Journal of Clinical Epidemiology 140 (2021) 149–158

Journal of Clinical Epidemiology

REVIEW

Missing data was handled inconsistently in UK prediction models: a review of method used

Antonia Tsvetanova[a,*], Matthew Sperrin[a], Niels Peek[a,b], Iain Buchan[a,d], Stephanie Hyland[c], Glen P. Martin[a]

[a] *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*

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

[c] *Microsoft Research Cambridge, Cambridge, UK*

[d] *Institute of Population Health, The University of Liverpool, Liverpool, UK*

Accepted 7 September 2021; Available online 11 September 2021

Abstract

Objectives: 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 descriptive cross-sectional meta-epidemiological study aiming 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: A total of 23 CPMs were included through "sampling strategy." 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. Three 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; Missing data handling approaches

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 [3,4]. A simple, but inefficient (and potentially bi-

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

* Corresponding author. Antonia Tsvetanova. Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Vaughan House, Manchester M13 9GB, UK. Tel.: +44 (0) 161 27 50179.

E-mail address: antonia.tsvetanova@manchester.ac.uk (A. Tsvetanova).

https://doi.org/10.1016/j.jclinepi.2021.09.008

What is new?

Key findings

- The optimal methods for handling missing data in the development stage of clinical prediction models are well-understood, however there is less clarity on methods that should be used in external validation and implementation stages.
- Missing data has been generally handled inconsistently across the pipeline of clinical prediction models (CPMs) used in the UK healthcare.
- Missing data and their handling has been poorly reported and accounted for.

What this study 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 were found.

What is the implication and what should change now?

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

ased), 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. In turn, this lack of guidance has potentially resulted in inconsistency between imputation methods used across the pipeline 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]. 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.

At the moment, it is unclear what the consequences of using different missing data handling techniques at different stages of the prediction model, are. Therefore, in this article, we aimed to: (1) review CPMs recommended for use in UK healthcare and to (2) summarize 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 performed a descriptive cross-sectional meta-epidemiological study, aiming to identify and describe existing clinical prediction models used in UK, with respect to missing data handling across their CPM pipeline.

2.1. Sampling strategy

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. 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 by descriptive statistics.

2.2. 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 [16]. 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. The search terms for validation papers included [name of the CPM] – both abbreviated and extended. To maintain a viable number of papers to screen, we selected the top 10 most cited (according to Google scholar as-of September 18, 2020) validation papers for each CPM. 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 (eg, 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 September 22, 2020.

2.3. Data extraction and synthesis

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

(i) General information such as: author, year of publication and title.
(ii) Source of data used (eg, cohort, case-control, registry), sample size and the outcome of interest.
(iii) The method (eg, logistic, survival), and software used for analysis.
(iv) Missing data handling approach (eg, CCA, imputation methods, other or none reported).
(v) 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.

2.4. The TRIPOD statement

We applied The Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement to all the articles (items 9, 13a and 13b [17]), to check how well missing data and their handling have been reported.

2.5. Definition of consistency across CPM pipeline

We defined "consistency" through discussions between the authors, 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 (eg, in our case, only permitting missing at random (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 Missing completely at random (MCAR) pattern, 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 missing at random pattern.

3. Results

We identified 23 clinical prediction models that met the eligibility criteria (Fig. 1). (See online Appendix A for the list of all initially identified CPMs). Descriptions of each CPM are summarized in Table 1.

In total, information from 233 articles was extracted. Development articles were available for 23 out of 24 CPMs. There was one article (Waterlow Score [18]), 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 [19–21], Thoracoscore [22], The Leicester practice risk score [23], FRAX [24], BOADICEA [25] and NEWS2 [26]).

3.1. Missing Data

Six missing data approaches were identified through the literature search within the development and validation papers, summarized 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 10 (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, missing data handling approach was not reported in 100 studies (48%).

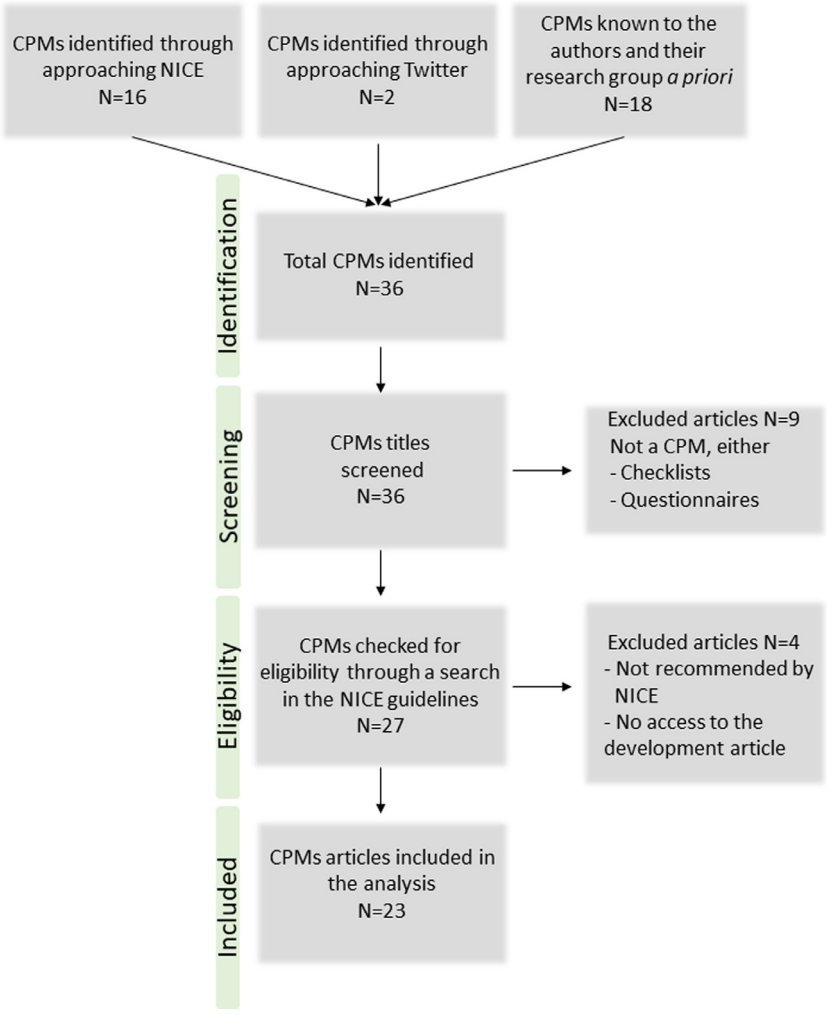

Fig. 1. CPMs sampling strategy.

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 imputation, additional category for missing values (missing indicator method) or missing values considered as normal (eg, 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 [21]) 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. Eight CPMs (35%) make the assumption for missing values to be the lowest/normal – ABCD2 [27], TIMI [28], CRB65 [29], EuroScore [30], FRAX [24], CHADVASC [31], DG-ROMA [32], NEWS2 [26]. Overall, 13 CPMs (57%) require that all data is present when making a prediction at implementation stage; these models were Thoracoscore [22], NPI [33], Leicester Risk Score [23], PREDICT [34], Blatchford [35], HAS-BLED [36], GRACE [37], Framingham [38], Gleason [39], Braden Scale [40], APACHE [41], APGAR [42], MTS [43]. The remaining CPM – BOADICEA [25], (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" were observed for

Table 1. List of CPMs included in this study

CPM	Description
QRISK [19–21]	10-yr risk of developing CVD
Thoracoscore [22]	NSCLC pre-operative risk of death
Nottingham Prognostic index [33]	Risk of recurrence and overall survival in breast cancer
The Leicester practice risk score [23]	Screening for undiagnosed T2DM
PREDICT [34]	Breast and prostate cancers
FRAX [24]	10-yr risk of developing osteoporotic & hip fracture
Manchester Triage System [43]	Assign clinical priority to patients
CRB65 [29]	Assessment of community acquired pneumonia
Blatchford [35]	Upper Gastrointestinal bleeding
APGAR [42]	Evaluate the prognosis of a newborn baby
ABCD2 [27]	Stroke/Transient ischaemic attack
GRACE [37]	Adverse CVD outcomes
APACHE [41]	ICU scoring systems for predicting mortality
CHADVASC [31]	Atrial fibrillation stroke risk
DG-ROMA [32]	Risk of ovarian malignancy
TIMI [28]	Thrombolysis in myocardial infarction
HAS-BLED [36]	Major Bleeding risk
BOADICEA [25]	Breast cancer risk prediction model
Gleason score [39]	Prostate cancer
Braden Scale [40]	Predicting pressure ulcer risk
EuroScore [30]	Short-term mortality after cardiac surgery
Framingham [38]	Risk of CVD over 10 yr
NEWS2 [26]	Identifying acutely ill patients

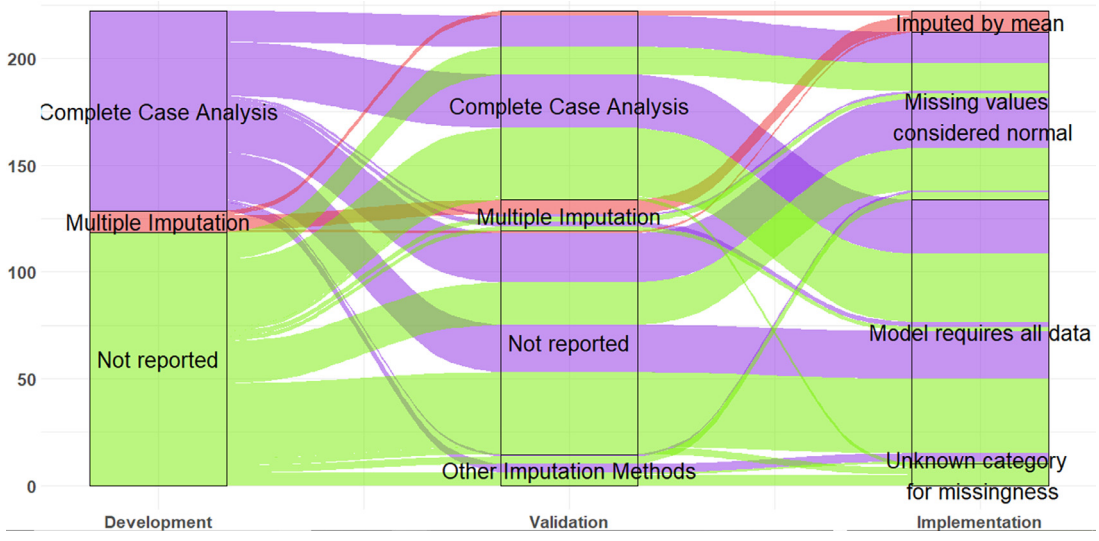

Fig. 2. Missing data handling across the pipeline of a CPM. A Sankey diagram, showing different "paths" of handling missing data across the three stages of a CPM's pipeline. The X axis shows the stages of a CPM's pipeline, whilst the Y axis shows the number possible combinations based on the number of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

only three CPMs: Thoracoscore [22], MTS [43], APGAR [42] score.

3.3. The TRIPOD statement

Across all the development and external validation articles, 27 studies (12%) stated that potential bias and inconsistency in the results might have occurred owing to

Table 2. Identified missing data handling methods

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

* missing values considered normal - for categorical predictors, where one would assume e.g. that the patient does not have a condition (similar to the "risk factor absent" approach), the lowest value possible was automatically imputed.

missing data. Of the development studies, 11 (48%) have followed point 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 point 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 (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 (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 [22], Manchester Triage System [43] and APGAR [42] 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 [44].

CCA was the commonest approach 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 [45]. This could be because most of the articles were published before the TRIPOD [17] 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 (MI) arise when the researcher only has access to published parameter estimates. 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 [46]. 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, 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 [47]. Another study has stated that the presence of a specific variable could have changed the coefficients in the remaining variables [48]. 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 [49,50]. Many studies have stated that missing data could have affected their findings [51–55]. 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 pipeline of CPMs recommended for use in the UK. 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, our choice to use Google Scholar to search for external validation papers might mean that some validation papers were missed. The reason we chose to use Google scholar is closely linked to our pragmatic decision to only include the top 10 most cited validation articles for each of the CPMs. Our motivation behind this choice was that (i) it would not have been possible to find all the validation articles for a certain CPM and (ii) we presume that the top 10 most cited articles would have had a bigger influence on this field, rather than choosing validation articles randomly. Therefore, we started our data collection process by asking NICE directly to pro-

vide 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. Hence, 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 articles included only the top 10 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. Lastly, the definition of consistency was created through discussions between the authors, which might make it subjective. Therefore, future work is required to firm up this definition.

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. A framework for handling missing data is needed to quality assure CPM pipelines. Through computer simulations and "real-world data," future work will be testing whether inconsistent handling of missing data has any effect on the CPMs' predictive performance.

Funding

This work was supported by EPSRC iCASE and Microsoft Research through its PhD Scholarship Programme. Grant number: 18000110.

Acknowledgments

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. We wish to thank two anonymous reviewers and the Editors for their constructive comments which helped improve the article upon revision.

Supplementary materials

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

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2021.09.008.

References

[1] Steyerberg EW. Clinical prediction models: application of prediction models. In: Springer Sci. + Bus. Media; 2009. p. 101–11. doi:10.1007/978-0-387-77244-8.
[2] Riley, R. D., van der Windt, D., 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. 2021
[4] Mercaldo SF, Blume JD. Missing data and prediction: the pattern submodel. Biostatistics 2020;21:236–52.
[5] Little, R. J. A. & Rubin, D. B. Statistical analysis with missing data. 2021
[6] Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38:1276–96.
[7] Janssen KJM, Rogier A, Donders T, Harrell FE Jr, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol. 2010;63:721–7.
[8] Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. Biometrical J 2015;57:614–32.
[9] Mertens BJA, Banzato E, Wreede LC. 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 2020;62:724–41.
[10] Enders CK. Multiple imputation as a flexible tool for missing data handling in clinical research. Behav Res Ther 2017;98:4–18.
[11] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ (Online) 2009;339:157–60.
[12] Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. Biometrical J 2015;57:614–32.
[13] Hoogland J, Barreveld Mvan, Debray TPA, Reitsma JB, Verstraelen TE, Dijkgraaf MGW, et al. Handling missing predictor values when validating and applying a prediction model to new patients. Stat Med 2020;39:3591–607.
[14] Sperrin M, Martin GP, 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 Mvan, Groenwold RHH, Moons KG. A cautionary note on the use of the missing indicator method for handling missing data in prediction research. J Clin Epidemiol. 2020;0:188–90.
[16] MDCalc - Medical calculators, equations, scores, and guidelines. https://www.mdcalc.com/, (access date 22 September 2020).
[17] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 2015;162:55–63.
[18] Waterlow J. Pressure sores: a risk assessment card. Nurs Times 1985;81:49–55.

[19] Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. Br Med J 2007;335:136–41.
[20] Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008;336:1475–82.
[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. doi:10.1136/bmj.j2099.
[22] Falcoz PE, Conti M, Brouchet L, Chocron S, Puyraveau M, Mercier M, 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 2007;133:325–32.
[23] Gray LJ, Taub NA, Khunti K, Gardiner E, Hiles S, Webb DR, Srinivasan BT, 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. 2010;27:887–95.
[24] Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX™ and the assessment of fracture probability in men and women from the UK. Osteoporos Int 2008;19:385–97.
[25] Antoniou AC, Pharoah PDP, McMullan G, Day NE, Stratton MR, Peto J, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 2002;86:76–83.
[26] National Early Warning Score (NEWS) 2 | RCP London. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2, (access date 22 September 2020).
[27] Rothwell PM, Giles MF, Flossmann E, Lovelock CE, Redgrave JNE, Warlow CP, et al. A simple score (ABCD) to identify individuals at high early risk of stroke after transient ischaemic attack. Lancet 2005;366:29–36.
[28] Morrow DA, Antman EM, Charlesworth A, Cairns R, Murphy SA, de Lemos JA, 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 2000;102:2031–7.
[29] Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003;58:377–82.
[30] Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardio-Thoracic Surg. 1999;16:9–13.
[31] Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. 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 2010;137:263–72.
[32] Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. 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. 1990;97:922–9.
[33] Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45:361–6.
[34] Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010;12:1–10.
[35] Blatchford O, Murray WR, Blatchford M. A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. Lancet 2000;356:1318–21.
[36] Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. 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 2010;138:1093–100.
[37] Eagle KA, Lim MJ, Dabbous OH, Pieper KS, Goldberg RJ, Van de Werf F, 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 2004;291:2727–33.
[38] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837–47.
[39] DF G, GT M. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 1974;111:58–64.
[40] The braden scale for predicting pressure sore risk - PubMed. https://pubmed.ncbi.nlm.nih.gov/3299278/, (access date 20 January 2020).
[41] WA K, EA D, DP W, JE Z. APACHE II: a severity of disease classification system. Crit Care Med 1985;13:818–29.
[42] Apgar V. A proposal for a new method of evaluation of the newborn infant. In: Anesthesia and analgesia, 120. Lippincott Williams and Wilkins; 2015. p. 1056–9.
[43] *Emergency triage. Emergency triage*. John Wiley & Sons Ltd; 2013. doi:101002/9781118299029.
[44] Sperrin M, Martin GP, Sisk R, Peek N. Missing data should be handled differently for prediction than for description or causal explanation. J Clin Epidemiol 2020;125:183–7.
[45] Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. BMC Med 2018;16:120.
[46] 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.
[47] 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 2014;4:5809.
[48] Chamogeorgakis TP, Connery CP, Bhora F, Nabong A, Toumpoulis IK. Thoracoscore predicts midterm mortality in patients undergoing thoracic surgery. J. Thorac Cardiovasc Surg 2007;134:883–7.
[49] Johnston SC, Rothwell PM, Nguyen-Huynh MN, Giles MF, Elkins JS, Bernstein AL, et al. Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. Lancet 2007;369:283–92.
[50] Carr E, Bendayan R, Bean D, Stammers M, Wang W, Zhang H, Searle T, et al. Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. BMC Med 2021 191 2021;19:1–16.
[51] Lambertini M, Pinto AC, Ameye L, Jongen L, Del Mastro L, Puglisi F, et al. The prognostic performance of Adjuvant! Online and Nottingham Prognostic Index in young breast cancer patients. Br J Cancer 2016;115:1471–8.
[52] Barber SR, Dhalwani NN, Davies MJ, Khunti K, Gray LJ. External national validation of the Leicester Self-Assessment score for Type 2 diabetes using data from the English longitudinal study of ageing. Diabet Med 2017;34:1575–83.
[53] Myint PK, Kamath AV, Vowler SL, Maisey DN, Harrison BDW. 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 2006;35:286–91.

[54] Greenland P, LaBree L, Azen SP, Doherty TM, Detrano RC. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. J Am Med Assoc 2004;291:210–15.

[55] Zachariasse JM, Seiger N, Rood PPM, Alves CF, Freitas P, Smit FJ, et al. Validity of the Manchester triage system in emergency care: a prospective observational study. PLoS One 2017;12(2) p.e0170811.