Literature Report

MA Barrowman

Contents

Last updated: 25 May

1 Introduction

2 Clinical Prediction Models

The idea of prognosis dates back to ancient Greece with the work of Hippocrates (Hippocrates and Adams 1886) 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 (Hippisley-Cox, Coupland, and Brindle 2017) 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(Hemingway et al. 2013)). 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 (Thygesen et al. 2007)) or a previously unknown subtype of a disease (such as Brugada syndrome as a type of cardiovascular disease(Probst et al. 2010))

In general, prognosis research can be broken down into four main categories, with three subcategories (Riley, van der Windt, et al. 2019):

- Type I: Fundamental prognosis research (Hemingway et al. 2013)
- Type II: Prognostic factor research (Riley et al. 2013)
- Type III: Prognostic model research (Steyerberg et al. 2013)
    - Model development (Royston et al. 2009)
    - Model validation (Altman et al. 2009)
    - Model impact evaluation (Moons et al. 2009)
- Type IV: Stratified Medicine (Hingorani et al. 2013)

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 (Collins et al. 2015; Moons et al. 2015). studies into all three of the subcategories of prognostic model research *should* be completed before a model is used in clinical practice (Riley et al. 2016), although this does not always occur (Steyerberg et al. 2013). External validation is considered by some to be more important than the actual deviration of the model as it demonstrates generalisability of the model (Collins et al. 2013), 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 (Riley et al. 2013). 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 (Royston, Altman, and

Sauerbrei 2006). 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 (Riley et al. 2013).

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 (Steyerberg et al. 2013) and they are deigned to augment the job of a clinician; and not to completely replace them (Moons et al. 2009). Diagnostic prediction model can be used to indicate whether a patient is likely to need further testing to establish the presence of a disease [Collins et al. (2015);~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 (Collins et al. 2015; Moons et al. 2015). 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 (Moons et al. 2009). 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 (Steyerberg et al. 2013). " 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 (Bleeker et al. 2003) 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 (Riley et al. 2013). 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 (Hanauer 2009).

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 (Hippisley-Cox et al. 2008).

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 (**???**; Peduzzi et al. 1996), know as the Events-per-Variable (EPV). Recently, this number has been superseded by a methods to evaluate a specific required sample size (Riley, Snell, et al. 2019) based on Events-per-Predictor (EPP), where categorical variables are transformed into dummy variables prior to calculation (therefore number of predictors is higher than the number of variables). 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) (Sauerbrei, Royston, and Binder 2007). In general, it is also recommended that this development dataset contain at least 100 events (regardless of number of potential predictors) (Riley et al. 2016; Vergouwe et al. 2005; **???**). A systematic review by Counsell et al (Counsell and Dennis 2001) 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 (Riley, Snell, et al. 2019) 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 (Royston et al. 2009). 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 (Riley et al. 2016). 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 (Riley et al. 2016). 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 (Riley et al. 2016). 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 (Altman et al. 2009; Ivanov et al. 2000). Derivation of the QRISK2 Score (Hippisley-Cox et al. 2007) (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 (Hippisley-Cox, Coupland, and Brindle 2017) The Nottingham Prognostic Index (NPI) was trained on the first 500 patients admitted to Nottingham City Hospital after the study began (Haybittle et al. 1982) and later validated on the next 320 patients to be admitted (Todd et al. 1987), 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 (Pate et al. 2019). Healthcare systems and lifestyles change over time, and so models developed and externally validated in an outdated population will drift (Bhatnagar et al. 2015) and so should be updated regularly, as with QRISK (Hippisley-Cox, Coupland, and Brindle 2017) or automatically with a dynamic model (Jenkins et al. 2018)

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 (Liquet, Timsit, and Rondeau 2012). 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 (Hemingway et al. 2013; Riley et al. 2016). Heterogeneity is particularly high when using multiple countries as a source of data (Snell et al. 2016) 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 (Liquet, Timsit, and Rondeau 2012). 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 (Moons et al. 2009).

A prediction model researcher needs to select clinically relevant potential predictors for use in the development of the model (Royston et al. 2009). 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 (Collins et al. 2015; Moons et al. 2015). Potential predictors with high levels of missingness should be excludes as this missingness can introduce bias (Royston et al. 2009). 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 (Sauerbrei, Royston, and Binder 2007). It is also worth noting that cut-points are often either arbitrarily decided or data-driven with the latter leading to overfitting (Sauerbrei, Royston, and Binder 2007). If categorisation is performed, clear rationale should be provided with an ackowledgement that this wil reduce performance (**???**; Collins et al. 2013). 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 (Collins et al. 2013). 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 (Hippisley-Cox et al. 2008). 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 (Sauerbrei, Royston, and Binder 2007). 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 (Sauerbrei, Royston, and Binder 2007). 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 (Royston et al. 2009). 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 (Royston et al. 2009). 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.

It is also important to assess non-linearity relationships between variables and outcomes to ensure the relationship is accurately modeled. This can be done using standard transformations (e.g. logarithms, sqauring) or using fractional polynomials [**Cite: LR - 7 or 76?**]. Interactions between terms also need to be checked for the same reasons, and when interactions are strong, it may be useful to completely stratify by a factor, rather than including as a covariate in the model. Strong interactions can be an indicator for a differential response amongst populations and so should be investigated directly [**Cite: LR - 4**]. If a predictor is ex-

pensive or invasive, it may be better to include a less significant predictor which is easier to come by [**Cite: LR - 40**]. A limiting factor for some prognostic models is that the prognostic factors they measure are not readily available or are not used in routine care [**Cite: LR - 3**]. The measurement (or lack thereof) can also be an indicator of patient health and so researchers need to be aware of these causal links when analysing measurements [**Cite: Rose's Work**].

Once developed, prognostic models can be used to create risk groups for a population. Risk groups should be defined by clinical knowledge rather than statistical criteria[**Cite: LR - 35**]. Grouping patients into risk groups is not as accurate as using the specific model to provide an estimated risk [**Cite: LR - 3**].

2.3.2   Model Validation

2.3.3   Impact Evaluation

2.4   Stratified Medicine

2.5   Examples

2.5.1   Reviews

Along with the EPV assessment mentioned earlier, Counsell et al's systematic review [**Cite: LR - 62**] assessed other criteria related to validity, evaluation and practicality. Of those eighty-three models, only four met their requirements, none of which had been externally validated. The other seven criteria were:

- Adequate inception cohort
- Less than 10% loss to followup
- Prospective data collection
- Valid and reliable outcome
- Age as a candidate predictor
- Severity of condition as a candidate predictor
- Use of stepwise regression

3   Competing Risks & Multi-State Models

4   Chronic Kidney Disease

4.1   Clinical Prediction Models

4.2   Multi-State Models

References

Altman, Douglas G., Yvonne Vergouwe, Patrick Royston, and Karel G. M. Moons. 2009. "Prognosis and Prognostic Research: Validating a Prognostic Model." *BMJ* 338 (May): b605. https://doi.org/10.1136/bmj.b605.

Bhatnagar, Prachi, Kremlin Wickramasinghe, Julianne Williams, Mike Rayner, and Nick Townsend. 2015. "The Epidemiology of Cardiovascular Disease in the UK 2014." *Heart* 101 (15): 1182–9. https://doi.org/10.1136/heartjnl-2015-307516.

Bleeker, S. E., H. A. Moll, E. W. Steyerberg, A. R. T. Donders, G. Derksen-Lubsen, D. E. Grobbee, and K. G. M. Moons. 2003. "External Validation Is Necessary in Prediction Research: A Clinical Example." *Journal of Clinical Epidemiology* 56 (9): 826–32. https://doi.org/10.1016/s0895-4356(03)00207-5.

Collins, Gary S., Omar Omar, Milensu Shanyinde, and Ly-Mee Yu. 2013. "A Systematic Review Finds Prediction Models for Chronic Kidney Disease Were Poorly Reported and Often Developed Using Inappropriate Methods." *Journal of Clinical Epidemiology* 66 (3): 268–77. https://doi.org/10.1016/j.jclinepi.2012.06.020.

Collins, Gary S., Johannes B. Reitsma, Douglas G. Altman, and Karel GM Moons. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement." *BMC Medicine* 13 (1): 1. https://doi.org/10.1186/s12916-014-0241-z.

Counsell, C., and M. Dennis. 2001. "Systematic Review of Prognostic Models in Patients with Acute Stroke." *Cerebrovascular Diseases (Basel, Switzerland)* 12 (3): 159–70. https://doi.org/10.1159/000047699.

Hanauer, Stephen B. 2009. "Exploring the Controversial Themes of IBD." *Inflammatory Bowel Diseases* 15 (S1): S1–S10. https://doi.org/10.1002/ibd.20945.

Haybittle, J. L., R. W. Blamey, C. W. Elston, J. Johnson, P. J. Doyle, F. C. Campbell, R. I. Nicholson, and K. Griffiths. 1982. "A Prognostic Index in Primary Breast Cancer." *British Journal of Cancer* 45 (3): 361–66.

Hemingway, Harry, Peter Croft, Pablo Perel, Jill A. Hayden, Keith Abrams, Adam Timmis, Andrew Briggs, et al. 2013. "Prognosis Research Strategy (PROGRESS) 1: A Framework for Researching Clinical Outcomes." *BMJ* 346 (February): e5595. https://doi.org/10.1136/bmj.e5595.

Hingorani, Aroon D., Daniëlle A. van der Windt, Richard D. Riley, Keith Abrams, Karel G. M. Moons, Ewout W. Steyerberg, Sara Schroter, Willi Sauerbrei, Douglas G. Altman, and Harry Hemingway. 2013. "Prognosis Research Strategy (PROGRESS) 4: Stratified Medicine Research." *BMJ* 346 (February): e5793. https://doi.org/10.1136/bmj.e5793.

Hippisley-Cox, Julia, Carol Coupland, and Peter Brindle. 2017. "Development and Validation of QRISK3 Risk Prediction Algorithms to Estimate Future Risk of Cardiovascular Disease: Prospective Cohort Study." *BMJ* 357 (May). https://doi.org/10.1136/bmj.j2099.

Hippisley-Cox, Julia, Carol Coupland, Yana Vinogradova, John Robson, Margaret May, and Peter Brindle. 2007. "Derivation and Validation of QRISK, a New Cardiovascular Disease Risk Score for the United Kingdom: Prospective Open Cohort Study." *BMJ (Clinical Research Ed.)* 335 (7611): 136. https://doi.org/10.1136/bmj.39261.471806.55.

Hippisley-Cox, Julia, Carol Coupland, Yana Vinogradova, John Robson, Rubin Minhas, Aziz Sheikh, and Peter Brindle. 2008. "Predicting Cardiovascular Risk in England and Wales: Prospective Derivation and Validation of QRISK2." *BMJ* 336 (7659): 1475–82. https://doi.org/10.1136/bmj.39609.449676.25.

Hippocrates, and Francis Adams. 1886. *The Genuine Works of Hippocrates;* New York, W. Wood and company.

Ivanov, J., M. A. Borger, T. E. David, G. Cohen, N. Walton, and C. D. Naylor. 2000. "Predictive Accuracy Study: Comparing a Statistical Model to Clinicians' Estimates of Outcomes After Coronary Bypass Surgery." *The Annals of Thoracic Surgery* 70 (1): 162–68. https://doi.org/10.1016/s0003-4975(00)01387-4.

Jenkins, David A., Matthew Sperrin, Glen P. Martin, and Niels Peek. 2018. "Dynamic Models to Predict Health Outcomes: Current Status and Methodological Challenges." *Diagnostic and Prognostic Research* 2 (1): 23. https://doi.org/10.1186/s41512-018-0045-2.

Liquet, Benoit, Jean-François Timsit, and Virginie Rondeau. 2012. "Investigating Hospital Heterogeneity with a Multi-State Frailty Model: Application to Nosocomial Pneumonia Disease in Intensive Care Units." *BMC Medical Research Methodology* 12 (June): 79. https://doi.org/10.1186/1471-2288-12-79.

Moons, Karel G. M., Douglas G. Altman, Johannes B. Reitsma, John P. A. Ioannidis, Petra Macaskill, Ewout W. Steyerberg, Andrew J. Vickers, David F. Ransohoff, and Gary S. Collins. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration." *Annals of Internal Medicine* 162 (1): W1. https://doi.org/10.7326/M14-0698.

Moons, Karel G. M., Patrick Royston, Yvonne Vergouwe, Diederick E. Grobbee, and Douglas G. Altman. 2009. "Prognosis and Prognostic Research: What, Why, and How?" *BMJ* 338 (February): b375. https://doi.org/10.1136/bmj.b375.

Pate, Alexander, Richard Emsley, Darren M. Ashcroft, Benjamin Brown, and Tjeerd van Staa. 2019. "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* 17 (1): 134. https://doi.org/10.1186/s12916-019-1368-8.

Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvan R. Feinstein. 1996. "A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis." *Journal of Clinical Epidemiology* 49 (12): 1373–9. https://doi.org/10.1016/S0895-4356(96)00236-3.

Probst, Veltmann C., Eckardt L., Meregalli P.G., Gaita F., Tan H.L., Babuty D., et al. 2010. "Long-Term Prognosis of Patients Diagnosed with Brugada Syndrome." *Circulation* 121 (5): 635–43. https://doi.org/10.1161/CIRCULATIONAHA.109.887026.

Riley, Richard D., Joie Ensor, Kym I. E. Snell, Thomas P. A. Debray, Doug G. Altman, Karel G. M. Moons, and Gary S. Collins. 2016. "External Validation of Clinical Prediction Models Using Big Datasets from E-Health Records or IPD Meta-Analysis: Opportunities and Challenges." *BMJ* 353 (June). https://doi.org/10.1136/bmj.i3140.

Riley, Richard D., Jill A. Hayden, Ewout W. Steyerberg, Karel G. M. Moons, Keith Abrams, Panayiotis A. Kyzas, Núria Malats, et al. 2013. "Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research." *PLoS Medicine* 10 (2). https://doi.org/10.1371/journal.pmed.1001380.

Riley, Richard D., Kym IE Snell, Joie Ensor, Danielle L. Burke, Frank E. Harrell Jr, Karel GM Moons, and Gary S. Collins. 2019. "Minimum Sample Size for Developing a Multivariable Prediction Model: PART II - Binary and Time-to-Event Outcomes." *Statistics in Medicine* 38 (7): 1276–96. https://doi.org/10.1002/sim.7992.

Riley, Richard D., Danielle van der Windt, Peter Croft, and Karel G. M. Moons. 2019. *Prognosis Research in Healthcare: Concepts, Methods, and Impact.* First. Oxford University Press.

Royston, Patrick, Douglas G. Altman, and Willi Sauerbrei. 2006. "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea." *Statistics in Medicine* 25 (1): 127–41. https://doi.org/10.1002/sim.2331.

Royston, Patrick, Karel G. M. Moons, Douglas G. Altman, and Yvonne Vergouwe. 2009. "Prognosis and Prognostic Research: Developing a Prognostic Model." *BMJ* 338 (March): b604. https://doi.org/10.1136/bmj.b604.

Sauerbrei, Willi, Patrick Royston, and Harald Binder. 2007. "Selection of Important Variables and Determination of Functional Form for Continuous Predictors in Multivariable Model Building." *Statistics in Medicine* 26 (30): 5512–28. https://doi.org/10.1002/sim.3148.

Snell, Kym I. E., Harry Hua, Thomas P. A. Debray, Joie Ensor, Maxime P. Look, Karel G. M. Moons, and Richard D. Riley. 2016. "Multivariate Meta-Analysis of Individual Participant Data Helped Externally Validate the Performance and Implementation of a Prediction Model." *Journal of Clinical Epidemiology* 69 (January): 40–50. https://doi.org/10.1016/j.jclinepi.2015.05.009.

Steyerberg, Ewout W., Karel G. M. Moons, Danielle A. van der Windt, Jill A. Hayden, Pablo Perel, Sara Schroter, Richard D. Riley, Harry Hemingway, Douglas G. Altman, and for the PROGRESS Group. 2013. "Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research." *PLOS Medicine* 10 (2): e1001381. https://doi.org/10.1371/journal.pmed.1001381.

Thygesen, Kristian, Joseph S. Alpert, Harvey D. White, and Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. 2007. "Universal Definition of Myocardial Infarction." *Journal of the American College of Cardiology* 50 (22): 2173–95. https://doi.org/10.1016/j.jacc.2007.09.011.

Todd, J. H., C. Dowle, M. R. Williams, C. W. Elston, I. O. Ellis, C. P. Hinton, R. W. Blamey, and J. L. Haybittle. 1987. "Confirmation of a Prognostic Index in Primary Breast Cancer." *British Journal of Cancer* 56 (4): 489–92. https://doi.org/10.1038/bjc.1987.230.

Vergouwe, Yvonne, Ewout W. Steyerberg, Marinus J. C. Eijkemans, and J. Dik F. Habbema. 2005. "Substantial Effective Sample Sizes Were Required for External Validation Studies of Predictive Logistic Regression Models." *Journal of Clinical Epidemiology* 58 (5): 475–83. https://doi.org/10.1016/j.jclinepi.2004.06.017.