Literature Report

MA Barrowman

Contents

1 Introduction **1**

2 Clinical Prediction Models **1**

2.1 Fundamental Prognosis Research 2

2.2 Prognostic Factor Research 2

2.3 Prognostic Model Research 3

2.4 Stratified Medicine 6

2.5 Examples 6

3 Competing Risks & Multi-State Models **6**

4 Chronic Kidney Disease **6**

4.1 Clinical Prediction Models 6

4.2 Multi-State Models 6

References **6**

Last updated: 01 May

1 Introduction

2 Clinical Prediction Models

The idea of prognosis dates back to ancient Greece with the work of Hippocrates [1] and is derived from the Greek for "know before" meaning to forecast the future. Within the sphere of healthcare, it is definde as the risk of future health outcomes in patients, particularly patients with a certain disease or health condition. Prognosis allows clinicians to provide patients with a prediction of how their disease will progress and is uaully given as a probability of having an event in a prespecified number of years. For example, QRISK3 [2] provides a probability that a patient will have a heart attack or stroke in the next 10 years. Prognostic research encompasses any work which enhances the field of prognosis, whether through methodological advancements, field-specific prognostic modelling or educational material designed to improve general knowledge of prognosis. Prognostic models come under the wider umbrella of predictive models which also includes diagnostic models; because of this most of the keys points in the field or prognostic modeling can be applied to diagnostic models with little to no change.

Prognosis allows clinicians to evaluate the natural history of a patient (i.e. the course of a patient's future without any intervention) in order to establish the ffect of screening for asymptomatic diseases (such as with mammograms[3]). Prognosis research can be used to develop new definitions of diseases, whether a redefinition of an existing disease (such as the extension to th definition of myocardial infarction to include non-fatal events [4]) or a previously unknown subtype of a disease (such as Brugada syndrome as a type of cardiovascular disease[5])

In general, prognosis research can be broken down into four main categories, with three subcategories [6]:

- Type I: Fundamental prognosis research [3]
- Type II: Prognostic factor research [7]
- Type III: Prognostic model research [8]
  - Model development [9]
  - Model validation [10]
  - Model impact evaluation [11]
- Type IV: Stratified Medicine [12]

For a particular outcome, prognostic research will usually progress through these types, beginning with papers designed to evaluate overall prognosis within a whole population and then focusing in on more specificity and granularity towards individualised, causal predictions.

The model development and validation will usually occur in the same paper [13], [14]. studies into all three of the subcategories of prognostic model research *should* be completed before a model is used in clinical practice [15], although this does not always occur [8]. External validation is considered by some to be more important than the actual deviration of the model as it demonstrates generalisability of the model [16], whereas a model on it's own may be highly susceptible to overfitting [**Cite: Something**].

2.1 Fundamental Prognosis Research

[**What is it? Old definition is incorrect, so will need to write this fresh**]

2.2 Prognostic Factor Research

The aim of prognostic factor research (Type II) is to discover which factors are associated with disease progression. This allows for the general attribution of relationships between predictors and clinical outcomes.

Predictive factor research can give researchers and clinicians an idea of which patient factors are important when assessing a disease. It is vital to the development of clinical predictive models as without an idea of what covariates *can* affect an outcome, we cannot figure out which variables *will* affect the outcome. For example, [**xxxx**] demonstrated that [**xxxx**] is correlated with [**xxxx**], which subsequently used as a covariate in the development of the [**xxxx**] model. Note the use of the word correlate here as prognostic relationships do not have to be causal ones [**Cite: Something**]. These factors may indeed represent an underlying causal pathway, but this is not a requirement and it would require aetiological methods to discern whether it were causal or not. For example, when predicting [**xxxx**], we can demonstrate that [**xxxx**] is a prognostic factor, [however since the arrow of causation is [**xxxx**]] [**OR**] [however since [**xxxx**] causes both [**xxxx**] and [**xxxx**]], the relationship is prognostic, but not causal. [**Previously used Apgar score here, reference 40**]

Counter to the idea that prognostic factors aren't always causal, they are *always* confounding factors for the event they predict. Thue prognostic factors should be taken into account when planning clinical trials as if they are wildly misbalanced across the arms (or not accounted for in some other manner), they can cause biases in the results [7]. Sometimes these factors are so strong that adjusting the results of a clinical trial by the factor can affect, or even reverse the interpretation of the results [17]. If a prognostic factor is causal, then by directly affecting the factor, it can causally affect the outcome. By discovering new prognostic

factors, and investigating their causality, we can potentially open the door to new directions of attack for treatments.

It is unfortunate, however, that Riley at al **???** found that only 35.5% of prognostic factor studies in paediatric oncology actually reported the size of the effect of the prognostic factor they reported on. This means that very little information can be drawn from these studies. It is also important that prognostic factor research papers consider and report on the implications of the factor they assess such as healthcare costs. These kinds of implications are rarely assessed, especially when compared to drugs or interventions [7].

2.3 Prognostic Model Research

Predictive factors can be combined into a predictive model, which is a much more specific measurement of the effect of a factor on an outcome [8] and they are deigned to augment the job of a clinician; and not to completely replace them [11]. Diagnostic prediction model can be used to indicate whether a patient is likely to need further testing to establish the presence of a disease [13;~moons_transparent_2015]. Prognostic prediction models can be used to decide on further treatment for that patient, whether as a member of a certain risk group, or under a stratied medicine approach [13], [14]. Outcomes being assessed in a prediction model should be directly relevant to the patient (such as mortality) or have a direct causal relationship with something that is [11]. There is a trend of researchers focusing on areas of improvement that are of less significance to the patient than it is to a physician **???**. For example, older patient's might prefer to have an improved quality of life than an increase in life expectancy, and thus models should be developed to account for this.

Creating a clinicaly useful model is not as simple as just using some availble data to develop a model, despite what a lot of researchers seem to believe [**Cite: Something**]. To quote Steyerberg et al [8]. " To be useful for clinicia,s a prognostic model needs to provide validated and accurate predictions and to improve patient outcomes and cost-effectiveness of care". This means that, although a mdel might appear to be useful, its effectiveness is only relevant to the population it was developed in. If your population is different, then the model will behave differently. Bleeker [18] developed a model to predict bacterial infections in febrile children with an unknown source. The model scored well when assessed for the predictive value in the development dataset, however it scored much worse in an external dataset implying that, though it worked well in the development population, it would be unwise to apply it to a new population.

2.3.1 Model Development

The first stage of having a useful model is to develop one. Clinical predictive models can take a variety of forms, such as logistic regression, cox models or some kind of machine learning. Regardless of the specific model type being used, there are certain universal truths than should be held up during model development which will be discussed here. The size of the dataset being used is of vital importance as it can combat overfitting of the data, but so is choosing which prognostic factors to be included in the final model. This section will discuss various ideas that researchers need to account for when developing a model from any source and can be applied to any model type.

By considering a multivariable approach to prediction models (as opposed to a univariable one), researchers can consider different combinations of predictive factors, usually refered to as potential predictors [7]. These can include factors where a direct relationship with the disease can be clearly seen, such as tumour size in the prediction of cancer mortality **???**, or ones which could have a more general effect on overall health, such as socioeconomic and ethnicity variables **???**. By ignoring any previous assumptions about a correlation between these potential predictors and the outcome of interest, we can cast a wider net in our analysis allowing us to catch relationships that might have otherwise been lost [19]. Prediction models should take into account as many predictive factors as possible. Demographic data should also be included as these are often found to be confounding factors, variables such as ethnicity and social deprivation risk exacerbating the existing inequality between groups [20].

When developing a predictive model, the size of the dataset being used in an important consideration. A typical "rule of thumb" is to have at least 10 events for every potential predictor **???**, [21], know as the Events-per-Variable (EPV). Recently, this number has been superseded by a methods to evaluate a specific required sample size [22]. If there aren't enough events to satisfy this criteria, then some potential predictors should be eliminated before any formal analysis takes place (for example using clinical knowledge) [23]. In general, it is also recommended that this development dataset contain at least 100 events (regardless of number of potential predictors) **???**, [15], [24]. A systematic review by Counsell et al [25] found that out of eighty-three prognostic models for acute stroke, less than 50% of them had more than 10 EPV, and the work by Riley et al [22] showed that less that [**Pull example from Riley EPV**]. Having a low EPV can lead to overfitting of the model which is a concern associated with having a small data set. Overfitting leads to a worse prediction when the model is used on a new population which essentially makes the model useless [9]. However, just because a dataset is large does not imply that it will be a *good* dataset if the quality of the data is lacking [15]. Having a large amount of data can lead to predictors being considered statistically significant when in reality they only add a small amount of information to the model [15]. The size of the effect of a predictor should therefore be taken into account in the final model and, if beneficial, some predictors can be dropped at the final stage.

Large datasets can be used for both development and validation if an effective subset is chosen. This subset should not be random or data driven and should be decided before data analysis is begun [15]. Randomly splitting a dataset set into a training set (for development) and a testing set (for internal validation) can result in optimistic results in the validation process in the testing set. This is due to the random nature of the splitting causing the two populations to be too exchangeable, which is similar to the logic behind the splitting of patients in a Randomised Control Trial (RCT). Splitting the population by a specific characteristic (such as geographic location or time period) can result in a better internal validation [10], [26]. Derivation of the QRISK2 Score [27] (known later as QRISK2-2008) randomly assigned two thirds of practices to the derivation dataset and the remainder to the validation dataset. This model was further externally validated **???**, and its most modern incarnation, QRISK3, performed the external validation in the same paper [2] The Nottingham Prognostic Index (NPI) was trained on the first 500 patients admitted to Nottingham City Hospital after the study began [28] and later validated on the next 320 patients to be admitted [29], this validation was not performed at the same time as the initial development and is thus an external validation.

As with any technology, clinicians and researchers should be wary of models becoming outdated [30]. Healthcare systems and lifestyles change over time, and so models developed and externally validated in an outdated population will drift [31] and so should be updated regularly, as with QRISK [2] or automatically with a dynamic model [32]

If a sufficient amount of data is available and it has been taken from multiple sources (practices, clinics or studies), then it should be clustered to account for heterogeneity across sources [33]. It is important that any sources of potential variability are identified (such as heterogeneity between centres) as this can have an impact on the results of any analysis [3], [15]. Heterogeneity is particularly high when using multiple countries as a source of data [34] or if a potential predictor is of a subjective nature, which leads to discrepancies between assessors **???**. Overlooking of this clustering can lead to incorrect inferences [33]. The generalisability of the sources of data should also be considered in the development of a model. For example, the inclusion and exclusion criteria of an RCT can greatly reduce generalisability if used as a data source [11].

A prediction model researcher needs to select clinically relevant potential predictors for use in the development of the model [9]. Once chosen, researchers need to be very specific about how these variables are treated. Any adjustments from the raw data should be reported in detail [13], [14]. Potential predictors with high levels of missingness should be excludes as this missingness can introduce bias [9]. One key fact that many experts agree on is that categoriation of continuous predictors should be avoided [**Cite: LOADS**] as it retains much more predictive information. The cut-points of these categorisations lead to artificial jumps in the outcome risk [23]. It is also worth noting that cut-points are often either arbitrarily decided or data-driven with the latter leading to overfitting [23]. If categorisation is performed, clear rationale should be provided with an ackowledgement that this wil reduce performance **???**, [16]. When applying a model to a new population, extrapolation of a model should be avoided **???** and so to aid in this, the ranges

of continuous variables, and the considered values of categorical variables should be reported [16]. this is especially true for age. QRISK2 was derived in a population ranging from 35 to 74 years of ages and so should not have been applied to patients out of this range [20]. This ranges was later extended with the updated version **???** and currently can be applied to patients aged 25-84 [**Update with QRISK3**].

When building a prediction model, we begin with a certain pool of potential predictors and try to establish which to include in the final model [23]. With k candidate variables, we have 2^k possible choices which can get unwieldy even for low values of k, with only 10 predictors (a very reasonable number), there are over 1,000 combinations. This doesn't include interactions or non-linear components which increases this number even more. Therefore, model-building techniques are important for anybody attempting to build an accurate prediction model. It is currently undecided what the "best" way to select predictors in a multivariable model is or even if it exists [23]. One method that researchers use to decide on which predictors to include is to analyse each potential predictor individually for a correlation with the outcome in a univariable analysis and keeping those which are considered to have a statistically significant correlation. The general consensus amongst researchers is that predictors should not be excluded in this way [9]. Univariables analysis does not account for any dependencies between potential predictors and so any cross correlations that exists between them can cause a bias in the results. Despite its clear weaknesses, any prognostic studies still use univariable analysis to build their models **???**.

Backwards elimination (BE) involves starting with all potential predictors in the model and removing ones which do not reach a certain level of statistical significant (for example, 5%) one at a time untill all remaining variables are significant. Forward selection begins wth no variables and adds one a a time based on similar criteria. Under either of these methods, a lower significance level will exxlucde more variables [9]. Backward elimination of variables is preferable over forward selection as users are less likely to end up in local minima **???**. A variant of these techniques is to use the Akaike Information Criteria (AIC) rather than statistical significance. This method avoids the comparison to p-values and so is often preferable to build robust models [**Cite: p-values be bad reference**]. For this method, to establish which predictors should be removed at each step, the model is re-built with each of the predictors individually removed, and the AIC is calculated. The model with the lowest AIC is chosen to be the new model and the process is repeated. This process is repeated until the removal of a predictor would increase the AIC (i.e. make the model's fit worse). This same technique can be applied to a forward selection style model or, if the computing power is available, a backward-forward elimination technique were predictors are added or removed at each stage. The advantage of this method is that it avoids local minima better by trying more combinations.

2.3.2 Model Validation

2.3.3 Impact Evaluation

2.4 Stratified Medicine

2.5 Examples

3 Competing Risks & Multi-State Models

4 Chronic Kidney Disease

4.1 Clinical Prediction Models

4.2 Multi-State Models

References

[1] Hippocrates and F. Adams, *The genuine works of Hippocrates;* New York, W. Wood and company, 1886.

[2] J. Hippisley-Cox, C. Coupland, and P. Brindle, "Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study," *BMJ*, vol. 357, May 2017, doi: 10.1136/bmj.j2099.

[3] H. Hemingway *et al.*, "Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes," *BMJ*, vol. 346, p. e5595, Feb. 2013, doi: 10.1136/bmj.e5595.

[4] K. Thygesen, J. S. Alpert, H. D. White, and Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction, "Universal definition of myocardial infarction," *Journal of the American College of Cardiology*, vol. 50, no. 22, pp. 2173–2195, Nov. 2007, doi: 10.1016/j.jacc.2007.09.011.

[5] Probst *et al.*, "Long-Term Prognosis of Patients Diagnosed With Brugada Syndrome," *Circulation*, vol. 121, no. 5, pp. 635–643, Feb. 2010, doi: 10.1161/CIRCULATIONAHA.109.887026.

[6] R. D. Riley, D. van der Windt, P. Croft, and K. G. M. Moons, *Prognosis Research in Healthcare: Concepts, Methods, and Impact*, First. Oxford University Press, 2019.

[7] R. D. Riley *et al.*, "Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research," *PLoS Medicine*, vol. 10, no. 2, Feb. 2013, doi: 10.1371/journal.pmed.1001380.

[8] E. W. Steyerberg *et al.*, "Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research," *PLOS Medicine*, vol. 10, no. 2, p. e1001381, Feb. 2013, doi: 10.1371/journal.pmed.1001381.

[9] P. Royston, K. G. M. Moons, D. G. Altman, and Y. Vergouwe, "Prognosis and prognostic research: Developing a prognostic model," *BMJ*, vol. 338, p. b604, Mar. 2009, doi: 10.1136/bmj.b604.

[10] D. G. Altman, Y. Vergouwe, P. Royston, and K. G. M. Moons, "Prognosis and prognostic research: Validating a prognostic model," *BMJ*, vol. 338, p. b605, May 2009, doi: 10.1136/bmj.b605.

[11] K. G. M. Moons, P. Royston, Y. Vergouwe, D. E. Grobbee, and D. G. Altman, "Prognosis and prognostic research: What, why, and how?" *BMJ*, vol. 338, p. b375, Feb. 2009, doi: 10.1136/bmj.b375.

[12] A. D. Hingorani *et al.*, "Prognosis research strategy (PROGRESS) 4: Stratified medicine research," *BMJ*, vol. 346, p. e5793, Feb. 2013, doi: 10.1136/bmj.e5793.

[13] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement," *BMC Medicine*, vol. 13, no. 1, p. 1, Jan. 2015, doi: 10.1186/s12916-014-0241-z.

[14] K. G. M. Moons *et al.*, "Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration," *Annals of Internal Medicine*, vol. 162, no. 1, p. W1, Jan. 2015, doi: 10.7326/M14-0698.

[15] R. D. Riley *et al.*, "External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: Opportunities and challenges," *BMJ*, vol. 353, Jun. 2016, doi: 10.1136/bmj.i3140.

[16] G. S. Collins, O. Omar, M. Shanyinde, and L.-M. Yu, "A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods," *Journal of Clinical Epidemiology*, vol. 66, no. 3, pp. 268–277, Mar. 2013, doi: 10.1016/j.jclinepi.2012.06.020.

[17] P. Royston, D. G. Altman, and W. Sauerbrei, "Dichotomizing continuous predictors in multiple regression: A bad idea," *Statistics in Medicine*, vol. 25, no. 1, pp. 127–141, Jan. 2006, doi: 10.1002/sim.2331.

[18] S. E. Bleeker *et al.*, "External validation is necessary in prediction research: A clinical example," *Journal of Clinical Epidemiology*, vol. 56, no. 9, pp. 826–832, Sep. 2003, doi: 10.1016/s0895-4356(03)00207-5.

[19] S. B. Hanauer, "Exploring the controversial themes of IBD," *Inflammatory Bowel Diseases*, vol. 15, no. S1, pp. S1–S10, 2009, doi: 10.1002/ibd.20945.

[20] J. Hippisley-Cox *et al.*, "Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2," *BMJ*, vol. 336, no. 7659, pp. 1475–1482, Jun. 2008, doi: 10.1136/bmj.39609.449676.25.

[21] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *Journal of Clinical Epidemiology*, vol. 49, no. 12, pp. 1373–1379, Dec. 1996, doi: 10.1016/S0895-4356(96)00236-3.

[22] R. D. Riley *et al.*, "Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes," *Statistics in Medicine*, vol. 38, no. 7, pp. 1276–1296, 2019, doi: 10.1002/sim.7992.

[23] W. Sauerbrei, P. Royston, and H. Binder, "Selection of important variables and determination of functional form for continuous predictors in multivariable model building," *Statistics in Medicine*, vol. 26, no. 30, pp. 5512–5528, 2007, doi: 10.1002/sim.3148.

[24] Y. Vergouwe, E. W. Steyerberg, M. J. C. Eijkemans, and J. D. F. Habbema, "Substantial effective sample sizes were required for external validation studies of predictive logistic regression models," *Journal of Clinical Epidemiology*, vol. 58, no. 5, pp. 475–483, May 2005, doi: 10.1016/j.jclinepi.2004.06.017.

[25] C. Counsell and M. Dennis, "Systematic review of prognostic models in patients with acute stroke," *Cerebrovascular Diseases (Basel, Switzerland)*, vol. 12, no. 3, pp. 159–170, 2001, doi: 10.1159/000047699.

[26] J. Ivanov, M. A. Borger, T. E. David, G. Cohen, N. Walton, and C. D. Naylor, "Predictive accuracy study: Comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery," *The Annals of Thoracic Surgery*, vol. 70, no. 1, pp. 162–168, Jul. 2000, doi: 10.1016/s0003-4975(00)01387-4.

[27] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, M. May, and P. Brindle, "Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study," *BMJ (Clinical research ed.)*, vol. 335, no. 7611, p. 136, Jul. 2007, doi: 10.1136/bmj.39261.471806.55.

[28] J. L. Haybittle *et al.*, "A prognostic index in primary breast cancer." *British Journal of Cancer*, vol. 45, no. 3, pp. 361–366, Mar. 1982.

[29] J. H. Todd *et al.*, "Confirmation of a prognostic index in primary breast cancer," *British Journal of Cancer*, vol. 56, no. 4, pp. 489–492, Oct. 1987, doi: 10.1038/bjc.1987.230.

[30] A. Pate, R. Emsley, D. M. Ashcroft, B. Brown, and T. van Staa, "The uncertainty with using risk prediction models for individual decision making: An exemplar cohort study examining the prediction of cardiovascular disease in English primary care," *BMC Medicine*, vol. 17, no. 1, p. 134, Jul. 2019, doi: 10.1186/s12916-019-1368-8.

[31] P. Bhatnagar, K. Wickramasinghe, J. Williams, M. Rayner, and N. Townsend, "The epidemiology of cardiovascular disease in the UK 2014," *Heart*, vol. 101, no. 15, pp. 1182–1189, Aug. 2015, doi: 10.1136/heartjnl-2015-307516.

[32] D. A. Jenkins, M. Sperrin, G. P. Martin, and N. Peek, "Dynamic models to predict health outcomes: Current status and methodological challenges," *Diagnostic and Prognostic Research*, vol. 2, no. 1, p. 23, Dec. 2018, doi: 10.1186/s41512-018-0045-2.

[33] B. Liquet, J.-F. Timsit, and V. Rondeau, "Investigating hospital heterogeneity with a multi-state frailty model: Application to nosocomial pneumonia disease in intensive care units," *BMC medical research methodology*, vol. 12, p. 79, Jun. 2012, doi: 10.1186/1471-2288-12-79.

[34] K. I. E. Snell *et al.*, "Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model," *Journal of Clinical Epidemiology*, vol. 69, pp. 40–50, Jan. 2016, doi: 10.1016/j.jclinepi.2015.05.009.