Seminars in Oncology Nursing 36 (2020) 151089

Contents lists available at ScienceDirect

Seminars in Oncology Nursing

journal homepage: https://www.journals.elsevier.com/seminars-in-oncology-nursing

How to Develop Statistical Predictive Risk Models in Oncology Nursing to Enhance Psychosocial and Supportive Care

Jenny Harris, BSc, MSc, PhD[a,*], Edward Purssell, RSCN, RGN, BSc, MSc, PhD[b], Emma Ream, RGN, BSc, MSc, PhD[a], Anne Jones, RGN, PGCAP, MSc, PhD[c], Jo Armes, RGN, BSc, MSc, PhD[a], Victoria Cornelius, BSc, PhD[d]

[a] School of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, United Kingdom
[b] School of Health Sciences, City, University of London, London, United Kingdom
[c] Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care, King's College London, London, United Kingdom
[d] Imperial Clinical Trials Unit, School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom

ARTICLE INFO

Key Words:
Predictive risk models
Regression models
Cancer
Psychosocial
Supportive care
Psychological
Distress

ABSTRACT

Objectives: Predictive risk models are advocated in psychosocial oncology practice to provide timely and appropriate support to those likely to experience the emotional and psychological consequences of cancer and its treatments. New digital technologies mean that large scale and routine data collection are becoming part of everyday clinical practice. Using these data to try to identify those at greatest risk for late psychosocial effects of cancer is an attractive proposition in a climate of unmet need and limited resource. In this paper, we present a framework to support the development of high-quality predictive risk models in psychosocial and supportive oncology. The aim is to provide awareness and increase accessibility of best practice literature to support researchers in psychosocial and supportive care to undertake a structured evidence-based approach.
Data Sources: Statistical prediction risk model publications.
Conclusion: In statistical modeling and data science different approaches are needed if the goal is to predict rather than explain. The deployment of a poorly developed and tested predictive risk model has the potential to do great harm. Recommendations for best practice to develop predictive risk models have been developed but there appears to be little application within psychosocial and supportive oncology care.
Implications for Nursing Practice: Use of best practice evidence will ensure the development and validation of predictive models that are robust as these are currently lacking. These models have the potential to enhance supportive oncology care through harnessing routine digital collection of patient-reported outcomes and the targeting of interventions according to risk characteristics.

Introduction

Oncology nurses have an important role in delivering high-quality and compassionate psychosocial and supportive care for people living with and affected by cancer.[1] An integral part of the cancer multidisciplinary team nurses are particularly well-placed to provide practical, informational, and emotional support in response to symptom and distress screening[2] and ongoing needs assessments.[3,4] The advent of digital health and increasingly routine collection of patient-reported outcomes (PROs)[5,6] present greater opportunities for oncology nurses to incorporate real-time patient feedback and probability-based risk assessments into their psychosocial and supportive care plans.

Predictive risk models (PRMs) also known as clinical prediction models, nomograms, risk indexes, or rules[7] are designed to predict an individual's risk of having, or developing, a specific condition or outcome based on multiple variables.[8-11] Well-known PRMs are used in practice to target the most appropriate screening programs, care, and treatments according to future risk such as the BOADICEA model used for risk stratification for breast cancer in the general population and for women with family history[12] and the Nottingham Prognostic Index (NPI)[13,14] or Adjuvant! Online[15,16] for the management of breast cancers. In cancer care, as with most areas of health care, the focus has been on developing PRMs related to primary disease outcomes such as death and occurrence or remission of disease.[17] However, other outcomes relevant to ongoing quality of life are beginning to receive more attention[9] including within supportive oncology.[18-21]

Developing PRMs to inform oncology nursing practice can be useful to inform and enhance follow-up care through heightened clinical awareness[9] or PRMs can be used alongside ongoing needs assessments

* Address correspondence to: Jenny Harris, BSc, MSc, PhD, School of Health Sciences, University of Surrey, Faculty of Health and Medical Sciences, Duke of Kent Building, Guildford, Surrey, United Kingdom GU2 7XH.
E-mail address: jen.harris@surrey.ac.uk (J. Harris).

https://doi.org/10.1016/j.soncn.2020.151089
0749-2081/

to target care.[22,23] Furthermore, as shared decision-making forms the cornerstone of modern oncology nursing practice,[24-26] PRMs have the potential to provide an evidence-based input to support shared decision-making[27] and patient-centered communication, particularly in the context of nurse led follow-up consultations and in considering supportive care options.[28]

However in psychosocial and supportive oncology model development, practice is not optimal and there appears to be a lack of awareness around differences between developing models for valid explanatory purposes (that is, causation or inference) and building powerful predictive models.[9,29,30] Our systematic review on predictors of anxiety after breast cancer treatment found that most existing research was cross-sectional, lacked reproducibility (including any model validation or testing in different contexts) and relied on methods best suited to explanation rather than prediction.[18] This focus on explanatory models is understandable; those involved in psychosocial oncology and supportive care research are more familiar with explanatory modeling as the goals are often to unpick the underlying mechanisms, to develop or confirm theories, and to identify treatments or interventions to target these mechanisms.[31] Additionally, training in most elementary courses in quantitative methods in health sciences cover only statistical inference (explanatory mechanisms) not prediction models.

Examples from across different sectors including health and social care demonstrate that the failure to apply best practice standards can lead to suboptimal models[32] and the deployment of poorly developed and inadequately tested PRM can lead not only to biased models ("garbage in, garbage out")[33] but suboptimal practices, care, and outcomes, and result in real harms to individuals.[34-36]

To develop powerful and safe predictive models for use in routine practice a different approach is required, with less emphasis on null hypothesis significance testing and greater emphasis on avoidance of overfitting the model to the nuances of the data. In this paper, we first explain why different approaches to model development are needed and then present a framework for psychosocial and supportive care oncology researchers to describe best practice for PRM development using a structured evidence-based approach. By providing this framework we hope to encourage greater uptake of good statistical practice. We draw on our own experience from using patient-reported outcomes and data, key literature, and best practice recommendations developed within the statistical community,[7,23,27,37-44] the latter of which have received relatively little application within the psychosocial and supportive cancer care. The intention is that this paper will provide a useful, accessible framework for a non-statistician oncology audience and related professionals involved in psychosocial oncology care and research.

Focus on Prediction, Not Explanation

Many researchers conceptually conflate explaining phenomena with the ability to generate robust predictions.[27,30] The roots of this may lie in the dominance of hypothetico-deductive modeling (falsification) and inferential hypothesis testing,[45] which is core to statistical training in nursing and health sciences,[46] and tends to equate the two. Predictive modeling is characterized by some notable elements summarized in Table 1.

Overall, three main types of models for prediction can be identified: regression, classification, and neural networks.[27] Here we focus on regression-based statistical techniques as they are mostly widely used in oncology nursing. Those interested in AI approaches to machine learning should refer to recent and upcoming EQUATOR guidance.[47-49]

A Guide Good Practice in Predictive Risk Model Development

A series of landmark articles recommended three main steps that should be completed before PRMs are used in routine clinical practice: (1) developing the model, (2) validating statistical performance, and (3) evaluating clinical performance.[7,9,23,37,38,50,51] Criteria for reporting the development of such models was further established by the EQUATOR network provided by the transparent reporting of multivariable prediction model for individual prognosis or diagnosis (TRPOID) statement[37,51] and the American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine established their own endorsement criteria.[43] This article focuses on the *development phase* including two distinct stages *preparation for modeling* describing three key processes and *analysis and modeling* describing six key processes (Table 2). Some of the processes are iterative and overlap,

TABLE 1
Six Key Characteristics of Predictive Risk Models

Characteristic	Explanation
Longitudinal data	In predictive modeling use of longitudinal data is required, ie, where the outcome occurs at a later calendar date that the predictors. Models are constructed for predicting outcomes based on predictors available a baseline. The number of individuals in the dataset is of key importance in order to be able to fit a suitable model, but it should be noted that standard power calculations for determining sample size are not appropriate.
Pragmatic considerations are important	Models need to be acceptable to patients and usable in clinical practice with a feasible number of predictors that have the potential to be easily collected (eg, between 2 and 20).
Bias-variance tradeoff	In explanatory modeling the focus is on minimizing bias to obtain accurate representation of the underlying theory[30]. In comparison, predictive modeling seeks to minimize the combination of bias and estimation variance, occasionally sacrificing the accuracy for improved precision and usefulness in practice.
Avoid overfitting the model to the data	Models should avoid overfitting, meaning where data closely follow a set of data points, this helps with generalizability. Typically, models are not applicable to new patient samples, even in similar populations or when the setting is very similar to the original setting and this is often because the original model has been fitted to reflect the idiosyncrasies in the data. The impact of overfitting can be reduced with modern statistical techniques (see Process 6)
Calibration and discrimination are important	Calibration refers to whether the models predicted probability is in line with observed outcomes and discrimination refers to the model's ability to correctly distinguish those with/without outcome. Both should be assessed. For PRMs to be used in oncology nursing practice they need to be sufficiently reliable. This means that there should be concordance between the predicted and observed outcomes, and for example, the PRM discriminates between those who are higher- and lower-risk patients.
Model validation is vital	Validation matters in a similar way to the development and testing of questionnaires or patient-reported outcome measures used in oncology nursing practice. Before a predictive model is used to inform decision-making, it should be validated, at least using so-called internal validation (assessing performance in a bootstrap sample, using cross-validation or split-sample (holdout) validation, see Process 9). Even better, an external sample or pilot, and evaluation of clinical performance in real-world practice. We have found a void in current psychosocial oncology literature evaluating performance using either resampling or in new samples.

TABLE 2
Summary of Guide to Good Practice for Predictive Risk Model Development

Processes (and subprocess, where relevant)	Details
Stage 1: Preparation for modeling	
1. *Choosing and defining an outcome*	Identify a clinically relevant and patient focused outcome to be predicted. Define its key parameters including suitable methods of valid and reliable measurement, type of outcome (single/combined), and time horizon including duration, if relevant, the number and timing of measurement and timing of outcome occurrence.
2. *Reviewing current knowledge and clinical practice*	Used to identify potential predictors, this should include (but not be limited to): Undertaking a systematic literature review and/or meta-analysis; Stakeholder and patient involvement.
3. *Assessing the data quality*	Data should be longitudinal and comprehensive. Sample size is important.
3.1 Longitudinal design and data sources	PRM should be developed using longitudinal (prospective or retrospective) designs only using either original research or secondary analysis. The sample includes people at risk of developing outcome of interest. PRM includes multiple candidate predictors measured at an appropriate baseline. Outcome(s) are measured after a sufficient interval (time horizons). Measures should be valid/reliable for target population.
3.2 Candidate predictors	Candidate predictors/outcome(s) can be interval, ordinal, or nominal. When developing models the conversion of continuous predictors/outcome(s) (eg, age) to categorical (eg, age group) should be avoided.
3.3 Sample size	An appropriate sample size is required in terms of the number of participants relative to the number of candidate predictors.
3.4 Establish type of multivariable regression method	The outcome of interest will determine the type of regression model to use. Typical models include linear (for continuous outcomes) logistic (for binary outcomes) and Cox (time to event).
Stage 2: Analysis and modeling	
4. *Assess data quality*	Examine descriptive statistics and characteristics of the sample. Explore amount/patterns of missingness. Failure to assess this may result in biased prediction models.
5. *Handle missing data*	Missing data can occur at different levels including for the baseline or outcome, or at item and scale level. If simultaneously missing data is ≥5% avoid case-wise deletion, mean or single imputation, which reduce data and increase bias. Often MICE sensitivity analysis will be the best approach. MICE regressions include all predictors of missing data and any predictors/outcome(s) from final analysis model. MICE rule of thumb: impute the number of datasets equivalent to the percentage of missing data.
6. *Predictor selection during modeling*	Prior to modeling this is informed by: Systematic review +/− meta analysis; Expert opinion/clinical practice/patient public involvement; Data quality/availability. Avoid selection based on significant *P* values. During modeling: Strong evidence base or specific research question: test full model (no selection) Exploratory research and many candidate predictors: consider penalized regressions. If selection of predictors is necessary, *avoid* selection using stepwise regression or *P* value screening.
7. *Assessing model performance*	Assess overall performance for example using R^2 (to describe the variance explained) and residual vs fitted plots (assessment of bias). Assess discrimination (ability to correctly distinguish those with/without outcome) and concordance statistics. Assess calibration (predicted probability is in line with observed outcomes) for example using RMSE (the SD of the residuals), calibration plots and the area under the ROC. Assess clinical utility and cost/benefits of using the model to inform practice with DCA.
8. *Provide a clear rationale if using risk grouping*	If using, model may be simplified, and appropriate cutoffs determined for risk grouping. Caution should be used in applying arbitrary cutoffs as this may lead to suboptimal care. The rationale for any risk grouping should be clearly described.
9. *Undertake internal model validation*	Use split samples, cross-validation or Bootstrap validation to avoid overfitting (ie, over-optimistic model performance). Bootstrapping is often appropriate in psychosocial oncology and supportive care research. Use the TRIPOD statement to report model development.

although for ease of explanation they are presented sequentially here.

Preparation for Modeling

Choosing and Defining an Outcome

Predictive risk models estimate the probability of an outcome[51] and so the first step involves not only identifying an outcome to be predicted but also defining its key parameters of interest. For example, for a researcher interested in long-term neutropenic symptoms this could include occurrence (classified as present/absent), severity of symptoms (rated on a PRO scale), progression, or change in symptoms over time (repeated-measures), or the absolute number of discrete symptoms (counted). All may be important depending on the specific research question, condition, domain, or time-horizon of interest. Choosing the most appropriate outcome will help identify potential data sources and determine modeling strategies. In psychosocial oncology the parameters of outcomes should be both clinically meaningful and patient focused (see Process 2) and will likely be informed by pragmatic concerns and local clinical context.

Other important parameters to consider when choosing an outcome are the availability of valid and reliable measurements including considering administration/collection mode, type of outcome (single/combined), duration, number, and the timing of measurements of outcome.[52] Linked to this is the need to clearly define clinically meaningful time horizon from baseline (time origin) to the eventual outcome prediction time point.[42]

To develop powerful PRMs, wherever possible researchers should avoid the temptation to apply dichotomies without careful consideration and/or arbitrary cutoffs of scales. In prediction research there is a strong tendency to classify outcomes so that individuals are “a

case" or "not a case". Such approaches to dichotomization, requiring use of logistic regression models to analyze these data, should generally be avoided in the development phase if the outcome is not a discrete event in order to avoid residual confounding (which occurs when there is inadequately adjustment for a variable in a model).[40,53]

The emergence of digitally assisted data collection techniques in practice means it is no longer necessary for researchers to impose arbitrary "cutoffs" in model development, where a score above an arbitrary threshold is determined to be a case. Instead, if patient or clinician usability/acceptability requires risk grouping, appropriate thresholds should be determined through implementation studies if risk grouping is necessary after development phase.

Reviewing Current Knowledge and Clinical Practice

To fully understand the context in which the model will be used and assist with some modeling decisions such as selecting candidate predictors, a thorough review of the literature, and investigation of current practice and patient perspectives should be undertaken.

The identification of candidate predictor variables[54] should be achieved through an in-depth review of current evidence and assessment of clinical practice including grey literature and local and/or national policies and provision. Reflecting the scope of oncology nursing practice, candidate predictors are unlikely to involve only patients' biomedical characteristics (tumor type and grade, metastases) but will typically span biopsychosocial domains of health.[18,55] The formal identification of candidate predictors should be informed by a risk prediction systematic review,[54] which uses specific methodologies focused on identifying and evaluating the strength of evidence from studies with longitudinal designs and multivariable analysis to distinguish the variables most strongly associated with an outcome of interest.[29,52,56,57] When evaluating the evidence base for statistical models it is important to assess original research against quality standards. The Quality in Prognosis Studies (QUIPS) tool provides a useful basis by which to systematically assess quality of previous research and thereby the strength of evidence.[39,58]

The involvement of people affected by cancer, cancer care professionals, and other stakeholders is also an essential component in developing models that are relevant and usable in practice.[8] Patient and public involvement (PPI) in the development of PRMs should be viewed as a priority[38] and reported according to best practice.[59,60] This particularly important for PRMs designed to inform supportive cancer care as acceptability (patient and clinicians subjective views) and feasibility (for example, ease of use and assessment in clinical setting) are often overlooked but are crucial if a PRM is ever to be implemented in routine clinical practice.[8] Similarly high-quality PPI is likely to becoming increasingly important in light of issues around ethical issues about the use of data, regulation and legislation, such as General Data Protection Regulation in the European Union. More formal techniques for establishing clinical consensus such as focus groups, expert surveys or modified Delphi,[61] may be required where evidence is unclear or original research is lacking.

Assessing the Data Quality

In this section, we detail three key subprocesses useful in determining high-quality data sources for predictive models. In the age of open science and open data, this may not only include original comprehensive observational research[9] designed specifically for the purpose of developing PRMs but increasingly secondary analysis of existing research data[62] and/or routinely collected data (eg, electronic health records) at a local or national level and/or data linkage through digital technologies[38]. However, some key challenges exist with use of routine data including losses to follow-up (missing data), errors or validity of coding/classifications, and ethical data protection issues around access, consent, and use of data.[63] As with any analysis it is important to be able to understand and describe the specific cancer diagnosis of the patients on which the PRM was developed and any inclusion/exclusion criteria used so users can understand the applicability of the model to individual patients.[43]

Longitudinal Design and Data Sources

It is common for psychosocial oncology studies to describe results as "predictive"when in reality they are correlational due to using a cross-sectional study design (even if using a regression-based modeling approach).[18] If we are to build powerful PRMs to inform the decision-making of people living with cancer and to guide their supportive care pathways, models should be built on longitudinal data. This could include prospective data collection, retrospective, or some combination of both. As longitudinal research is expensive, an acceptable and efficient solution is to use preexisting datasets for secondary analysis.[62] Ideally data would be observational (this could incorporate routine data and patient records).[9] Secondary analysis of trial data can be appropriate in psychosocial and supportive care research, but case control data should be used with caution because participants are selected based whether an outcome was achieved and so their risk profile may not reflect the target population.[42]

When designing a new study or secondary analysis, multiple candidate predictors need to be measured at an appropriate baseline and the outcome(s), time-horizon predicted, measured after a sufficient interval.[42,64] These time horizons will depend on the research question and intended purpose of the model being developed. For example, a PRM designed to predict health-related quality of life (HRQoL) after adjuvant chemotherapy for women with breast cancer might include baseline measurements (such as demographic, clinical characteristics, and baseline QoL) before or at the start of treatment plus time-dependent covariates (occurrence of symptoms/toxicities during the first 6 cycles of treatment), to predict HRQoL 3 months after treatment has completed. In this example, this time horizon might be chosen as a patient-focused and clinically relevant time point to coincide with the timing of the nurse led post-chemotherapy follow-up appointment.

Candidate Predictors

Data sources should include important candidate predictors (informed by Processes 1 and 2) and will often include routinely collected patient characteristics such as age, sex, comorbidities, and baseline measurement of outcome. Where secondary analysis is involved, it is likely that some important candidate predictors will be omitted from the data source and this should be acknowledged as a limitation. As with any study it is important to ensure that measurements are valid/reliable for the target population including those that are routinely collected.[38,64] Candidate predictors can be interval, ordinal or nominal but the transformation of variables from continuous to categories should generally be avoided to develop robust models and avoid residual confounding.[27,40,65] As with outcomes, in psychosocial oncology candidate predictors should be both clinically meaningful and their inclusion will likely be informed by similar pragmatic concerns and local clinical context. For example, in practice PRMs are unlikely to be adopted if the baseline measurement required is difficult to measure or requires additional resource to collect.[8]

Sample Size

To develop a robust PRM and avoid overfitting the model to the data, an appropriate sample size is required in terms of the number of participants relative to the number of candidate predictors.[50] In this context it is not widely understood that traditional "power calculations" are inappropriate because there is no hypothesis test.[50] For

binary (categorical) or time-to-event outcomes in the past the general *rule of thumb* for sample size was 10 events (ie, cases) per candidate predictor in the model.[54] However, this is now understood to be more complex,[66-68] in certain circumstances this could be lower[69] and more recent publications have shown the number is often likely to be much higher.[50,70,71] There are similar and complex considerations for linear regression models in which traditional approaches to sample size calculation often underestimate numbers required and will fail to achieve necessary precision.[72] Specifically, researchers will often underestimate the sample size required in the presence of categorical predictors, which are common in psychosocial and support care research.[72] In certain circumstances the effects of a low events-per-predictor may be mitigated with modern regression shrinkage techniques (discussed later) but their application requires careful consideration. Therefore, our advice is that sample size rule of thumbs should be avoided. We recommend using Riley et al's[50] four-step guidance on calculating sample sizes for PRMs and that expert advice should always be sought when considering sample sizes for prediction, to ensure precise estimates of predictors, avoid an overly optimistic model, and reduce overfitting.

Establish Type of Multivariable Regression Method

The outcome of interest will determine the type of regression model to use. For continuous outcomes, a linear regression model is often developed to predict an outcome value based on the values of multiple predictors, which can be continuous, dichotomous, or categorical.[72] For binary outcomes logistic regression models are developed to predict an event as present or absent conditional on values of multiple predictors. Other classes of regression used in oncology research include Cox's proportional hazard regression models (the time for the event [outcome] to occur) and less commonly Poisson regression models (for counted outcomes data) and polynomial (when outcomes are continuous but associations are not assumed to be linear). The key message is that the type of model will be determined by the research question, choice of outcome, and the data source.

Advances in modern computing and data science mean that researchers can also consider using newer regression techniques that may be more appropriate for developing generalizable PRMs. Penalized (shrinkage) regressions such as Ridge,[73,74] LASSO (least absolute shrinkage and selection operator),[75] or Elastic Nets[76] all introduce a penalty term into the regression and can help reduce overfitting. These approaches may particularly useful for; identifying predictors of "rarer" psychological outcomes or symptom profiles in cases of low events per predictor,[77] when there is multicollinearity of predictors and situations in which parsimonious models may be more practical and variable selection is required.[78]

Prior to commencing analysis, a statistical analysis plan should be written, preferably in consultation with a medical statistician, to avoid data dredging (where large values of data are analyzed and any spurious associations are presented as important)[79] and deviations from this plan should be noted in any publications.[80] In some cases, including PRMs using data linkage and big data, the involvement of data managers and computer and/or data scientists will also be important.

Analysis and Modeling

Assess Data Quality

As with all analyses, is important to first use descriptive statistics to get to know the data and describe the sample, if possible, compared with known population estimates.[51] This is important because systematic biases in data may lead to biased models. For example, in our experience in psychosocial oncology people from black and minority ethnic communities are often under-represented in datasets[81] and so in assessing the utility we should have an awareness of inherent biases with respect to the individual.

Many research and clinical datasets have missing data and so it is essential to explore the amount, patterns, and mechanism of missingness.[82-86] It is often unrealistic to assume that the values for missing data for patients are the same as those with complete data, particularly in cancer, when patients often experience multiple-morbidities alongside social, economic, and psychological challenges. Failure to analyze missing data may result in biased prediction models[82] and many software packages readily support exploration of missing data with built in functions to assist with this. The quality of data may affect the ability to include candidate predictors, for example if a potential predictor includes a high amount of missing data or lacks variability (floor–ceiling effects).[40] When this is the case, it should always be acknowledged as a methodological limitation.

Handle Missing Data

The best approach to missing data is to try to avoid it in the first place, through better design and data collection methods; however, missing data are an inevitable part of health research or clinical datasets.[82,84-86] Furthermore, as much research in psychosocial oncology involves PROs and other measurement scales derived from self-report questionnaires, data can be missing in many ways, including for outcomes and predictors, and at the item level or for complete questionnaires or scales.[82] For PRO and questionnaires, original authors' guidelines should be consulted to handle item-level missingness, however, we have found that some widely used measures, such as the Hospital Anxiety and Depression Scale[87] provide no guidance on handling missing data and specialist statistical studies may need to be consulted where they exist.[83]

Unfortunately, in psychosocial oncology model building typically relies on only patients with complete data, so called case-wise deletion, even when more than 5% of data is simultaneously missing.[18] This should generally be avoided, as should mean or single imputation which reduce efficiency of data use and increase bias[88] as these methods assume data is missing completely at random which is often unrealistic. There are alternative approaches that make less strong assumptions for the missing data. One of these is multiple imputation using chained equations (MICE),[89,90] which makes a missing at random assumption and provides a flexible approach to missing data. This can be implemented using statistical software (including SPSS, SAS, Stata, and R) and is well suited to longitudinal data generated using patient-reported outcomes.

In many studies model-building procedures, such as variable selection, is undertaken on complete-case datasets, even when multiple imputation (MI) has been used. This is an emerging field of research but generally this approach should be avoided as the results can be biased and lack power.[90,91] Alternative approaches such as averaging the results over all imputed datasets are available but are complex to implement and would require expert statistical support. Ultimately, we suggest that where the proportion of missing predictors or outcomes becomes too large (for example, >40%)[92] this may prohibit meaningful model development and anything more than exploratory analysis.

Predictor Selection During Modeling

If there are too many candidate predictors it will be necessary to reduce the number to those with the most predictive utility. This process may seem strange to those more familiar with explanatory modeling but emphasizes the pragmatic characteristic of predictive modeling favoring usability and parsimony (Table 1). Therefore if a PRM is to be used in everyday oncology nursing practice, it will often be necessary to reduce the number of candidate predictors from a long list (eg, >20) to a more manageable number determined by PPI

and stakeholder engagement to reduce burden for both staff and patients. As described, predictor selection prior to modeling is important and should be informed by Processes 1, 2, and 5. However, often further reduction is required to ensure the PRM is reliable, generalizable, and usable.[8,40,93] A commonly used but flawed approach to variable selection in this context is stepwise regression, when inclusion of variables in a model is based on some predetermined criteria such as *P* values below a threshold, and less commonly Akaike information criterion (AIC) or Bayesian information criterion (BIC). However, over the last 2 decades the arguments against its use have grown and stepwise regression is not recommended by many leading statisticians despite its continued use.[40,94] See Harrell[40] for an excellent overview of the problems with stepwise regression.

Often univariate regression screening has been used to determine potential predictors, as indicated by *F* test or α above a certain threshold (eg, 0.1).[27] However, for the purposes of predictive as opposed to explanatory modeling, this has been criticized as just another form of stepwise selection "through the backdoor" leading to over-fitted estimates and biased error.[27] Part of the problem is that this approach ignores the fact that predictors do not exist in isolation from each other. Another problem with such approaches is that predictors shown to be important in previous research may not be included if they fail to reach thresholds (perhaps affecting the acceptability and generalizability of the model).[42]

Although univariate screening may be a reasonable approach when the purpose of the modeling is explanatory, for predictive modeling modern regularization techniques provide a more appropriate and powerful approach. In summary these approaches help to minimize prediction error and overfitting of data by reducing the coefficients by introducing a penalty term to the regression. An advantageous technique to be aware of is LASSO,[75] which introduces a penalty term equal to the sum of absolute coefficient, meaning all coefficients are reduced and some reduced to zero. When used for variable selection, variables with non-zero coefficients after the shrinkage process are selected to be part of the model[78] and those with values of zero are effectively dropped. This helps develop more parsimonious models, can also afford a higher events-to-predictor ratio (eg, $\leq$5) and overcome the limitations associated with highly correlated variables, which in our experience can be common in research involving patient-reported outcomes and QoL measures (eg, high variance and beta values). However, even if using shrinkage techniques such as LASSO, those candidate predictors with a strong case for inclusion as suggested by Process 2 (reviewing current knowledge and clinical practice) should be included in the model.[42] New approaches by which to apply these techniques to MI data are being explored[95] but require expertise statistical support to implement.

Interactions between potential predictors should be explored but should be prespecified if analysis is exploratory or limited to those with evidence from wider literature to avoid spurious associations. Although often, for ease of use in clinical practice, interactions tend to be excluded from final models (because models assume effects of all predictors are additive)[27] with the increasing development and availability of online/mobile tools for calculating risk, the use of interaction terms will become increasingly feasible and so should be considered if they improve model performance.

Assessing Model Performance

It is important that the assumptions and appropriateness of the PRM are examined. This can be undertaken by defining residuals (ie, the error in the model) and examining residual plots,[96] and performance can be assessed for discrimination (ability to correctly distinguish those with/without outcome) and calibration (predicted probability is in line with observed outcomes). In development studies the focus would usually be on discrimination, because, by definition, models will usually be well calibrated on the dataset in which they are developed.[51] Specific assessment approaches will be determined by the model type. For example, linear regression's overall performance will often be assessed using R^2 to establish the variance explained by the model.[51,97] Calibration can be assessed using root mean square error (RMSE; difference between values predicted by the model and the observed values) and visually using locally weighted scatterplot smoothing and other calibration plots.[27]

For logistic regression models, discrimination can be assessed with area under the receiver operating curve (ROC; similar to C index for binary outcomes), pseudo R^2, or equivalent tests.[29] It can be useful to determine the model's overall correct classification percentage, sensitivity (true-positive rate), specificity (true-negative rate), positive predictive values, and negative predictive values. Calibration can be assessed by plotting the observed proportion of events against the predicted risk by groups defined by ranges of individual predicted risks and using goodness-of-fit Hosmer–Lemeshow or equivalent tests.[29,51]

It is worth noting that although these traditional performance statistics assess accuracy, they provide no information the consequences of using the model in practice. Newer techniques such as decision curve analysis (DCA) can be used to assess the potential benefits and costs of using the model, by evaluating the clinical consequence of using the model to intervene or not.[98,99] Importantly DCA provides an estimate of net benefit, which combines the number of true-positives and false-positives, weighted by a factor of false-positives relative to false-negatives, into a net number (similar to net profit) in which the larger the number the better the model. This is plotted against a range of probability thresholds, which is important because it is possible to consider patient preferences for risk. Therefore, if a PRM is ever planned to be used in clinical decision-making, an assessment of DCA is recommended although not yet common in practice.[51,97] For an introduction to the basic principles of DCA see Vickers et al.[100]

Provide a Clear Rationale If Using Risk Grouping

Predictive risk models are often used to identify risk groupings, individuals at low, moderate, or high risk of an outcome, through calculation of risk score/algorithm or index. At this stage the model can be simplified and appropriate cutoffs determined based on best practice regarding predictive accuracy and clinical utility.[40] Such approaches are tempting when using PRO measures (eg, cancer patients referred for support group if they score above a certain threshold), however, researchers should be cautious because a cutoff can quickly become standard practice without any clear rationale. Further in practice it is often assumed that individuals within each group have equivalent risk but in reality, there can be considerable diversity, which can lead to suboptimal care not in the best interests of patients.[51] It is therefore important to provide details on how the risk groups are defined with rationales to avoid arbitrary decisions.[51]

Undertake Internal Model Validation

The process in the model development phase involves internal validation using cross-validation, splitting the sample or holdout samples, or bootstrap validation to avoid overfitting. For smaller studies (n<1000) typical in psychosocial oncology, we suggest using bootstrapping as it attempts to account for model overfitting or uncertainty in the entire model development process by generating a new sample of data from the original sample. Hence with limited sample size bootstrapping can be superior to other approaches as it uses *all* available data (unlike cross-validation or split-sample validation).[51] The importance of this often-overlooked process is to estimate the bias in the model and therefore "test" its predictive power in a "new" sample.

All PRM development and internal validation should be reported in line with established guidelines provided by the transparent

reporting of multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.[51]

After the Model Development Phase: External Validation and Evaluating Clinical Performance

PRMs tend to perform less well in completely new samples and can perform differently in new clinical contexts and populations.[27] Testing models in new samples, known as external validation, allows for generalizability to be fully understood and test adaptations/extensions to an existing model. We are not aware of any such studies for applied PRMs in psychosocial and supportive oncology research. Similarly, there is a lack of research evaluation of PRM clinical performance and implementation in practice.[27,44] This is important because even when models are well developed, they might still not lead to improvements in patient outcomes and experience,[101] and so it is important that future research evaluates their efficacy and effectiveness rather than assumes they will lead to improvements. We are unaware of any impact studies (evaluating PRM use in clinical practice) in psychosocial oncology.

Discussion

We have collated and presented a succinct summary of best practice research for PRMs to assist researchers and help them navigate the vast and sometimes conflicting literature available. Although these recommendations for best practice in developing predictive risk models are available there appears to be relatively little application within psychosocial oncology and supportive care. If high-quality PRMs are to be safely deployed as part of routine practice, there needs to be greater awareness of the quality standards and our framework aims to raise awareness of some of the key processes for best practice in predictive risk model development in the context of oncology nursing.

Over the next decade the delivery of cancer care, particularly in follow-up and survivorship phases, will be reoriented away from hospital settings, toward the community and stratified follow-up approaches for people with the most common cancer,[102] and this is likely to accelerate post-COVID-19. Follow-up care will increasingly comprise remote monitoring and personalized supported self-management for a large (low risk of recurrence) majority. The use of risk prediction in oncology nursing practice offers exciting opportunities that are yet to be realized, to offer patients more personalized, responsive, supportive cancer care. However, PRM use presents some intrinsically linked key challenges where oncology nurses are particularly well placed to lead the research agenda and practice debate including:

- *Ensuring PRM are developed for areas that are important to people affected by cancer*. Oncology nurses' holistic approach to care means they are well-placed to inform the research agenda by ensuring models are predicting outcomes that truly make difference to patients' quality of life and well-being into survivorship.
- *Good communication and understanding of PRMs*. Predicted probabilities and percentages are often misunderstood[103-105] even amongst health professionals.[106] Oncology nurses have skills and training in advanced communication skills and are particularly well-placed to develop and deliver strategies to ensure PRMs are used in a way that can be widely understood and to allow for genuine choice and patient-preference in decision-making regarding psychosocial and supportive care options.
- *Ethical care as the focus of PRMs.* Understanding how and when PRMs can be used in an ethical way is a key challenge. As with psychosocial screening,[24,107] PRMs will only be helpful if services have clear pathways for care and self-management advice that are fit for purpose to meet patients' needs, particularly of those identified as "higher risk".

PRMs have the potential to be used to complement clinical practice and could be linked with existing holistic needs assessment and digital systems for symptom monitoring. In the future routinely collected digital patient-reported predictors and outcomes could be used to facilitate the use of real-time PRMs in practice to enhance the quality of life through stratified supportive care packages.

However, the failure to develop predictive risk models using best evidence and deployment without proper testing can cause great harm. If properly developed and validated, predictive risk models have the potential to enhance supportive oncology care through harnessing routine digital collection of patient-reported outcomes and the targeting of interventions according to risk characteristics.

Funding

J.H. was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Disclosures

None.

Acknowledgments

We thank Independent Cancer Patients' Voice and KCL's Supportive Cancer Care group for their feedback on our research.

References

1. Legg MJ. What is psychosocial care and how can nurses better provide it to adult oncology patients. *Austr J Adv Nurs*. 2011;28:61
2. Fitch MI. Psychosocial management of patients with recurrent ovarian cancer: treating the whole patient to improve quality of life. *Semin Oncol Nur*. 2003;19:40–53
3. Mitchell AJ, Hussain N, Grainger L, Symonds P. Identification of patient-reported distress by clinical nurse specialists in routine oncology practice: a multicentre UK study. *Psychooncology*. 2011;20:1076–1083
4. Mitchell AJ, Lord K, Slattery J, Grainger L, Symonds P. How feasible is implementation of distress screening by cancer clinicians in routine clinical care. *Cancer*. 2012;118:6260–6269
5. Maguire R, Fox PA, McCann L, et al. The eSMART study protocol: a randomised controlled trial to evaluate electronic symptom management using the advanced symptom management system (ASyMS) remote technology for patients with cancer. *BMJ Open*. 2017;7: e015016
6. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA*. 2017;318:197–198
7. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10: e1001381
8. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606
9. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how. *BMJ*. 2009;338:b375
10. Hutchings HA, Evans BA, Fitzsimmons D, et al. Predictive risk stratification model: a progressive cluster-randomised trial in chronic conditions management (PRISMATIC) research protocol. *Trials*. 2013;14:301
11. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA*. 2016;316:533–534
12. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*. 2019;21:1708–1718
13. Haybittle J, Blamey R, Elston C, et al. A prognostic index in primary breast cancer. *Br J Cancer*. 1982;45:361
14. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat*. 1992;22:207–219
15. Olivotto IA, Bajdik CD, Ravdin PM, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol*. 2005;23:2716–2725
16. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19:980–991

17. Hippisley-Cox J, Coupland C. Development and validation of risk prediction equations to estimate survival in patients with colorectal cancer: cohort study. *BMJ*. 2017;357:j2497
18. Harris J, Cornelius V, Ream E, Cheevers K, Armes J. Anxiety after completion of treatment for early-stage breast cancer: a systematic review to identify candidate predictors and evaluate multivariable model development. *Suppor Care Cancer*. 2017;25:2321–2333
19. Kumar S, Rana ML, Verma K, et al. PrediQt-Cx: post treatment health related quality of life prediction model for cervical cancer patients. *PLoS One*. 2014;9:e89851
20. Beltran-Alacreu H, López-de-Uralde-Villanueva I, Calvo-Lobo C, et al. Prediction models of health-related quality of life in different neck pain conditions: a cross-sectional study. *Patient Prefer Adherence*. 2018;12:657–666
21. Révész D, van Kuijk SMJ, Mols F, et al. Development and internal validation of prediction models for colorectal cancer survivors to estimate the 1-year risk of low health-related quality of life in multiple domains. *BMC Med Inform Decis Mak*. 2020;20:54
22. Watson E, Rose PW, Neal RD, et al. Personalised cancer follow-up: risk stratification, needs assessment or both. *Br J Cancer*. 2012;106:1579–1580
23. Hingorani AD, van der Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ*. 2013:346:e5793.
24. Heathcote LC, Goldberg DS, Eccleston C, et al. Advancing shared decision making for symptom monitoring in people living beyond cancer. *Lancet Oncol*. 2018;19:e556–e563
25. LeBlanc TW. Shared Decision-making in Acute Myeloid Leukemia. *Semin Oncol Nurs*. 2019;35: 150958
26. Taylor C, Finnegan-John J, Green JS. No decision about me without me" in the context of cancer multidisciplinary team meetings: a qualitative interview study. *BMC Health Serv Res*. 2014;14:488
27. Steyerberg EW. *Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating*. New York: Springer Science & Business Media; 2008
28. de Leeuw J, Larsson M. Nurse-led follow-up care for cancer patients: what is known and what is needed. *Support Care Cancer*. 2013;21:2643–2649
29. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604
30. Shmueli G. To explain or to predict. *Stat Sci*. 2010;25:289–310
31. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect Psychol Sci*. 2017;12:1100–1122
32. Wynants L, Van Calster B, Bonten MM, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328
33. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. *Obstetr Gynecol*. 2010;116:1018–1019
34. Chan MK, Bhatti H, Meader N, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. *Br J Psychiatry*. 2016;209:277–283
35. Mulder R, Newton-Howes G, Coid JW. The futility of risk prediction in psychiatry. *Br J Psychiatry*. 2016;209:271–272
36. Gillingham P. Predictive risk modelling to prevent child maltreatment: insights and implications from Aotearoa/New Zealand. *J Public Child Welf*. 2017;11:150–165
37. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med*. 2015;350:g7594
38. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. 2013;346:e5595.
39. Hayden JA, van der Windt DA, Cartwright JL, CÃ P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158:280–286
40. Harrell F. *Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. New York: Springer; 2015
41. Kuhn M, Johnson K. *Applied Predictive Modeling*. 26. New York: Springer; 2013
42. Kattan MW, Gerds TA. A framework for the evaluation of statistical prediction models. *Chest*. 2020;158:S29–S38
43. Kattan MW, Hess KR, Amin MB, et al. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin*. 2016;66:370–374
44. Vickers AJ. Prediction models in cancer care. *CA Cancer J Clin*. 2011;61:315–326
45. Hempel C, Oppenheim P. Studies in the logic of explanation. *Philos Sci*. 1948;15:135–175
46. Purssell E, While A. *P* = nothing, or why we should not teach healthcare students about statistics. *Nurs Educ Today*. 2011;31:837–840
47. STARD-AI extension: reporting guidelines for diagnostic accuracy studies evaluating artificial intelligence interventions. Available at: https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-other-study-designs/#STARD-AI-Abs. (Accessed September 12, 2020).
48. STARD-AI for abstracts extension: reporting guidelines for diagnostic accuracy studies evaluating artificial intelligence interventions. Available at: https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-other-study-designs/#STARD-AI-Abs . Accessed Sept 12, 2020.
49. Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020;370:m3210
50. Riley RD, Ensor J, Snell KI, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441
51. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–W73
52. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11: e1001744
53. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25:127–141
54. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. 2016:353:i3235.
55. Bours MJ, van der Linden BW, Winkels RM, et al. Candidate predictors of health-related quality of life of colorectal cancer survivors: a systematic review. *Oncologist*. 2016;21:433–452
56. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med*. 2006;144:427–437
57. Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. *J Clin Epidemiol*. 2009;62:781–796.e781
58. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *Bmj*. 2015;350:h870
59. Staniszewska S, Brett J, Mockford C, Barber R. The GRIPP checklist: strengthening the quality of patient and public involvement reporting in research. *Int J Technol Assess Health Care*. 2011:391–399
60. Staniszewska S, Brett J, Simera I, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *Res Involv Engagem*. 2017;3:13
61. Murphy M, Black N, Lamping D, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998;2:1–88
62. Peat G, Riley RD, Croft P, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Med*. 2014;11: e1001671
63. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12: e1001885
64. Hendriksen J, Geersing G, Moons K, De Groot J. Diagnostic and prognostic prediction models. *J Thromb Haemost*. 2013;11:129–141
65. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med*. 2016;35:4124–4135
66. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. *J Clin Epidemiol*. 1995;48:1495–1501
67. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis: II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503–1510
68. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II—binary and time—to—event outcomes. *Stat Med*. 2019;38:1276–1296
69. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 2007;165:710–718
70. van Smeden M, de Groot JA, Moons KG, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol*. 2016;16:163
71. van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res*. 2019;28:2455–2474
72. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part I. Continuous outcomes. *Stat Med*. 2019;38:1262–1275
73. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970;12:55–67
74. Hoerl AE, Kennard RW. Ridge regression: applications to nonorthogonal problems. *Technometrics*. 1970;12:69–82
75. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996:267–288
76. Zou H, Hastie T. Regression shrinkage and selection via the elastic net, with applications to microarrays. *JR Stat Soc Series B Stat Methodol*. 2003;67:301–320
77. McNeish DM. Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivariate Behav Res*. 2015;50:471–484
78. Fonti V, Belitser E. Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics. 2017 Mar 30;30:1—25. Available at: https://beta.vu.nl/nl/Images/werkstuk-fonti_tcm235-836234.pdf (Accessed September 12, 2020).
79. Smith GD, Ebrahim S. Data dredging, bias, or confounding. *BMJ*. 2002;325:1437–1438
80. Hiemstra B, Keus F, Wetterslev J, Gluud C, van der Horst IC. DEBATE-statistical analysis plans for observational studies. *BMC Med Res Methodol*. 2019;19:233
81. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004;291:2720–2726
82. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. 2014;23:440–459
83. Bell ML, Fairclough DL, Fiero MH, Butow PN. Handling missing items in the Hospital Anxiety and Depression Scale (HADS): a simulation study. *BMC Res Notes*. 2016;9:479
84. Bell ML, Fiero M, Horton NJ, Hsu C-H. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol*. 2014;14:118
85. Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health Qual Life Outcomes*. 2009;7:57

86. Fielding S, Maclennan G, Cook JA, Ramsay CR. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials*. 2008;9:51
87. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67:361–370
88. Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW. Missing data: a systematic review of how they are reported and handled. *Epidemiology*. 2012;23:729–732
89. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatric Res*. 2011;20:40–49
90. White IR, Royston P, Wood A. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011:30
91. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data. *Stat Med*. 2008;27:3227–3246
92. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17. 162-162
93. Subramanian J, Simon R. Overfitting in prediction models—is it a problem only in high dimensions. *Contemp Clin Trials*. 2013;36:636–641
94. Flom PL, Cassell DL. Stopping stepwise: why stepwise and similar selection methods are bad, and what you should use. *Paper presented at: NorthEast SAS Users Group Inc 20th Annual Conference: 11-14th November 2007*. Baltimore,MD; 2007
95. Musoro JZ, Zwinderman AH, Puhan MA, Ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. *BMC Med Res Methodol*. 2014;14:116
96. Acock AC. *A gentle introduction to Stata*. College Station, TX: Stata Press; 2008
97. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21:128
98. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26:565–574
99. Vickers AJ, Cronin AM. Everything you always wanted to know about evaluating prediction models (but were too afraid to ask). *Urology*. 2010;76:1298–1301
100. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3:18
101. Connell A, Raine R, Martin P, et al. Implementation of a digitally enabled care pathway (part 1): impact on clinical outcomes and associated health care costs. *J Med Internet Res*. 2019;21:e13147
102. England N. Long term plan. Available at: www.england.nhs.uk/long-term-plan. (Accessed September 12, 2020).
103. Han PK, Klein WM, Lehman TC, Massett H, Lee SC, Freedman AN. Laypersons' responses to the communication of uncertainty regarding cancer risk estimates. *Med Decis Mak*. 2009;29:391–403
104. Han PKJ, Lehman TC, Massett H, Lee SJC, Klein WMP, Freedman AN. Conceptual problems in laypersons' understanding of individualized cancer risk: a qualitative study. *Health Expect*. 2009;12:4–17
105. Gigerenzer G, Galesic M. Why do single event probabilities confuse patients. *BMJ*. 2012;344:e245
106. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest*. 2007;8:53–96
107. Meijer A, Roseman M, Delisle VC, et al. Effects of screening for psychological distress on patient outcomes in cancer: a systematic review. *J Psychosom Res*. 2013;75:1–17