. Child.* **96**, 513–518 (2011).
82. Van Ierland, Y., Seiger, N., Van Veen, M., Moll, H. A. & Oostenbrink, R. Alarming signs in the Manchester Triage System: A tool to identify febrile children at risk of hospitalization. *J. Pediatr.* **162**, (2013).
83. Nijman, R. G. *et al.* Can urgency classification of the Manchester triage system predict serious bacterial infections in febrile children? *Arch. Dis. Child.* **96**, 715–722 (2011).
84. Zachariasse, J. M., Kuiper, J. W., de Hoog, M., Moll, H. A. & van Veen, M. Safety of the Manchester Triage System to Detect Critically Ill Children at the Emergency Department. *J. Pediatr.* **177**, 232-237.e1 (2016).
85. Nishi, F. A., Polak, C. & Cruz, D. de A. L. M. da. Sensitivity and specificity of the Manchester Triage System in risk prioritization of patients with acute myocardial infarction who present with chest pain. *Eur. J. Cardiovasc. Nurs.* **17**, 660–666 (2018).
86. Zaboli, A. *et al.* Triage of patients with fever: The Manchester triage system's predictive validity for sepsis or septic shock and seven-day mortality. *J. Crit. Care* **59**, 63–69 (2020).
87. Lim, W. S. *et al.* Defining community acquired pneumonia severity on presentation to hospital: An international derivation and validation study. *Thorax* **58**, 377–382 (2003).
88. Capelastegui, A. *et al.* Validation of a predictive rule for the management of community-acquired pneumonia. *Eur. Respir. J.* **27**, 151–157 (2006).
89. Kruger, S. *et al.* Procalcitonin predicts patients at low risk of death from community-acquired pneumonia across all CRB-65 classes. *Eur. Respir. J.* **31**, 349–355 (2008).

90. Bauer, T. T., Ewig, S., Marre, R., Suttorp, N. & Welte, T. CRB-65 predicts death from community-acquired pneumonia. *J. Intern. Med.* **260**, 93–101 (2006).
91. Ewig, S. *et al.* Validation of predictive rules and indices of severity for community acquired pneumonia. *Thorax* **59**, 421–427 (2004).
92. Menéndez, R. *et al.* Biomarkers improve mortality prediction by prognostic scales in community-acquired pneumonia. *Thorax* **64**, 587–591 (2009).
93. Shin, Y. M. *et al.* Prospective comparison of three predictive rules for assessing severity of community-acquired pneumonia in Hong Kong. *Thorax* **62**, 348–353 (2007).
94. Barlow, G., Nathwani, D. & Davey, P. The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia. *Thorax* **62**, 253–259 (2007).
95. Myint, P. K., Kamath, A. V., Vowler, S. L., Maisey, D. N. & Harrison, B. D. W. Severity assessment criteria recommended by the British Thoracic Society (BTS) for community-acquired pneumonia (CAP) and older patients. Should SOAR (systolic blood pressure, oxygenation, age and respiratory rate) criteria be used in older people? A compilation study of two prospective cohorts. *Age Ageing* **35**, 286–291 (2006).
96. Chalmers, J. D., Singanayagam, A. & Hill, A. T. Systolic blood pressure is superior to other haemodynamic predictors of outcome in community acquired pneumonia. *Thorax* **63**, 698–702 (2008).
97. Bont, J., Hak, E., Hoes, A. W., Macfarlane, J. T. & Verheij, T. J. M. Predicting death in elderly patients with community-acquired pneumonia: A prospective validation study reevaluating the CRB-65 severity assessment tool. *Archives of Internal Medicine* vol. 168 1465–1468 (2008).
98. Blatchford, O., Murray, W. R. & Blatchford, M. A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. *Lancet* **356**, 1318–1321 (2000).
99. Stanley, A. J. *et al.* Outpatient management of patients with low-risk upper-gastrointestinal haemorrhage: multicentre validation and prospective evaluation. *Lancet* **373**, 42–47 (2009).
100. Pang, S. H. *et al.* Comparing the Blatchford and pre-endoscopic Rockall score in predicting the need for endoscopic therapy in patients with upper GI hemorrhage. *Gastrointest. Endosc.* **71**, 1134–1140 (2010).
101. Hyett, B. H. *et al.* The AIMS65 score compared with the Glasgow-Blatchford score in predicting outcomes in upper GI bleeding. *Gastrointest. Endosc.* **77**, 551–557 (2013).
102. Bryant, R. V. *et al.* Performance of the Glasgow-Blatchford score in predicting clinical outcomes and intervention in hospitalized patients with upper GI bleeding.

Gastrointest. Endosc. **78**, 576–583 (2013).

103. Laursen, S. B., Hansen, J. M. & Schaffalitzky de Muckadell, O. B. The Glasgow Blatchford Score Is the Most Accurate Assessment of Patients With Upper Gastrointestinal Hemorrhage. *Clin. Gastroenterol. Hepatol.* **10**, 1130-1135.e1 (2012).
104. Chandra, S. *et al.* External validation of the Glasgow-Blatchford bleeding score and the Rockall score in the US setting. *Am. J. Emerg. Med.* **30**, 673–679 (2012).
105. Srirajaskanthan, R., Conn, R., Bulwer, C. & Irving, P. The Glasgow Blatchford scoring system enables accurate risk stratification of patients with upper gastrointestinal haemorrhage. *Int. J. Clin. Pract.* **64**, 868–874 (2010).
106. Robertson, M. *et al.* Risk stratification in acute upper GI bleeding: Comparison of the AIMS65 score with the Glasgow-Blatchford and Rockall scoring systems. *Gastrointest. Endosc.* **83**, 1151–1160 (2016).
107. Schiefer, M. *et al.* Predictive validity of the Glasgow Blatchford Bleeding Score in an unselected emergency department population in continental Europe. *Eur. J. Gastroenterol. Hepatol.* **24**, 382–387 (2012).
108. Stephens, J. R. *et al.* Management of minor upper gastrointestinal haemorrhage in the community using the Glasgow Blatchford Score. *Eur. J. Gastroenterol. Hepatol.* **21**, 1340–1346 (2009).
109. Apgar, V. A proposal for a new method of evaluation of the newborn infant. in *Anesthesia and Analgesia* vol. 120 1056–1059 (Lippincott Williams and Wilkins, 2015).
110. Casey, B. M., McIntire, D. D. & Leveno, K. J. The Continuing Value of the Apgar Score for the Assessment of Newborn Infants. *N. Engl. J. Med.* **344**, 467–471 (2001).
111. Thorngren-Jerneck, K. & Herbst, A. Low 5-minute Apgar score: A population-based register study of 1 million term births. *Obstet. Gynecol.* **98**, 65–70 (2001).
112. Moster, D., Lie, R. T., Irgens, L. M., Bjerkedal, T. & Markestad, T. The association of Apgar score with subsequent death and cerebral palsy: A population-based study in term infants. *J. Pediatr.* **138**, 798–803 (2001).
113. Iliodromiti, S., MacKay, D. F., Smith, G. C. S., Pell, J. P. & Nelson, S. M. Apgar score and the risk of cause-specific infant mortality: A population-based cohort study. *Lancet* **384**, 1749–1755 (2014).
114. Lie, K. K., Grøholt, E. K. & Eskild, A. Association of cerebral palsy with Apgar score in low and normal birthweight infants: Population based cohort study. *BMJ* **341**, 817 (2010).
115. Laptook, A. R. *et al.* Outcome of term infants using Apgar scores at 10 minutes following hypoxic-ischemic encephalopathy. *Pediatrics* **124**, 1619–1626 (2009).
116. Moster, D., Lie, R. T. & Markestad, T. Joint association of Apgar scores and early

neonatal symptoms with minor disabilities at school age. *Arch. Dis. Child. Fetal Neonatal Ed.* **86**, F16 (2002).

117. Regenbogen, S. E. *et al.* Utility of the surgical apgar score: Validation in 4119 patients. *Arch. Surg.* **144**, 30–36 (2009).
118. Li, F. *et al.* The Apgar Score and Infant Mortality. *PLoS One* **8**, (2013).
119. Sun, Y., Vestergaard, M., Pedersen, C. B., Christensen, J. & Olsen, J. Apgar scores and long-term risk of epilepsy. *Epidemiology* **17**, 296–301 (2006).
120. Rothwell, P. M. *et al.* A simple score (ABCD) to identify individuals at high early risk of stroke after transient ischaemic attack. *Lancet* **366**, 29–36 (2005).
121. Johnston, S. C. *et al.* Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet* **369**, 283–292 (2007).
122. Giles, M. F. *et al.* Early stroke risk and ABCD2 score performance in tissue- Vs time-defined TIA: A multicenter study. *Neurology* **77**, 1222–1228 (2011).
123. Tsivgoulis, G. *et al.* Validation of the ABCD score in identifying individuals at high early risk of stroke after a transient ischemic attack: A hospital-based case series study. *Stroke* **37**, 2892–2897 (2006).
124. Sheehan, O. C. *et al.* Population-based study of ABCD2 score, carotid stenosis, and atrial fibrillation for early stroke prediction after transient ischemic attack: The North Dublin TIA study. *Stroke* **41**, 844–850 (2010).
125. Sciolla, R. & Melis, F. Rapid identification of high-risk transient ischemic attacks: Prospective validation of the ABCD score. *Stroke* **39**, 297–302 (2008).
126. Perry, J. J. *et al.* Prospective validation of the ABCD2 score for patients in the emergency department with transient ischemic attack. *CMAJ* **183**, 1137–1145 (2011).
127. Kiyohara, T. *et al.* ABCD3 and ABCD3-I scores are superior to ABCD2 score in the prediction of short- and long-term risks of stroke after transient ischemic attack. *Stroke* **45**, 418–425 (2014).
128. Koton, S. & Rothwell, P. M. Performance of the ABCD and ABCD2 scores in TIA patients with carotid stenosis and atrial fibrillation. *Cerebrovasc. Dis.* **24**, 231–235 (2007).
129. Tsivgoulis, G. *et al.* Multicenter external validation of the ABCD2 score in triaging TIA patients. *Neurology* **74**, 1351–1357 (2010).
130. Fothergill, A., Christianson, T. J. H., Brown, R. D. & Rabinstein, A. A. Validation and refinement of the ABCD2 score: A population-based analysis. *Stroke* **40**, 2669–2673 (2009).
131. Eagle, K. A. *et al.* A validated prediction model for all forms of acute coronary syndrome estimating the risk of 6-month postdischarge death in an international registry. *J. Am. Med. Assoc.* **291**, 2727–2733 (2004).

132. De Araújo Gonçalves, P., Ferreira, J., Aguiar, C. & Seabra-Gomes, R. TIMI, PURSUIT, and GRACE risk scores: Sustained prognostic value and interaction with revascularization in NSTE-ACS. *Eur. Heart J.* **26**, 865–872 (2005).
133. Tang, E. W., Wong, C. K. & Herbison, P. Global Registry of Acute Coronary Events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome. *Am. Heart J.* **153**, 29–35 (2007).
134. Elbarouni, B. *et al.* Validation of the Global Registry of Acute Coronary Event (GRACE) risk score for in-hospital mortality in patients with acute coronary syndrome in Canada. *Am. Heart J.* **158**, 392–399 (2009).
135. Gale, C. P. *et al.* Evaluation of risk scores for risk stratification of acute coronary syndromes in the myocardial infarction national audit project (MINAP) database. *Heart* **95**, 221–227 (2009).
136. Ramsay, G., Podogrodzka, M., McClure, C. & Fox, K. A. A. Risk prediction in patients presenting with suspected cardiac pain: The GRACE and TIMI risk scores versus clinical evaluation. *QJM* **100**, 11–18 (2007).
137. Lyon, R., Morris, A. C., Caesar, D., Gray, S. & Gray, A. Chest pain presenting to the Emergency Department-to stratify risk with GRACE or TIMI? *Resuscitation* **74**, 90–93 (2007).
138. Bradshaw, P., Ko, D. T., Newman, A. M., Donovan, L. R. & Tu, J. V. Validity of the GRACE (Global Registry of Acute Coronary Events) acute coronary syndrome prediction model for six month post-discharge death in an independent data set. *Heart* **92**, 905–909 (2006).
139. Meune, C. *et al.* The GRACE score's performance in predicting in-hospital and 1-year outcome in the era of high-sensitivity cardiac troponin assays and B-type natriuretic peptide. *Heart* **97**, 1479–1483 (2011).
140. Abu-Assi, E. *et al.* Do GRACE (Global Registry of Acute Coronary events) risk scores still maintain their performance for predicting mortality in the era of contemporary management of acute coronary syndromes? *Am. Heart J.* **160**, (2010).
141. Kozieradzka, A. *et al.* GRACE, TIMI, Zwolle and CADILLAC risk scores - Do they predict 5-year outcomes after ST-elevation myocardial infarction treated invasively? *Int. J. Cardiol.* **148**, 70–75 (2011).
142. APACHE II: A severity of disease classification system : Critical Care Medicine. https://journals.lww.com/ccmjournal/Abstract/1985/10000/APACHE_II__A_severity_of_disease_classification.9.aspx.
143. Larvin, M. & Mcmahon, M. J. APACHE-II SCORE FOR ASSESSMENT AND MONITORING OF ACUTE PANCREATITIS. *Lancet* **334**, 201–205 (1989).
144. Wilson, C., Heath, D. I. & Imrie, C. W. Prediction of outcome in acute pancreatitis: A

comparative study of APACHE II, clinical assessment and multiple factor scoring systems. *Br. J. Surg.* **77**, 1260–1264 (1990).

145. Rowan, K. M. *et al.* Intensive Care Society's APACHE II study in Britain and Ireland - II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. *Br. Med. J.* **307**, 977–981 (1993).
146. Kruse, J. A., Thill Baharozian, M. C. & Carlson, R. W. Comparison of Clinical Assessment With APACHE II for Predicting Mortality Risk in Patients Admitted to a Medical Intensive Care Unit. *JAMA J. Am. Med. Assoc.* **260**, 1739–1742 (1988).
147. Bohnen, J. M. A., Mustard, R. A., Oxholm, S. E. & Schouten, B. D. APACHE II Score and Abdominal Sepsis: A Prospective Study. *Arch. Surg.* **123**, 225–229 (1988).
148. Escarce, J. J. & Kelley, M. A. Admission Source to the Medical Intensive Care Unit Predicts Hospital Death Independent of APACHE II Score. *JAMA J. Am. Med. Assoc.* **264**, 2389–2394 (1990).
149. Wong, D. T., Crofts, S. L., Gomez, M., McGuire, G. P. & Byrick, R. J. Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients. *Crit. Care Med.* **23**, 1177–1183 (1995).
150. Johnson, C. D., Toh, S. K. C. & Campbell, M. J. Combination of APACHE-II score and an obesity score (APACHE-O) for the prediction of severe acute pancreatitis. *Pancreatology* **4**, 1–6 (2004).
151. Cerra, F. B., Negro, F. & Abrams, J. APACHE II Score Does Not Predict Multiple Organ Failure or Mortality in Postoperative Surgical Patients. *Arch. Surg.* **125**, 519–522 (1990).
152. Chatzicostas, C. *et al.* Comparison of Ranson, APACHE II and APACHE III scoring systems in acute pancreatitis. *Pancreas* **25**, 331–335 (2002).
153. Lip, G. Y. H. *et al.* Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: The Euro Heart Survey on atrial fibrillation. *Chest* **137**, 263–272 (2010).
154. Olesen, J. B. *et al.* Validation of risk stratification schemes for predicting stroke and thromboembolism in patients with atrial fibrillation: Nationwide cohort study. *BMJ* **342**, 320 (2011).
155. Piccini, J. P. *et al.* Renal dysfunction as a predictor of stroke and systemic embolism in patients with nonvalvular atrial fibrillation: Validation of the R2CHADS2 index in the ROCKET AF. *Circulation* **127**, 224–232 (2013).
156. Coppens, M. *et al.* The CHA2DS2-VASc score identifies those patients with atrial fibrillation and a CHADS2 score of 1 who are unlikely to benefit from oral anticoagulant therapy. *Eur. Heart J.* **34**, 170–176 (2013).
157. Chao, T. F. *et al.* CHADS 2 and CHA 2DS 2-VASc scores in the prediction of clinical

outcomes in patients with atrial fibrillation after catheter ablation. *J. Am. Coll. Cardiol.* **58**, 2380–2385 (2011).

158. Chao, T. F. *et al.* Using the CHA2DS2-VASc score for refining stroke risk stratification in 'low-risk' Asian patients with Atrial fibrillation. *J. Am. Coll. Cardiol.* **64**, 1658–1665 (2014).
159. Ntaios, G. *et al.* CHADS2, CHA2DS2-VASc, and long-term stroke outcome in patients without atrial fibrillation. *Neurology* **80**, 1009–1017 (2013).
160. Mitchell, L. B. *et al.* Prediction of stroke or TIA in patients without atrial fibrillation using CHADS2 and CHA2DS2-VASc scores. *Heart* **100**, 1524–1530 (2014).
161. Van Den Ham, H. A., Klungel, O. H., Singer, D. E., Leufkens, H. G. M. & Van Staa, T. P. Comparative Performance of ATRIA, CHADS2, and CHA2DS2-VASc Risk Scores Predicting Stroke in Patients With Atrial Fibrillation: Results From a National Primary Care Database. *J. Am. Coll. Cardiol.* **66**, 1851–1859 (2015).
162. Welles, C. C. *et al.* The CHADS2 score predicts ischemic stroke in the absence of atrial fibrillation among subjects with coronary heart disease: Data from the Heart and Soul Study. *Am. Heart J.* **162**, 555–561 (2011).
163. Poçi, D. *et al.* Role of the CHADS2 score in acute coronary syndromes: Risk of subsequent death or stroke in patients with and without atrial fibrillation. *Chest* **141**, 1431–1440 (2012).
164. JACOBS, I. *et al.* A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. *BJOG An Int. J. Obstet. Gynaecol.* **97**, 922–929 (1990).
165. Tingulstad, S. *et al.* Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. *BJOG An Int. J. Obstet. Gynaecol.* **103**, 826–831 (1996).
166. Moore, R. G. *et al.* Comparison of a novel multiple marker assay vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. *Am. J. Obstet. Gynecol.* **203**, 228.e1-228.e6 (2010).
167. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals - PubMed. https://pubmed.ncbi.nlm.nih.gov/10074998/.
168. Davies, A. P., Jacobs, I., Woolas, R., Fish, A. & Oram, D. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. *BJOG An Int. J. Obstet. Gynaecol.* **100**, 927–931 (1993).
169. Anton, C. *et al.* A comparison of CA125, HE4, risk ovarian malignancy algorithm (ROMA), and risk malignancy index (RMI) for the classification of ovarian masses. *Clinics* **67**, 437–441 (2012).
170. Manjunath, A. P., Pratapkumar, Sujatha, K. & Vani, R. Comparison of three risk of

malignancy indices in evaluation of pelvic masses. *Gynecol. Oncol.* **81**, 225–229 (2001).

171. Andersen, E. S., Knudsen, A., Rix, P. & Johansen, B. Risk of Malignancy Index in the preoperative evaluation of patients with adnexal masses. *Gynecol. Oncol.* **90**, 109–112 (2003).
172. Van Gorp, T. *et al.* Subjective assessment by ultrasound is superior to the risk of malignancy index (RMI) or the risk of ovarian malignancy algorithm (ROMA) in discriminating benign from malignant adnexal masses. *Eur. J. Cancer* **48**, 1649–1656 (2012).
173. van den Akker, P. A. J. *et al.* Evaluation of the risk of malignancy index in daily clinical management of adnexal masses. *Gynecol. Oncol.* **116**, 384–388 (2010).
174. Obeidat, B. R., Amarin, Z. O., Latimer, J. A. & Crawford, R. A. Risk of malignancy index in the preoperative evaluation of pelvic masses. *Int. J. Gynecol. Obstet.* **85**, 255–258 (2004).
175. Morrow, D. A. *et al.* TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation: An Intravenous nPA for Treatment of Infarcting Myocardium Early II trial substudy. *Circulation* **102**, 2031–2037 (2000).
176. Morrow, D. A. *et al.* Application of the TIMI risk score for ST-elevation MI in the National Registry of Myocardial Infarction 3. *J. Am. Med. Assoc.* **286**, 1356–1359 (2001).
177. Aragam, K. G. *et al.* Does simplicity compromise accuracy in ACS risk prediction? A retrospective analysis of the TIMI and GRACE risk scores. *PLoS One* **4**, (2009).
178. Scirica, B. M. *et al.* Validation of the Thrombolysis In Myocardial Infarction (TIMI) risk score for unstable angina pectoris and non-ST-elevation myocardial infarction in the TIMI III registry. *Am. J. Cardiol.* **90**, 303–305 (2002).
179. Hess, E. P. *et al.* Prospective validation of a modified thrombolysis in myocardial infarction risk score in emergency department patients with chest pain and possible acute coronary syndrome. *Acad. Emerg. Med.* **17**, 368–375 (2010).
180. Correia, L. C. L. *et al.* Prognostic value of TIMI score versus grace score in st-segment elevation myocardial infarction. *Arq. Bras. Cardiol.* **103**, 98–106 (2014).
181. TIMI Risk Score accurately predicts risk of death in 30-day and one-year follow-up in STEMI patients treated with primary percutaneous coronary interventions - PubMed. https://pubmed.ncbi.nlm.nih.gov/17694460/.
182. Selvarajah, S. *et al.* An Asian validation of the TIMI risk score for ST-segment elevation myocardial infarction. *PLoS One* **7**, (2012).
183. Macdonald, S. P. J., Nagree, Y., Fatovich, D. M. & Brown, S. G. A. Modified TIMI risk

score cannot be used to identify low-risk chest pain in the emergency department: A multicentre validation study. *Emerg. Med. J.* **31**, 281–285 (2014).

184. Williams, B. A. *et al.* External validation of the TIMI risk score for secondary cardiovascular events among patients with recent myocardial infarction. *Atherosclerosis* **272**, 80–86 (2018).
185. Pisters, R. *et al.* A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: The euro heart survey. *Chest* **138**, 1093–1100 (2010).
186. Lip, G. Y. H., Frison, L., Halperin, J. L. & Lane, D. A. Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: The HAS-BLED (hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile INR, elderly, drugs/alcohol concomitantly) score. *J. Am. Coll. Cardiol.* **57**, 173–180 (2011).
187. Apostolakis, S., Lane, D. A., Guo, Y., Buller, H. & Lip, G. Y. H. Performance of the HEMORR 2HAGES, ATRIA, and HAS-BLED bleeding risk-prediction scores in patients with atrial fibrillation undergoing anticoagulation: The AMADEUS (Evaluating the use of SR34006 compared to warfarin or acenocoumarol in patients with atrial fibrillation) study. *J. Am. Coll. Cardiol.* **60**, 861–867 (2012).
188. Roldán, V. *et al.* Predictive value of the HAS-BLED and ATRIA bleeding scores for the risk of serious bleeding in a 'real-world' population with atrial fibrillation receiving anticoagulant therapy. *Chest* **143**, 179–184 (2013).
189. Roldán, V. *et al.* The HAS-BLED score has better prediction accuracy for major bleeding than CHADS2 or CHA2DS2-VASc scores in anticoagulated patients with atrial fibrillation. *J. Am. Coll. Cardiol.* **62**, 2199–2204 (2013).
190. Omran, H., Bauersachs, R., Rübenacker, S., Goss, F. & Hammerstingl, C. The HAS-BLED score predicts bleedings during bridging of chronic oral anticoagulation: Results from the national multicentre BNK online bridging registry (BORDER). *Thromb. Haemost.* **108**, 65–73 (2012).
191. Gallego, P. *et al.* Relation of the HAS-BLED bleeding risk score to major bleeding, cardiovascular events, and mortality in anticoagulated patients with atrial fibrillation. *Circ. Arrhythmia Electrophysiol.* **5**, 312–318 (2012).
192. Senoo, K., Proietti, M., Lane, D. A. & Lip, G. Y. H. Evaluation of the HAS-BLED, ATRIA, and ORBIT Bleeding Risk Scores in Patients with Atrial Fibrillation Taking Warfarin. *Am. J. Med.* **129**, 600–607 (2016).
193. Apostolakis, S., Lane, D. A., Buller, H. & Lip, G. Y. H. Comparison of the CHADS2, CHA2DS2 -VASc and HAS-BLED scores for the prediction of clinically relevant bleeding in anticoagulated patients with atrial fibrillation: The AMADEUS trial. *Thromb.*

Haemost. **110**, 1074–1079 (2013).

194. Okumura, K. *et al.* Validation of CHA2DS2-VASc and HAS-BLED scores in Japanese patients with nonvalvular atrial fibrillation - An analysis of the J-RHYTHM registry. *Circ. J.* **78**, 1593–1599 (2014).
195. Kooiman, J. *et al.* The HAS-BLED score identifies patients with acute venous thromboembolism at high risk of major bleeding complications during the first six months of anticoagulant treatment. *PLoS One* **10**, (2015).
196. Antoniou, A. C. *et al.* A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br. J. Cancer* **86**, 76–83 (2002).
197. Fischer, C. *et al.* Evaluating the performance of the breast cancer genetic risk models BOADICEA, IBIS, BRCAPRO and Claus for predicting BRCA1/2 mutation carrier probabilities: A study based on 7352 families from the German hereditary breast and ovarian cancer consortium. *J. Med. Genet.* **50**, 360–367 (2013).
198. Panchal, S. M., Ennis, M., Canon, S. & Bordeleau, L. J. Selecting a BRCA risk assessment model for use in a familial cancer clinic. *BMC Med. Genet.* **9**, (2008).
199. Arver, B. *et al.* Bilateral prophylactic mastectomy in swedish women at high risk of breast cancer: A national survey. *Annals of Surgery* vol. 253 1147–1154 (2011).
200. Muranen, T. A. *et al.* Polygenic risk score is associated with increased disease risk in 52 Finnish breast cancer families. *Breast Cancer Res. Treat.* **158**, 463–469 (2016).
201. Kwong, A. *et al.* Accuracy of BRCA1/2 mutation prediction models for different ethnicities and genders: Experience in a southern Chinese cohort. *World J. Surg.* **36**, 702–713 (2012).
202. Moghadasi, S. *et al.* Performance of BRCA1/2 mutation prediction models in male breast cancer patients. *Clin. Genet.* **93**, 52–59 (2018).
203. Gleason, D. F., Mellinger, G. T. & Ardving, L. J. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J. Urol.* **111**, 58–64 (1974).
204. Lattouf, J. B. & Saad, F. Gleason score on biopsy: Is it reliable for predicting the final grade on pathology? *BJU Int.* **90**, 694–698 (2002).
205. Djavan, B., Kadesky, K., Klopukh, B., Marberger, M. & Roehrborn, C. G. Gleason scores from prostate biopsies obtained with 18-gauge biopsy needles poorly predict gleason scores of radical prostatectomy specimens. *Eur. Urol.* **33**, 261–270 (1998).
206. Green, G. A., Hanlon, A. L., Al-Saleem, T. & Hanks, G. E. A gleason score of 7 predicts a worse outcome for prostate carcinoma patients treated with radiotherapy. *Cancer* **83**, 971–976 (1998).
207. Benaim, E. A., Pace, C. M. & Roehrborn, C. G. Gleason score predicts androgen independent progression after androgen deprivation therapy. *Eur. Urol.* **42**, 12–17

(2002).

208. Lucca, I. *et al.* Validation of tertiary Gleason pattern 5 in Gleason score 7 prostate cancer as an independent predictor of biochemical recurrence and development of a prognostic model. *Urol. Oncol. Semin. Orig. Investig.* **33**, 71.e21-71.e26 (2015).
209. He, J. *et al.* Validation of a Contemporary Five-tiered Gleason Grade Grouping Using Population-based Data. *Eur. Urol.* **71**, 760–763 (2017).
210. Pompe, R. S. *et al.* Population-Based Validation of the 2014 ISUP Gleason Grade Groups in Patients Treated With Radical Prostatectomy, Brachytherapy, External Beam Radiation, or no Local Treatment. *Prostate* **77**, 686–693 (2017).
211. Djaladat, H. *et al.* Oncological Outcomes After Radical Prostatectomy for High-Risk Prostate Cancer Based on New Gleason Grouping System: A Validation Study From University of Southern California With 3,755 Cases. *Prostate* **77**, 743–748 (2017).
212. Faraj, S. F. *et al.* Clinical validation of the 2005 isup gleason grading system in a cohort of intermediate and high risk men undergoing radical prostatectomy. *PLoS One* **11**, (2016).
213. Rao, V. *et al.* Validation of the WHO 2016 new Gleason score of prostatic carcinoma. *Urol. Ann.* **10**, 324–329 (2018).
214. The Braden Scale for Predicting Pressure Sore Risk - PubMed. https://pubmed.ncbi.nlm.nih.gov/3299278/.
215. Bergstrom, N., Braden, B., Kemp, M., Champagne, M. & Ruby, E. Predicting Pressure Ulcer Risk: A Multisite Study of the Predictive Validity of the Braden Scale. *Nurs. Res.* **47**, 261–269 (1998).
216. Braden, B. J. & Bergstrom, N. Predictive validity of the braden scale for pressure sore risk in a nursing home population. *Res. Nurs. Health* **17**, 459–470 (1994).
217. Halfens, R. J. G., Van Achterberg, T. & Bal, R. M. Validity and reliability of the Braden scale and the influence of other risk factors: A multi-centre prospective study. *Int. J. Nurs. Stud.* **37**, 313–319 (2000).
218. Jun Seongsook, R. N., Jeong Ihnsook, R. N. & Lee Younghee, R. N. Validity of pressure ulcer risk assessment scales; Cubbin and Jackson, Braden, and Douglas scale. *Int. J. Nurs. Stud.* **41**, 199–204 (2004).
219. Kwong, E. *et al.* Predicting pressure ulcer risk with the modified Braden, Braden, and Norton scales in acute care hospitals in Mainland China. *Appl. Nurs. Res.* **18**, 122–128 (2005).
220. Bergstrom, N. & Braden, B. J. Predictive validity of the Braden Scale among black and white subjects. *Nurs. Res.* **51**, 398–403 (2002).
221. Lyder, C. H. *et al.* The braden scale for pressure ulcer risk: Evaluating the predictive validity in black and latino/hispanic elders. *Appl. Nurs. Res.* **12**, 60–68 (1999).

222. Hyun, S. *et al.* Predictive validity of the braden scale for patients in intensive care units. *Am. J. Crit. Care* **22**, 514–520 (2013).
223. VandenBosch, T., Montoye, C., Satwicz, M., Durkee-Leonard, K. & Boylan-Lewis, B. Predictive validity of the braden scale and nurse perception in identifying pressure ulcer risk. *Appl. Nurs. Res.* **9**, 80–86 (1996).
224. Chan, W. S., Pang, S. M. C. & Kwong, E. W. Y. Assessing predictive validity of the modified Braden scale for prediction of pressure ulcer risk of orthopaedic patients in an acute care setting. *J. Clin. Nurs.* **18**, 1565–1573 (2009).
225. Nashef, S. A. M. *et al.* European system for cardiac operative risk evaluation (EuroSCORE). *Eur. J. Cardio-Thoracic Surg.* **16**, 9–13 (1999).
226. Nashef, S. A. M. *et al.* Validation of European System for Cardiac Operative Risk Evaluation (EuroSCORE) in North American cardiac surgery. in *European Journal of Cardio-thoracic Surgery* vol. 22 101–105 (Eur J Cardiothorac Surg, 2002).
227. Osswald, B. R. *et al.* Overestimation of aortic valve replacement risk by EuroSCORE: Implications for percutaneous valve replacement. *Eur. Heart J.* **30**, 74–80 (2009).
228. Wendt, D. *et al.* Society of Thoracic Surgeons Score Is Superior to the EuroSCORE Determining Mortality in High Risk Patients Undergoing Isolated Aortic Valve Replacement. *Ann. Thorac. Surg.* **88**, 468–475 (2009).
229. Yap, C. H. *et al.* Validation of the EuroSCORE model in Australia. in *European Journal of Cardio-thoracic Surgery* vol. 29 441–446 (Eur J Cardiothorac Surg, 2006).
230. Roques, F. *et al.* Does EuroSCORE work in individual European countries? *Eur. J. Cardio-thoracic Surg.* **18**, 27–30 (2000).
231. Barili, F. *et al.* Does EuroSCORE II perform better than its original versions? A multicentre validation study. *Eur. Heart J.* **34**, 22–29 (2013).
232. Nilsson, J., Algotsson, L., Höglund, P., Lührs, C. & Brandt, J. EuroSCORE predicts intensive care unit stay and costs of open heart surgery. *Ann. Thorac. Surg.* **78**, 1528–1534 (2004).
233. Chalmers, J. *et al.* Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. *Eur. J. Cardio-thoracic Surg.* **43**, 688–694 (2013).
234. Nilsson, J., Algotsson, L., Höglund, P., Lührs, C. & Brandt, J. Early mortality in coronary bypass surgery: The EuroSCORE versus the Society of Thoracic Surgeons risk algorithm. *Ann. Thorac. Surg.* **77**, 1235–1239 (2004).
235. Biancari, F. *et al.* Validation of EuroSCORE II in patients undergoing coronary artery bypass surgery. *Ann. Thorac. Surg.* **93**, 1930–1935 (2012).
236. Wilson, P. W. F. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–1847 (1998).
237. D'Agostino, R. B., Grundy, S., Sullivan, L. M. & Wilson, P. Validation of the

Framingham coronary heart disease prediction scores: Results of a multiple ethnic groups investigation. *J. Am. Med. Assoc.* **286**, 180–187 (2001).

238. Greenland, P., LaBree, L., Azen, S. P., Doherty, T. M. & Detrano, R. C. Coronary Artery Calcium Score Combined with Framingham Score for Risk Prediction in Asymptomatic Individuals. *J. Am. Med. Assoc.* **291**, 210–215 (2004).
239. Pencina, M. J., D'Agostino, R. B., Larson, M. G., Massaro, J. M. & Vasan, R. S. Predicting the 30-year risk of cardiovascular disease: The framingham heart study. *Circulation* **119**, 3078–3084 (2009).
240. Liu, J. *et al.* Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-provincial Cohort Study. *J. Am. Med. Assoc.* **291**, 2591–2599 (2004).
241. Hense, H. W., Schulte, H., Löwel, H., Assmann, G. & Keil, U. Framingham risk function overestimates risk of coronary heart disease in men and women from Germany - Results from the MONICA Augsburg and the PROCAM cohorts. *Eur. Heart J.* **24**, 937–945 (2003).
242. Lloyd-Jones, D. M. *et al.* Framingham risk score and prediction of lifetime risk for coronary heart disease. *Am. J. Cardiol.* **94**, 20–24 (2004).
243. Marrugat, J. *et al.* An adaptation of the Framingham coronary heart disease risk function to European Mediterranean areas. *J. Epidemiol. Community Health* **57**, 634–638 (2003).
244. Albert, M. A., Glynn, R. J. & Ridker, P. M. Plasma concentration of C-reactive protein and the calculated Framingham Coronary Heart Disease Risk Score. *Circulation* **108**, 161–165 (2003).
245. Empana, J. P. *et al.* Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. *Eur. Heart J.* **24**, 1903–1911 (2003).
246. Marrugat, J. *et al.* Validity of an adaptation of the Framingham cardiovascular risk function: The VERIFICA study. *J. Epidemiol. Community Health* **61**, 40–47 (2007).
247. National Early Warning Score (NEWS) 2 | RCP London. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2.
248. Fernando, S. M. *et al.* Prognostic accuracy of the Hamilton Early Warning Score (HEWS) and the National Early Warning Score 2 (NEWS2) among hospitalized patients assessed by a rapid response team. *Crit. Care* **23**, 60 (2019).
249. Mellhammar *et al.* NEWS2 is Superior to qSOFA in Detecting Sepsis with Organ Dysfunction in the Emergency Department. *J. Clin. Med.* **8**, 1128 (2019).
250. Beane, A. *et al.* Evaluation of the feasibility and performance of early warning scores to identify patients at risk of adverse outcomes in a low-middle income country setting.

BMJ Open **8**, (2018).

251. Carr, E. *et al.* Evaluation and Improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. *medRxiv* 2020.04.24.20078006 (2020) doi:10.1101/2020.04.24.20078006.
252. Frost, F., Bradley, P., Tharmaratnam, K. & Wootton, D. G. The utility of established prognostic scores in COVID-19 hospital admissions: a multicentre prospective evaluation of CURB-65, NEWS2, and qSOFA. *medRxiv* 2020.07.15.20154815 (2020) doi:10.1101/2020.07.15.20154815.
253. Martín-Rodríguez, F. *et al.* The Value of Prehospital Early Warning Scores to Predict in - Hospital Clinical Deterioration: A Multicenter, Observational Base-Ambulance Study. *Prehospital Emerg. Care* (2020) doi:10.1080/10903127.2020.1813224.
254. QRISK3. https://qrisk.org/three/.
255. ABCD^2 Score for TIA - MDCalc. https://www.mdcalc.com/abcd2-score-tia.
256. TIMI Risk Score for UA/NSTEMI - MDCalc. https://www.mdcalc.com/timi-risk-score-ua-nstemi.
257. CURB-65 Score for Pneumonia Severity - MDCalc. https://www.mdcalc.com/curb-65-score-pneumonia-severity.
258. New EuroSCORE II (2011). http://www.euroscore.org/calc.html.
259. (No Title). https://www.sheffield.ac.uk/FRAX/tool.aspx?country=1.
260. CHA_2DS_2-VASc Score for Atrial Fibrillation Stroke Risk - MDCalc. https://www.mdcalc.com/cha2ds2-vasc-score-atrial-fibrillation-stroke-risk.
261. Risk of Malignancy Index (RMI) for Ovarian Cancer - MDCalc. https://www.mdcalc.com/risk-malignancy-index-rmi-ovarian-cancer.
262. National Early Warning Score (NEWS) 2 - MDCalc. https://www.mdcalc.com/national-early-warning-score-news-2.
263. The Thoracic Surgery Scoring System | pmidCALC online calculators. http://www.pmidcalc.org/?sid=17258556&newtest=Y.
264. Nottingham Prognostic Index | pmidCALC online calculators. http://www.pmidcalc.org/?sid=3689666&newtest=Y.
265. Diabetes UK – Know Your Risk of Type 2 diabetes. https://riskscore.diabetes.org.uk/start.
266. Predict Breast. https://breast.predict.nhs.uk/tool.
267. Glasgow-Blatchford Bleeding Score (GBS) - MDCalc. https://www.mdcalc.com/glasgow-blatchford-bleeding-score-gbs.
268. HAS-BLED Score for Major Bleeding Risk - MDCalc. https://www.mdcalc.com/has-bled-score-major-bleeding-risk.
269. GRACE ACS Risk and Mortality Calculator - MDCalc. https://www.mdcalc.com/grace-

acs-risk-mortality-calculator.

270. Framingham Risk Score for Hard Coronary Heart Disease - MDCalc. https://www.mdcalc.com/framingham-risk-score-hard-coronary-heart-disease.
271. Gleason Score for Prostate Cancer - MDCalc. https://www.mdcalc.com/gleason-score-prostate-cancer.
272. (No Title). https://www.in.gov/isdh/files/Braden_Scale.pdf.
273. APACHE II Score - MDCalc. https://www.mdcalc.com/apache-ii-score.
274. APGAR Score - MDCalc. https://www.mdcalc.com/apgar-score.
275. BOADICEA risk estimation on the World Wide Web. https://pluto.srl.cam.ac.uk/cgi-bin/bd3/v3/userReg2.web2.cgi.
276. Sperrin, M., Martin, G. P., Sisk, R. & Peek, N. Missing data should be handled differently for prediction than for description or causal explanation. *Journal of Clinical Epidemiology* vol. 125 183–187 (2020).
277. Heus, P. *et al.* Poor reporting of multivariable prediction model studies: Towards a targeted implementation strategy of the TRIPOD statement. *BMC Med.* **16**, 120 (2018).
278. Swj, N., Tkj, G. & Jjl, J. Journal Pre-proof Real-time imputation of missing predictor values improved the application of prediction models in daily practice. *J. Clin. Epidemiol.* (2021) doi:10.1016/j.jclinepi.2021.01.003.

TABLE TITLES

Table 1. List of CPMs included in this study.

Table 2. Identified missing data handling methods.

FIGURE TITLES AND DESCRIPTION

Figure 1. CPMs Eligibility criteria.

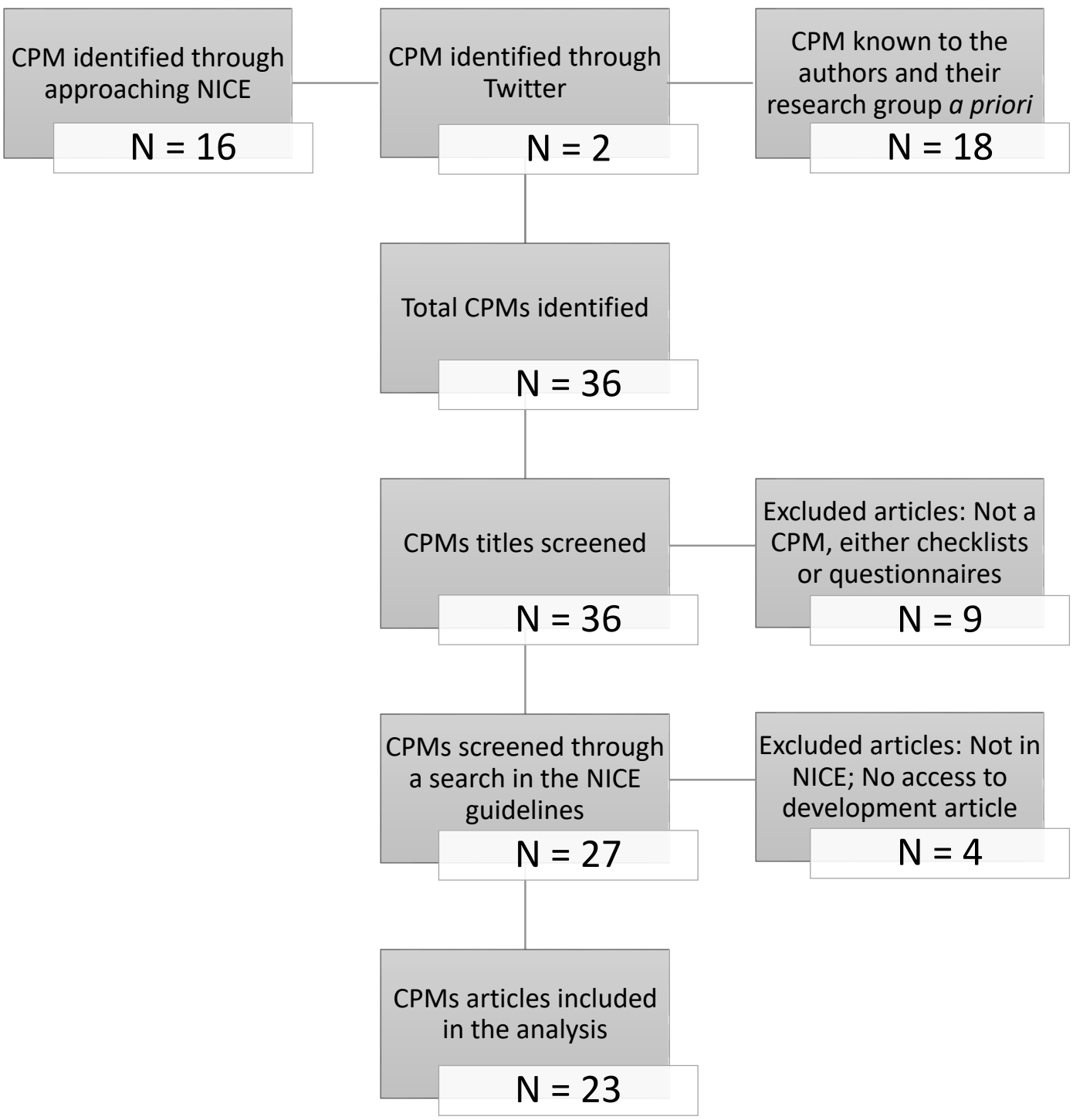
Figure 2. Missing data handling across the pipeline of a CPM. A Sankey diagram, showing different 'paths' of handling missing data across the three stages of a CPM's pipeline. The X axis shows the stages of a CPM's pipeline, whilst the Y axis shows the number possible combinations based on the number of validation papers (0 to 210)

SUPPLEMENTARY MATERIAL CAPTION

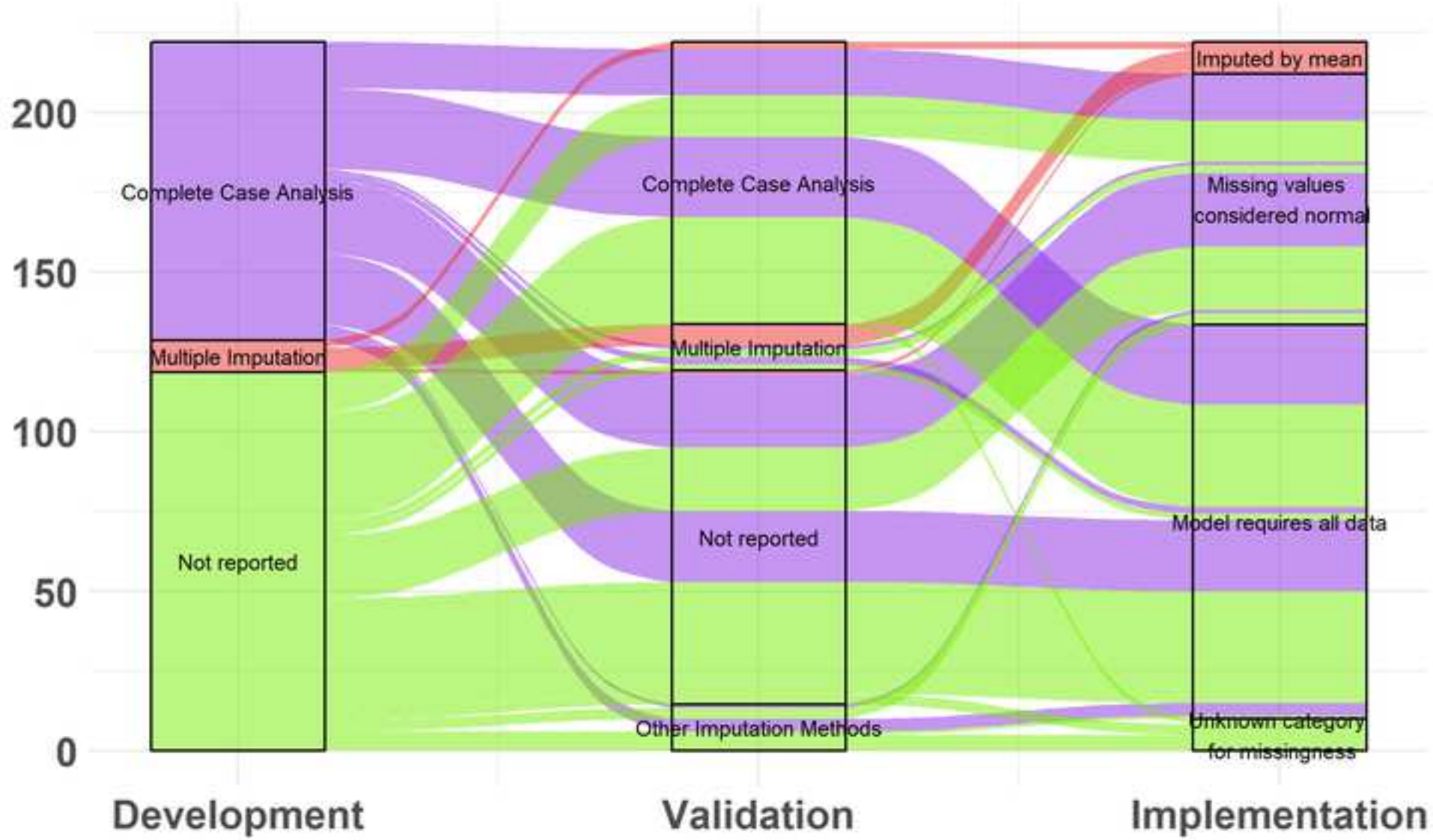
Full list of combinations of missing data handling methods across the pipelines of the CPMs included in this study showing consistent/inconsistent 'paths'.

Figure

Click here to access/download;Figure;Fig.1_Antonia Tsvetanova.pdf



Figure



Table

Click here to access/download;Table;Table 1_Antonia Tsvetanova.docx

CPM	Description	Validation articles
QRISK[18–20]	10-year risk of developing CVD	21–29
Thoracoscore[30]	NSCLC pre-operative risk of death	31–39
Nottingham Prognostic index[40]	Risk of recurrence and overall survival in breast cancer	41–50
The Leicester practice risk score[51]	Screening for undiagnosed T2DM	52,53
PREDICT[54]	Breast and prostate cancers	55–64
FRAX[65]	10-year risk of developing osteoporotic & hip fracture	66–74
Manchester Triage System[75]	Assign clinical priority to patients	76–85
CRB65[86]	Assessment of community acquired pneumonia	87–96
Blatchford[97]	Upper Gastrointestinal bleeding	98–107
APGAR[108]	Evaluate the prognosis of a newborn baby	109–118
ABCD2[119]	Stroke/Transient ischaemic attack	120–129
GRACE[130]	Adverse CVD outcomes	131–140
APACHE[141]	ICU scoring systems for predicting mortality	142–151
CHADVASC[152]	Atrial fibrillation stroke risk	153–162
DG-ROMA[163]	Risk of ovarian malignancy	164–173
TIMI[174]	Thrombolysis in myocardial infarction	175–183,
HAS-BLED[184]	Major Bleeding risk	185–194
BOADICEA[195]	Breast cancer risk prediction model	196–201
Gleason score[202]	Prostate cancer	203–212
Braden Scale[213]	Predicting pressure ulcer risk	214–223
EuroScore[224]	Short-term mortality after cardiac surgery	225–234

Framingham[235]	Risk of CVD over 10 years	236–245
NEWS2[246]	Identifying acutely ill patients	247–252

Table

Method	***Development***		***Validation***		***Implementation***	
	Pros	*Cons*	*Pros*	*Cons*	*Pros*	*Cons*
Complete Case Analysis	Simple	Loss of information	Simple	Selection bias	N/A Equivalent of CCA for Implementation: "Model requires all data"	N/A Equivalent of CCA for Implementation: "Model requires all data"
Mean Imputation	Short computation time	Only works for the average individual	Short computation time	Only works for the average individual	Computationally achievable	Only works for the average individual
Multiple Imputation	Original data/Conditional distribution	High computational cost; Large bias/trade-off for MNAR	Resembling a 'real-world' situation	High computational cost; Large bias/trade-off for MNAR	Original data/Conditional distribution	Cannot be applied to an individual patient; Outcome required
KNN imputation	Can be more accurate than mean/median imputation	High computational cost; Sensitive to outliers	Can be more accurate than mean/median imputation	High computational cost; Sensitive to outliers	N/A	N/A
Additional Category for Missingness	Simple	Known to be biased, even in MCAR	Simple	Unstable to changes in missingness mechanism	No loss of information	Unstable to changes in missingness mechanism
Missing values considered as Normal	None	Simple	Will be biased even in MCAR	simple Biased	Simple	Bias

Model requires all data	All information needed	N/A Only applicable to implementation stage	N/A Only applicable to implementation stage	N/A Only applicable to implementation stage	N/A Only applicable to implementation stage	No loss of information	Cannot be applied to individuals with missing values

CPM review Antonia Tsvetanova

Conflicts of interest: none to declare