



Models for Predicting Melanoma Outcome

Lauren E. Haydu, Phyllis A. Gimotty, Daniel G. Coit,
John F. Thompson, and Jeffrey E. Gershenwald

Contents

Introduction	300
Prediction Tools and Statistical Models	300
Personalized Prognosis	301
Clinical Applications	301
The Link Between Prediction Tools and Staging Systems	301
The Relevance of Prediction Tools for Clinical Trials	301
Brief History of Melanoma Prediction Tools	302
Prediction Tools Developed from AJCC Databases	302
Other Prediction Tools	302

L. E. Haydu (✉)

Department of Surgical Oncology, The University of Texas
MD Anderson Cancer Center, Houston, TX, USA
e-mail: LEHaydu@mdanderson.org

P. A. Gimotty

Department of Biostatistics, Epidemiology and
Informatics, Perelman School of Medicine, The University
of Pennsylvania, Philadelphia, PA, USA

D. G. Coit

Department of Surgery, Memorial Sloan Kettering Cancer
Center, New York, NY, USA

J. F. Thompson

Melanoma Institute Australia, Faculty of Medicine and
Health, The University of Sydney, Sydney, NSW, Australia

Department of Melanoma and Surgical Oncology, Royal
Prince Alfred Hospital, Sydney, NSW, Australia

J. E. Gershenwald

Departments of Surgical Oncology and Cancer Biology,
Melanoma and Skin Center, The University of Texas MD
Anderson Cancer Center, Houston, TX, USA

Planning to Build a Prediction Model	302
Reporting Prediction Models	302
Criteria for Building Prediction Models	303
Selection of a Patient Population	303
Selection of an Outcome to Predict	305
Conditional Survival	306
Probability of Binary Outcome	306
Considering the Treatment Landscape	307
Selection of Relevant and Clearly Defined Predictors	307
Model Development	308
Cox Proportional Hazards Model	308
Hazard Function	309
Relative Risk	309
The Proportional Hazards Assumption	309
When the Proportional Hazards Assumption Is Violated	309
Model Validation and Performance	309
Internal Validation	310
External Validation	310
Putting Contemporary Models to the Test	311
Challenges and Opportunities	311
Sample Size	311
Selection Bias	311
Missing Data	312
Survival from Metastatic Melanoma	312
The Future	312
Cross-References	313
References	313

Abstract

Clinical and pathological features that impact melanoma patient survival have been studied extensively for decades at major melanoma centers around the world. With the aid of powerful statistical techniques and computational methods, remarkable progress has been made in the identification of dominant factors that are linked to the natural history of melanoma and associated clinical outcome. A wide array of clinical prediction tools have been promulgated, primarily focused on forecasting survival outcomes across the melanoma continuum, with the exception of distant metastatic (Stage IV) melanoma. Recent changes in melanoma clinical practice resulting from the availability of new targeted and immune therapies that are effective in both metastatic and adjuvant settings, as well as level I evidence demonstrating no

survival benefit for completion lymph node dissection after a positive sentinel lymph node biopsy, have together changed the melanoma landscape and will no doubt impact on approaches to outcome prediction. Against this contemporary and ever-evolving backdrop, we present clinical applications, criteria, challenges, and opportunities for interpreting and building tools for predicting melanoma outcomes.

Introduction

Prediction Tools and Statistical Models

In the field of clinical research, broadly, a prediction tool is the implementation of a statistical model into a readily usable format such as a nomogram, classification tree, or electronic calculator that forecasts patient

outcomes. The underlying statistical model for a given prediction tool is typically built using multiple regression methods that calculate the probability of patient survival (or hazard of death). Statistical models, in general, are built utilizing a training data set of independent variables, also known as predictors, together with known patient outcomes (representing the dependent variable, e.g., survival or sentinel lymph node positivity). A prognostic tool is a specific type of prediction tool that focuses on survival-based patient outcomes. Once a statistical model is developed, it is necessary to validate the model, which is the process by which the performance, or accuracy (of the predictions), of the model is assessed using an independent data set of predictors and patient outcomes.

Personalized Prognosis

The concept of personalized medicine, defined by the National Cancer Institute (NCI) at the National Institutes of Health (NIH) on their website (2018), involves the use of specific information about a patient's "genes, proteins, and environment to prevent, diagnose, plan treatment, find out how well treatment is working, or make a prognosis." It is not possible to exactly predict the outcome of an individual patient without a crystal ball; however, the extent to which we *approach* outcome prediction for an individual patient is the concept of personalized prognosis (i.e., individualized prognosis). The spectrum of outcome prediction ranges from a model that estimates a broad range of survival for a heterogeneous group of patients with melanoma, for example, to a model that estimates a more specific range of outcomes for a homogeneous group of patients (e.g., 50-year-old female patient presenting with 1.5 mm Breslow thickness melanoma and a negative sentinel lymph node biopsy). The latter end of the spectrum described is considered a *more* "personalized" or "individualized" prognosis compared with the former.

Clinical Applications

Prediction of the likely clinical course of a disease and the expected treatment outcome is an essential part of medical practice. Clinicians face daily decisions that include selection and recommendation of an "optimal" treatment for an individual patient. Follow-up evaluation strategies can vary according to a patient's prognosis. A prediction tool may generate a prognostic summary analysis of survival and disease recurrence rates for an individual patient based on presenting characteristics. Clinicians may use these projections in conjunction with other factors, including likelihood of treatment response and the morbidity or potential toxicity associated with treatment, to guide treatment decision-making and the frequency and duration of follow-up.

The Link Between Prediction Tools and Staging Systems

The importance of staging and classification of melanoma patients is described in chapter ▶ "Melanoma Prognosis and Staging." The staging system is necessarily structured in order for it to serve as a common global language of cancer classification; however, due at least in part to the de facto constrained (i.e., TNM-based) approach, it cannot incorporate some potentially important disease-specific prognostic factors, may not aggregate groups of patients solely according to risk, nor assign a more individualized prognosis. A prediction tool that incorporates factors such as patient age and sex, as well as other factors beyond those used in the TNM classification, may better facilitate personalized clinical decision-making at important treatment and surveillance decision points.

The Relevance of Prediction Tools for Clinical Trials

Tools that predict patient outcomes based on contemporary prognostic estimates derived from

historical, standard-of-care cohorts are useful during the planning of clinical trials to facilitate informed power calculations and to select patient eligibility criteria and stratification variables. For clinical trials not employing pre-randomization stratification or those stratifying patients according to inappropriate factors, a prediction tool can be used to determine whether any differences in outcome are the result of treatment or are caused by an imbalance in prognostic factors.

In analyzing clinical trials, treatments are often compared within various subgroups. These subgroups are mostly defined using combinations of known relevant prognostic factors. A well-known limitation of subgroup analysis is the difficulty in extending analysis beyond two or three variables unless a very large sample is available. In the analysis of survival data, this is further complicated by the possibility of varied censoring patterns and durations of patient follow-up within the different subgroups. No adequate statistical inferences can be drawn from “overstretched” subgroup analyses. As in pre-randomization stratification, a prediction tool can be used to classify patients into an adequate number of subgroups for analysis after completion of the study.

Brief History of Melanoma Prediction Tools

Dr. Seng-jaw Soong (1985) developed the first prognostic tool in the form of a scoring system for predicting survival outcome for patients with localized melanoma. The validity of the underlying statistical model was tested in independent data sets, and its high degree of predictability made it clinically useful. Next, Soong et al. (1992) developed generalized multivariable prognostic models for patients with localized melanoma to address survival at diagnosis, survival after a disease-free interval, and probability of recurrence. The models were developed by analyzing a combined database of 4568 patients from the Sydney Melanoma Unit (SMU) and the University of Alabama at Birmingham (UAB).

Prediction Tools Developed from AJCC Databases

In 2000, the American Joint Committee on Cancer (AJCC) assembled a melanoma database that contained details of 17,600 patients, allowing the development of a new, evidence-based sixth edition melanoma staging system, as well as new prognostic models at diagnosis for both localized and regional melanoma (Balch et al. 2001a, b; Soong et al. 2003). Multivariable analyses were later performed on a substantially enhanced AJCC melanoma database developed for revision of the AJCC staging system and for development of statistical models. The results of these analyses, used to update the prognostic tools, including model validation with an independent data set and parametric modeling, and including hazard function estimation for localized melanoma, were published in the previous (5th) edition of this book (Soong et al. 2009).

Other Prediction Tools

In addition to the pioneering prognostic modeling efforts of Soong and colleagues, as well as those initiated by the AJCC, multiple tools for predicting melanoma survival have been subsequently developed over the past two decades and published in the form of regression models, including coefficients or formulas, prognostic classification trees, nomograms or scoring systems, or as internet-based (online) electronic calculators (Table 1). These tools focused on endpoints of melanoma-specific survival (MSS), disease-specific survival (DSS), or overall survival (OS), and predicted outcomes for patients presenting with localized and/or locoregionally metastatic melanoma.

Planning to Build a Prediction Model

Reporting Prediction Models

Standardized recommended approaches have been developed for Transparent Reporting of multivariable prediction models for Individual

Table 1 Melanoma-specific (MSS) or overall survival (OS) prediction tools published in the prior two decades. (Adapted from Mahar et. al. 2016)

Year	First author	Population	Sample size	Dates of cohort	Survival outcome	Type of tool
2000	Cochran	Stage I–III	1042	1980–1990	MSS	Individualized risk formula
2007	Gimotty	Thin melanoma (i.e. Breslow thickness ≤ 1 mm)	26,921	1988–2001	MSS	Prognostic classification tree
2007	Gimotty	Thin melanoma (i.e. Breslow thickness ≤ 1 mm)	2389	1972–2001	MSS	Prognostic classification tree
2010	Soong	Stage I–II	14,760	Not reported	MSS	Online tool
2011	Michaelson	Stage I–III	Not reported	Not reported	MSS	Online tool
2012	Callender	Clinically negative lymph nodes; Breslow thickness ≥ 1 mm	2507	1997–2003	OS	Online tool
2013	Lyth	Thin melanoma (i.e. Breslow thickness ≤ 1 mm)	11,165	1990–2008	MSS	Prognostic subgroups
2014	Khosrotehrani	AJCC seventh edition stage IIIB and IIIC	494	2000–2011	MSS	Nomogram
2014	Maurichi	Thin melanoma (i.e. Breslow thickness ≤ 1 mm)	2243	1996–2004	OS	Nomogram

Prognosis or Diagnosis (TRIPOD), broadly across disease types (Collins et al. 2015). In their groundbreaking work, Collins et al. engaged a multi-disciplinary panel of methodologists, as well as healthcare and editorial professionals, to produce the TRIPOD statement, a checklist of 22 items (Table 2) required to be reported in order for prognostic or diagnostic tools to be transparently and properly published (Collins et al. 2015).

Criteria for Building Prediction Models

In 2016, as the work product of a workshop conducted by the Precision Medicine Core of the American Joint Committee on Cancer, Kattan et al. published criteria for endorsement of cancer prognostic models by the AJCC. More specifically addressing statistical methodologies for cancer prediction models, Kattan et al. (2016) described 13 inclusion and 3 exclusion criteria (Table 3) for potential endorsement by the AJCC of an individualized prognostic calculator for use in cancer clinical decision-making. As an extension of these published criteria, this section provides additional

guidance for the development and interpretation of a robust model for predicting melanoma patient outcomes.

Selection of a Patient Population

A useful prediction model provides estimates that are relevant to a specific patient population and selected outcome measure (e.g., melanoma-specific survival or overall survival) relevant to the clinical decision-making process. Multiple models, reflecting distinct patient populations and/or outcome measures, may be incorporated into electronic tools or calculators that provide options for the user to select the specific prediction model relevant to their patient. According to the TRIPOD statement, for each prediction model, it is necessary to report the source of data, participants, and predicted outcome measure that was the focus of the development and validation of the model. Such information includes whether the cohort was from a clinical trial, registry, or other source; specific study dates (i.e., start and end dates of patient accrual); the study

Table 2 Summary of the TRIPOD checklist items. (Adapted from Collins et al. 2015)

Item number	Section	Checklist item to include
1	Title	Development or validation model; target population; outcome to be predicted
2	Abstract	Summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions
3a	Introduction	Medical context and rationale
3b	Introduction	Objectives including whether development or validation model
4a	Methods – source of data	Study design and type of data source
4b	Methods – source of data	Key study dates; start/end accrual, follow-up
5a	Methods – participants	Study setting – number and type of centers
5b	Methods – participants	Eligibility criteria
5c	Methods – participants	Details of treatment(s) received
6a	Methods – outcome	Clearly define outcome predicted
6b	Methods – outcome	Any blind outcome assessment
7a	Methods – predictors	Define all predictors and how they are measured
7b	Methods – predictors	Any blind predictor assessment
8	Methods – sample size	How sample size determined
9	Methods – missing data	How missing data are handled
10a	Statistical methods	How predictors are handled in the model (development only)
10b	Statistical methods	Type of model, model-building procedure, and internal validation method (development only)
10c	Statistical methods	Describe how predictions were calculated (validation only)
10d	Statistical methods	Methods to assess model performance and/or model comparisons
10e	Statistical methods	Model updating (validation only)
11	Methods – risk groups	Details on creation of risk groups
12	Methods – development vs. validation	Differences from development model (validation only)
13a	Results – participants	Describe participants with and without outcome, and follow-up time
13b	Results – participants	Characteristics of participants including missing data
13c	Results – participants	Comparison with development data set (validation only)
14a	Results – model development	Number of participants and outcome events (development only)
14b	Results – model development	Unadjusted association between predictors and outcome (development only)
15a	Results – model specification	Full prediction model (development only)
15b	Results – model specification	How to use the prediction model (development only)
16	Results – model performance	Model performance and confidence intervals
17	Results – model updating	Model updating results (validation only)
18	Discussion – limitations	Limitations of the study
19a	Discussion – interpretation	Discuss results with reference to development model performance (validation only)
19b	Discussion – interpretation	Overall results interpretation
20	Discussion – implications	Potential clinical use
21	Supplementary information	Availability of supplementary resources
22	Funding	Source of funding and role of funders

Table 3 Summary of the AJCC prognostic model building criteria. (Adapted from Kattan et al. 2016)

Inclusion criteria
1. Overall survival, disease-specific survival, and disease-specific mortality are the outcomes to be predicted
2. Prediction model should address a clinically relevant question
3. Include relevant predictors, or explain why a relevant predictor was not included
4. Specify which patients were used to develop the model, and how patients were selected to validate the model
5. Assess the model for generalizability or external validation
6. Define the time-zero from which prognosis is assessed
7. Predictors are to be known at time-zero and well defined
8. Provide details of the model such that it can be implemented, or free access
9. Report a measure of discrimination
10. Report calibration curves
11. Validation should occur over a contemporary and relevant time frame and clinical practice setting
12. Indicate which initial treatment(s) is applied, if applicable
13. Development and/or validation of the model must be published in a peer-reviewed journal
Exclusion criteria
1. Substantial proportion of patients in the validation data set had no follow-up
2. No reporting of missing data in the validation data set
3. Small number of events in the validation data set

setting (e.g., general practice versus tertiary-referral center); and eligibility criteria and treatment details (Collins et al. 2015).

The treatment center and patient referral pathway can impact the mix of treatment options for any cancer patient. Since the type of treatment a patient receives is directly related to their outcome, it is important to be aware of treatment differences in any data set used for model building or validation. In particular, newly established guidelines or contemporary evidence may be more rapidly implemented in a tertiary-referral, specialist-laden cancer center. If patient data have been captured at a tertiary-referral center, review of their preceding treatment and diagnosis in the community will be needed in order to

confirm that the case is in line with the clinical management and diagnosis was reported as a prerequisite for the prediction tool.

Finally, for the use of international data sets, the issues outlined above can be exacerbated due to differences in clinical practice, with respect not only to treatment availability and decision-making but also, potentially, to differences in the standard conduct and availability of dermatology, pathology (e.g., specimen sectioning methods and standard immunohistochemistry staining practices), nuclear medicine, and imaging (e.g., availability of magnetic resonance imaging (MRI) or positron-emission tomography (PET)). It is therefore important to have an in-depth understanding of the clinical practice of each institution that contributes patient data for model building. It may be possible to create and use a new variable to describe and account for any such differences deemed to significantly impact the model.

Selection of an Outcome to Predict

As introduced above, whether the utility of the model is intended for decision-making to guide routine follow-up, specific treatment options, or clinical trial or public health planning, the type of prognostic estimates that the model generates needs to be designed with contemporary practice in mind. Across the melanoma continuum, the changing landscape of treatment options, including surgical, intralesional, and systemic therapies, has resulted in new priorities and points of decision-making, for which robust prognostic estimates are needed. Although AJCC endorsement of prediction models is currently limited to overall and disease-specific survival outcomes (Kattan et al. 2016), there is a need to create tools that predict time to melanoma relapse, either at any anatomic site or at a specific site (e.g., probability of melanoma brain metastasis within 3 years of initial diagnosis), as well as non-survival-based outcomes, such as the probability of tumor-involved non-sentinel lymph nodes following a positive sentinel lymph node biopsy (SLNB).

Time to Relapse After Initial Disease Management

Time-to-relapse analysis has many pseudonyms in the scientific literature, including, but not limited to, relapse-free survival, time to recurrence, recurrence-free survival, disease-free survival, progression-free survival, time to progression, etc. In addition, various ways in which these analyses are defined may lead to challenges in the comparison and meta-analysis of results across multiple studies for any given cohort. An initiative to standardize definitions for clinical trial endpoints is known as the DATECAN project, or Definitions for the Assessment of Time-to-event Endpoints in CANcer trials (Gourgou-Bourgade et al. 2015). For the purpose of building prediction models, utilizing established standard definitions and transparent reporting are both critical for the patient and/or clinician end user to be confident with their interpretation of the estimates generated for the patient's outcome. Published guidance for industry by the US Department of Health and Human Services (2007) provides a definition for disease-free survival as time to the first of either recurrence or death from any cause and notes that this endpoint may be useful for the assessment of adjuvant treatments (in the context of gaining approval for biologics) when the disease-free survival benefit outweighs any observed toxicity.

It is important to note that the standard definition of an event in any relapse-type survival analysis, such as disease-free, recurrence-free, or relapse-free survival, includes death from any cause as an event in addition to the relapse event. Care must be taken in the interpretation of these endpoints, because a perceived inflated risk of recurrence of disease may result, particularly in low-risk populations, in which death from other causes would be significantly more likely to occur. For example, survival analysis may not be optimal for estimating time to a specific disease relapse, or site of relapse, in older patients that have a higher likelihood of death from another cause (Berry et al. 2010). Other methods, such as the cumulative incidence function, may be more appropriate to estimate the proportion of patients that will experience a specific type of relapse (or any relapse of disease), while

considering death from any cause as a competing risk (Pintilie 2006; Kim 2007).

Conditional Survival

Melanoma survival models are useful for generating prognostic estimates from the time of initial diagnosis of the patient's primary melanoma presentation. This information is essential for accurate patient staging and to guide decision-making for initial clinical management. However, estimates that are generated based on clinical information at the time of initial diagnosis do not indicate how a patient's prognosis may change throughout follow-up. In this case, conditional survival estimates are more useful and are derived from modeling patient cohorts that have survived for specified periods of time (e.g., patients are entered into the model on the condition of surviving 1 year or any other predefined time point). For cohorts of patients that have survived for some specified time after diagnosis, the baseline predictors may no longer be significantly associated with these patient's future survival. Conditional survival models may elucidate subsets of these survivors who are still at significant risk, for example, of death or relapse, while defining those at lower risk and providing them with some relative reassurance (Collett 2015; Hosmer et al. 2007; Haydu et al. 2017; Hieke et al. 2015).

Probability of Binary Outcome

There is also interest and applicability of other, non-survival type multivariable regression models in melanoma, such as binary logistic regression (Hosmer and Lemeshow 2000). Probably the most investigated binary endpoint in melanoma is the histological status of sentinel lymph nodes (i.e., positive or negative) and, likewise, the histological status of non-sentinel lymph nodes (i.e., predicting whether a completion lymph node dissection (CLND) procedure will identify additional metastatic melanoma deposits in non-SLNs; Murali et al. 2010; Cadili et al. 2010). Since patients undergoing SLN biopsy, by definition, have no clinical evidence

of regional lymph node metastasis, a prediction model is very useful to determine the indication for the procedure for all patient groups, but particularly in lower-risk patient groups, such as those with a thin primary melanoma (≤ 1 mm).

Considering the Treatment Landscape

Clinical care milestones such as the introduction and routine uptake of SLNB in the 1990s (► [“Melanoma Prognosis and Staging”](#)), the FDA approval of BRAF inhibitors and immune checkpoint inhibitors beginning in 2011 (► [“Cytokines \(IL-2, IFN, GM-CSF, etc.\) Melanoma”](#) and ► [“Managing Checkpoint Inhibitor Symptoms and Toxicity for Metastatic Melanoma”](#)), and recent evidence indicating no overall survival benefit for CLND versus observation (Faries et al. 2017; Leiter et al. 2016) impact not only the decision-making process for melanoma patients and clinicians but also the relevance of underlying data sets that are relied upon for training and validation of prediction models. For example, the current clinical paradigm has shifted from completion lymph node dissection (CLND) to observation, as well as consideration of adjuvant treatment for SLNB-positive melanoma patients. In this new treatment paradigm, the probability of relapse for the SLNB-positive patient, and the factors that contribute to their elevated risk of relapse, will guide decisions about how they should be followed (e.g., frequency and extent of imaging) and whether and when to recommend potentially toxic and costly adjuvant therapy. Models incorporating the number of positive non-sentinel lymph nodes as a predictor of patient outcome are no longer relevant since these data are no longer collected.

Furthermore, these milestones are not fixed time points at which clinical practice changes but generally indicate the commencement of a phase of change management when considering the uptake of new procedures (e.g., SLNB). Other examples include the effective implementation of new clinical trial evidence for surgery or radiotherapy (e.g. CLND or adjuvant radiotherapy), and the on protocol, compassionate use, and standard-of-care phases of use of new

intralesional or systemic agents either in the adjuvant or therapeutic settings. For example, although SLNB was formally introduced in 1992 (Morton et al. 1992), due to the fact that it is a complex procedure requiring a multi-disciplinary team including surgeon, pathologist, and nuclear medicine clinician working in concert, there was significant evolution of technique throughout the mid- to late 1990s, depending on the treatment center (► [“Melanoma Prognosis and Staging”](#)). Furthermore, the indication for SLNB in specific patient groups, e.g., patients with thin primary melanomas (≤ 1 mm), varies to this day and is usually based on a clinician-dependent set of factors known to increase risk of positivity (e.g., young age, ulcerated tumors, increased mitotic rate, etc.).

Selection of Relevant and Clearly Defined Predictors

The a priori selection of independent predictors of outcome for inclusion in the development of a statistical model is critical to the performance of the implemented prediction tool. In general, the selection of relevant predictors to test should include those firmly established to have an impact on the patient outcome to be predicted, or for which there is a solid biological or theoretical basis for inclusion (► [“Models for Predicting Melanoma Outcome”](#)). In conjunction with this is the need for selected, tested predictors to be objectively measurable, and within reach of the intended users of the prediction tool, i.e., not obscure factors that are subjectively measured, costly, or otherwise prohibitive to obtain.

Predictors that may be out of reach of the intended patient or clinician user group for the tool might include, for example, a somatic mutation that is not yet clinically actionable as standard of care and not available on routine-targeted sequencing panels or by immunohistochemistry. For melanoma, a BRAF mutation would be considered within reach for Stage III and Stage IV melanoma patients (i.e., clinically actionable and

currently a standard targeted sequencing test, with a specific V600E BRAF mutation capable of being assayed by immunohistochemistry), whereas a p53 mutation would not be a generally available test at this time for a patient with any stage of melanoma. The AJCC staging and classification system also has this de facto constraint in order to enable a common and accessible framework within which cancer patients around the world are staged.

Finally, the availability of the predictor is important to ensure that sample sizes are sufficient to produce robust estimates of patient outcomes without the potential biases introduced by missing data, a topic addressed later. For melanoma, based on decades of evidence collected by the AJCC, and consistently demonstrated in the scientific literature to influence prognosis, the foundational elements of a survival prediction tool should include primary tumor features of Breslow thickness (mm) and ulceration status, with strong consideration of mitotic rate (per mm²); regional metastasis features, i.e., the presence or absence of in-transit, satellite, or microsatellite metastasis(es), number of tumor-involved regional lymph nodes, and whether metastatic lymph nodes were clinically detected or clinically occult; and distant metastasis features including the presence and anatomic site and serum lactate dehydrogenase (LDH) level (► “Models for Predicting Melanoma Outcome”). Additionally, patient sex, age, and primary tumor anatomic site have been demonstrated by Soong and others to impact survival and should also be tested for inclusion in prediction models for survival outcomes. After inclusion of the aforementioned predictors, any newly tested predictor should meet the criteria of either being more objectively or more readily measurable and explain additional variability related to the outcome.

Model Development

The development of a prediction model is also known as training the model, and typically involves using a statistical multiple regression

method to develop a best-fitting model using a designated training data set. The training data set contains a cohort of patients with observed outcomes (dependent variable, e.g., survival time) annotated with a corresponding set of potential predictors or independent variables (also known as covariates or biomarkers). During training, statistical methodology is employed to decide whether the inclusion of each independent variable contributes to explaining the variability of the dependent variable (also referred to as the primary outcome) and to select a final set of predictors. In survival analysis, also known as time-to-event modeling, the primary outcome event being modeled for OS is death from any cause, and for MSS (DSS) is death specifically from melanoma (disease). Patients alive at last follow-up assessment are censored in the OS analysis, as is the case in DSS survival analysis where patients deceased from other causes are also censored. There are many statistical modeling techniques that may be employed for time-to-event analysis, each with pros and cons. However, as noted in the prior edition of this book, and by others, the Cox proportional hazards model generally produces results consistent with other methods and is a standard for multivariable survival analysis under the assumption of proportional hazards (Collett 2015; Hosmer et al. 2007).

Cox Proportional Hazards Model

The most frequently employed statistical method for predicting survival in patients with localized or locoregionally metastatic melanoma is the proportional hazards model proposed by Cox (1972), also known as Cox regression. The introduction of Cox regression was the most important methodologic development in survival data analysis. It permits semi-parametric assessment of survival data and allows the statistical inference to be restricted to the relative effect of concomitant information (e.g., prognostic factors) without knowledge of the form of the survival distribution under the assumption of proportional hazards (discussed below). It allows for inclusion of multiple categorical or

continuous covariates (i.e., predictors) and models the time to a specific event (e.g., death). One rule of thumb for a Cox regression model is to include a *minimum* of ten events per covariate, although this may not guarantee a robust model. The Cox model can be expressed in terms of a hazard function or a survival function and is well suited to serve as a basis for developing prediction tools for melanoma (Soong 1985, 1992; Soong et al. 1992; Balch et al. 2001b; Soong et al. 2003).

Hazard Function

The hazard function at time t is defined as the instantaneous risk of death or failure at time t among those patients who are at risk (i.e., death or failure has not already occurred). It can be roughly interpreted as the rate of death or failure at a specific time. In the multifactor analysis of survival data, the hazard is often expressed as a function of the associated information related to patients' survival times.

Relative Risk

The relative risk (equivalently, a hazard ratio) is the ratio of the risk of death per unit of time for patients with a given set of characteristics (defined using predictors in the model) to the risk for patients in the reference group (i.e., patients with continuous predictors at their average values and categorical predictors equal to the reference category).

The Proportional Hazards Assumption

One important limitation of the Cox regression model is the inherent assumption that the influence of the predictors do not vary over time. There are multiple graphical methods available to assess whether a predictor violates the proportional hazards assumption, including two techniques readily accessible in most statistical software packages:

visual inspection of the log $[-\log(\text{Survival})]$ (i.e., log-cumulative hazard) plot, stratified by the groups of each predictor, to check for parallel curves; and visual inspection of a plot of Schoenfeld residuals plotted against survival time to look for nonrandom patterns (Hess 1995; Collett 2015).

When the Proportional Hazards Assumption Is Violated

When a predictor violates the proportional hazards assumption, two strategies can be employed that may allow use of the Cox regression model: (1) stratification of the model across the groups of the violating predictor and (2) entry of the predictor as a time-dependent or time-varying covariate. The former strategy is limited by sample size constraints, and the latter requires an observed value of the predictor at each point in time for every observation included in the model. An example of a time-varying covariate is the diagnosis with brain metastasis after initial diagnosis, since this event occurs at different times over the time period for each observation (e.g., patients who relapse in the brain would not be expected to relapse on the same date relative to their initial diagnosis). Knowledge of the date of brain metastasis allows for calculation of the interval from initial diagnosis to brain metastasis. This date is required to enter the brain metastasis predictor as a time-varying covariate in the model.

Model Validation and Performance

The model development phase, discussed above, involves testing the selected predictors using the appropriate statistical method (e.g., Cox regression) on the training data set, to determine the finite set of predictors and regression coefficients to be used in the final prediction formula. Model validation involves measuring the precision with which the final prediction formula produces estimates that match the observed outcomes in a validation data set. The most critical steps in implementing a prediction tool, in general, are

the internal and external validation of the model following initial development (i.e., model training).

Model performance is assessed by evaluating model discrimination and model calibration (Harrell et al. 1996; Hosmer et al. 2007; Alba et al. 2017), with most evaluation methods readily available in common statistical software packages. In general, model discrimination refers to the ability of the model to distinguish between patients having an event and those that did not, which can be evaluated using the receiver operating characteristic (ROC) curves, the area under the curve (AUC), or c-statistic. Model calibration or performance, also known as goodness-of-fit of the model, indicates the degree of accuracy of generated probabilities or prognostic estimates, i.e., an assessment of the difference between observed and predicted outcomes. A calibration curve, which is constructed by plotting observed and predicted outcome values for corresponding values of each predictor, is a visual method that provides insight as to which range of outcomes are being predicted well (overlapping curves) versus those that are not (divergent curves). The Hosmer-Lemeshow test statistic may be used to determine whether differences between observed versus predicted outcomes are statistically significant; however, it does not provide information as to the magnitude or potential subgroup of patients for whom the model does not provide good fit. Sometimes a prediction model fits very well in the training data but fails to predict well in the validation data set because of overfitting. Overfitting is defined as modeling that too closely approximates complex idiosyncrasies of a specific data set and is therefore not generalizable to other data sets. Conducting both internal and external validation helps avoid overfitting issues, by informing how well the model approximates outcomes across multiple and diverse data sets. In this context, it is possible to calculate a degree of optimism of the validation of a model, defined as the actual level of discrimination of the model subtracted from the apparent discrimination (Harrell et al. 1996; Hosmer et al. 2007; Alba et al. 2017).

The successful performance of a model across different settings, both temporally and

geographically, is indicative of a robust and generalizable prediction model. The publication of validation studies is important to provide clinicians and patients with context regarding the applicability of the model and the strengths and weaknesses of the predictions across varied settings, in order to guide appropriate interpretation of prediction or prognostic model estimates.

Internal Validation

Internal validation involves the reuse of the training data set to evaluate the discrimination of the model with respect to the precision of observed versus predicted outcomes. This requires that, *prior* to model training, an internal validation strategy is planned. Split-sample internal validation involves randomly splitting the development data set into two parts, one for training and one for validation. Generally, this method of internal validation produces overly optimistic measures of the model performance and poorer model performance in general, since the sample size of the training data set has been reduced to allow for a separate (smaller) validation data set. Bootstrapping, or bootstrap validation, is the preferred internal validation technique and utilizes readily available computational power to conduct iterative model validation based on a method of routine sampling of the training data set (Harrell et al. 1996).

External Validation

External validation, of which there are several types, involves the application of the prediction model to new data sets to further assess precision and generalizability assessing whether the developed model continues to predict outcomes in the new data set. The validation data set may differ from the development data set in many ways. Examples of validation data set types include *temporal* (e.g., data from a more contemporary data set from the same source), *geographical* (e.g., outcomes from a similarly acquired data set in a different institution or country), or *other*

types, such as generalizability across treatment settings (e.g., from a specialist center to a community hospital or registry data set or vice versa).

Putting Contemporary Models to the Test

Mahar et al. (2016) reviewed a set of contemporary prediction tools to determine whether the underlying statistical models were developed and validated using appropriate methodologies and suitable training and validation data sets. They conducted a systematic search and a subsequent, detailed, critical assessment of the statistical models corresponding to 17 identified melanoma prognostic tools. Their work reviewed the following key aspects: study cohort, sample size, dates of collection, duration of follow-up, number of events, outcome, final model covariates, and internal and external validation techniques (Mahar et al. 2016). They found that there was not significant consistency in the selection of covariates in the final models across the 17 assessed studies and that in less than half was either internal or external validation performed, although, when done, the validation methodologies that were employed were appropriate.

In a similar manner, Zabor et al. (2018) conducted a comparison of three internet-based melanoma prognostic tools by validating each of them against their own single-institution data set. Measures of discrimination (C-index) and calibration curves were reported and demonstrated significant variability among the tools (Zabor et al. 2018). In a clinical setting, for example, patients might receive varying estimates of prognosis depending on the model employed and the tool-specific range of predicted prognosis.

Challenges and Opportunities

The biggest challenges for the development of any prediction or prognostic tool relate to the underlying data set used for model development. These

challenges include obtaining sufficient sample size and number of events, minimizing selection bias, and reducing missing data.

Sample Size

Performance of a good model increases with increasing sample size, given that there are sufficient numbers of events (e.g., the rule of thumb of a minimum of ten events per covariate in the model). Models cannot predict an event with any degree of certainty when there are only a few events to train the model. The number of events is also a function of the median follow-up duration of the training data set. While sample size is important for training any model, it is also a challenge in developing models for conditional survival, since each future time point assessed (e.g., conditional survival at 1-, 2-, or 5-year post diagnosis) includes successively fewer patients entered into the model. Also, without mature follow-up, only earlier conditional time points may be assessed. The specific time points that can be assessed in a conditional survival analysis, therefore, are limited by sample size, number of events, and length of overall follow-up for the patient cohort.

Selection Bias

Selection bias, in the context of model building, is defined as artificial weighting introduced by a study cohort that does not represent the entire population at risk, but just a selected subset. For example, if patient records are more likely to be updated and available from a single-institution database for patients who received a particular type of treatment, those with a specific characteristic or stage of disease, or those with more frequent follow-up, nonrandom biases may occur and potentially skew model predictions toward those outcomes most represented. Thus, it is important to understand and minimize these potential biases that are often found in clinical databases, by determining the total number of patients at risk in the time frame and potential characteristics that may be

significantly impacting model development and validation.

Missing Data

Missing data (for dependent or independent variables, e.g., outcome or predictor) are a regular concern when utilizing older data sets, particularly if certain variables were not historically collected (e.g., SLN tumor burden) or collected with differing definitions (e.g., mitotic rate per high-power field [an old definition] versus per square millimeter [more contemporary definition]). Missing data present significant issues for multivariable modeling, since one missing covariate within a prediction model necessarily excludes the entire case from analysis in standard regression analyses. Multiple imputation provides a method for imputing sets of missing values via Bayesian sampling and conducting iterative statistical model assessment that can be combined into a single prediction formula (Molenberghs and Kenward 2007) without eliminating any data.

With respect to survival-based outcome, missing cause of death may restrict the model to predicting only overall instead of melanoma-specific survival. As noted previously, overall survival may be a more problematic outcome to interpret for certain groups of patients, such as older adults (Berry et al. 2010), or, alternatively, those at lower risk of dying from disease. In such cases, interpretation of any-cause death may not provide an accurate prediction of the patient's outcome specifically with respect to their disease.

Further complicating outcome analyses is the prevalence of missing data of known or putative prognostic factors. In the stage IV melanoma arena, significant missing LDH data in many institutional data sets and limitations in the coding of anatomic sites of metastases, both covariates known to significantly impact metastatic melanoma patient prognosis, limit stage IV outcome analyses. To address this issue, the eighth edition staging system for melanoma was updated with the specific intention of facilitating more robust and complete data capture of these factors in the future.

Survival from Metastatic Melanoma

Over the past decade, new, efficacious, targeted, and immune systemic therapies have revolutionized the treatment of patients diagnosed with regional and metastatic melanoma and have led to significantly improved outcomes (► “Cytokines (IL-2, IFN, GM-CSF, etc.) Melanoma” and ► “Managing Checkpoint Inhibitor Symptoms and Toxicity for Metastatic Melanoma”). This evolving landscape also creates specific challenges for estimating prognosis for patients with distant metastasis, since not all treatments are applicable to all patients with metastatic melanoma (e.g., only patients with particular BRAF mutations are eligible for treatment with BRAF inhibitors), and various agents can be administered concomitantly or sequentially. The resulting heterogeneous and time-dependent patterns of treatment introduce new layers of complexity for building multiple regression models. As the treatment landscape for patients with distant metastatic melanoma has been in rapid evolution over the past 5–10 years, analysis of patients with distant metastatic melanoma was not conducted for the eighth edition AJCC melanoma staging and classification system (Chap. 10) to allow for additional maturation of these exciting developments (Gershenwald et al. 2017a, b).

The Future

Complete data sets for training and validation capturing known prognostic factors for melanoma, together with large sample sizes to overcome the inevitable subgrouping that will be necessary to stratify for treatment eligibility, as well as an opportunity to develop novel statistical approaches, will all be essential to build robust, contemporary prediction models for metastatic melanoma patients in the future. For melanoma, successful integration of known clinical and pathological factors together with new molecular and immune biomarkers into prediction models is the next bar to reach to approach a more personalized prediction of a melanoma patient's outcome.

Cross-References

- ▶ [Biopsy of the Sentinel Lymph Node](#)
- ▶ [Checkpoint Inhibitors in the Treatment of Metastatic Melanoma](#)
- ▶ [Classification and Histopathology of Melanoma](#)
- ▶ [Melanoma Prognosis and Staging](#)
- ▶ [Radiotherapy for Primary and Regional Melanoma](#)
- ▶ [Surveillance and Follow-Up of Melanoma Patients](#)
- ▶ [Targeted Therapies for BRAF-Mutant Metastatic Melanoma](#)

References

- Alba AC, Agoritsas T, Walsh M et al (2017) Discrimination and calibration of clinical prediction models users' guides to the medical literature. *JAMA* 318(14):1377–1384
- Balch CM, Buzaid AC, Soong SJ et al (2001a) Final version of the American joint committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol* 19(16):3635–3648
- Balch CM, Soong SJ, Gershenwald JE et al (2001b) Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on cancer melanoma staging system. *J Clin Oncol* 19(16):3622–3634
- Berry SD, Ngo L, Samelson EJ et al (2010) Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc* 58(4):783–787
- Cadili A, Dabbs K, Scolyer RA et al (2010) Re-evaluation of a scoring system to predict nonsentinel-node metastasis and prognosis in melanoma patients. *J Am Coll Surg* 211(4):522–525
- Callender GG, Gershenwald JE, Egger ME et al (2012) A novel and accurate computer model of melanoma prognosis for patients staged by sentinel lymph node biopsy: comparison with the American Joint Committee on Cancer model. *J Am Coll Surg* 214(4):608–617
- Cochran AJ, Elashoff D, Morton DL et al (2000) Individualized prognosis for melanoma patients. *Hum Pathol* 31(3):327–331
- Collett D (2015) *Modelling survival data in medical research*, 3rd edn. CRC Press, Boca Raton
- Collins GS, Reitsma JB, Altman DG et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol* 68(2):134–143
- Cox DR (1972) Regression models and life tables (with discussion). *J R Stat Soc B* 34:187
- Faries MB, Thompson JF, Cochran AJ et al (2017) Completion dissection or observation for sentinel-node metastasis in melanoma. *N Engl J Med* 376(23):2211–2222
- Gershenwald JE, Scolyer RA, Hess KR et al (2017a) Melanoma of the skin. In: Amin M, Edge SB, Greene FL et al (eds) *AJCC cancer staging manual*, 8th edn. Springer, Cham, pp 563–585
- Gershenwald JE, Scolyer RA, Hess KR et al (2017b) Melanoma staging: evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 67(6):472–492
- Gimotty PA, Elder DE, Fraker DL et al (2007) Identification of high-risk patients among those diagnosed with thin cutaneous melanoma. *J Clin Oncol* 25(9):1129–1134
- Gourgou-Bourgade S, Cameron D, Poortmans P et al (2015) Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (definition for the assessment of time-to-event endpoints in CANcer trials). *Ann Oncol* 26(5):873–879
- Harrell FE, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15(4):361–387
- Haydu LE, Scolyer RA, Lo S et al (2017) Conditional survival: an assessment of the prognosis of patients at time points after initial diagnosis and treatment of locoregional melanoma metastasis. *J Clin Oncol* 35(15):1721–1731
- Hess KR (1995) Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Stat Med* 14(15):1707–1723
- Hieke S, Kleber M, König C et al (2015) Conditional survival: a useful concept to provide information on how prognosis evolves over time. *Clin Cancer Res* 21(7):1530–1537
- Hosmer DW, Lemeshow S (2000) *Applied logistic regression*, 2nd edn. Wiley, New York
- Hosmer DW, Lemeshow S, May S (2007) *Applied survival analysis regression modeling of time-to-event data*, 2nd edn. Wiley, Hoboken
- Kattan MW, Hess KR, Amin MB et al (2016) 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* 66(5):370–374
- Khosrotehrani K, van der Ploeg AP, Siskind V et al (2014) Nomograms to predict recurrence and survival in stage IIIB and IIIC melanoma after therapeutic lymphadenectomy. *Eur J Cancer* 50(7):1301–1309
- Kim HT (2007) Cumulative incidence in competing risks data and competing risks regression analysis. *Clin Cancer Res* 13(2):559–565
- Leiter U, Stadler R, Mauch C et al (2016) Complete lymph node dissection versus no dissection in patients with sentinel lymph node biopsy positive melanoma

- (DeCOG-SLT): a multicentre, randomised, phase 3 trial. *Lancet Oncol* 17(6):757–767
- Lyth J, Hansson J, Ingvar C et al (2013) Prognostic sub-classifications of T1 cutaneous melanomas based on ulceration, tumour thickness and Clark's level of invasion: results of a population-based study from the Swedish melanoma register. *Br J Dermatol* 168(4):779–786
- Mahar AL, Compton C, Halabi S et al (2016) Critical assessment of clinical prognostic tools in melanoma. *Ann Surg Oncol* 23(9):2753–2761
- Maurichi A, Miceli R, Camerini T et al (2014) Prediction of survival in patients with thin melanoma: results from a multi-institution study. *J Clin Oncol* 32(23):2479–2485
- Michaelson JS (2011) Melanoma outcome calculator. <http://lifemath.net/cancer/melanoma/outcome/index.php>
- Molenberghs G, Kenward MG (2007) Missing data in clinical studies. Wiley, West Sussex
- Morton DL, Wen DR, Wong JH et al (1992) Technical details of intraoperative lymphatic mapping for early stage melanoma. *Arch Surg* 127:392–399
- Murali R, Sesilva C, Thompson JF, Scolyer RA (2010) Non-sentinel node risk score (N-SNORE): a scoring system for accurately stratifying risk of non-sentinel node positivity in patients with cutaneous melanoma with positive sentinel lymph nodes. *J Clin Oncol* 28(29):4441–4449
- National Cancer Institute of the National Institutes of Health (2018.) <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/personalized-medicine>. Accessed 05 Nov 2018
- Pintilie M (2006) Competing risks a practical perspective. Wiley, West Sussex
- Soong SJ (1985) A computerized mathematical model and scoring system for predicting outcome in melanoma patients. In: Balch CM, Milton GW (eds) *Cutaneous melanoma: clinical management and treatment results worldwide*. JB Lippincott, Philadelphia, p 353
- Soong SJ (1992) A computerized mathematical model and scoring system for predicting outcome in patients with localized melanoma. In: Balch CM, Houghton AN, Milton GW et al (eds) *Cutaneous melanoma*, 2nd edn. JB Lippincott, Philadelphia, p 200
- Soong SJ, Shaw HM, Balch CM et al (1992) Predicting survival and recurrence in localized melanoma: a multivariate approach. *World J Surg* 16(2):191–195
- Soong SJ, Zhang Y, Desmond R (2003) Models for predicting outcome. In: Balch CM, Houghton AN, Sober A, Soong SJ (eds) *Cutaneous melanoma*, 4th edn. Quality Medical Publishing, St. Louis, p 77
- Soong SJ, Ding S, Coit DG et al (2009) Models for predicting melanoma outcome. In: Balch CM, Houghton AN, Sober AJ et al (eds) *Cutaneous melanoma*, 5th edn. Quality Medical Publishing, St. Louis, pp 87–104
- Soong SJ, Ding S, Coit D et al (2010) Predicting survival outcome of localized melanoma: an electronic prediction tool based on the AJCC melanoma database. *Ann Surg Oncol* 17(8):2006–2014
- U.S. Department of Health and Human Services, Food and Drug Administration (2007) Guidance for industry clinical trial endpoints for the approval of cancer drugs and biologics. <https://www.fda.gov/downloads/drugs/GuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>. Accessed 7 Oct 2018
- Zabor EC, Coit DG, Gershenwald JE et al (2018) Variability in predictions from online tools: a demonstration using internet-based melanoma predictors. *Ann Surg Oncol* 25(8):2172–2177