The identification of prognostic factors and building of risk assessment prognostic models will continue to play a major role in 21st century medicine in patient management and decision making. Investigators often are interested in examining the relationship among host, tumor-related, and environmental variables in predicting clinical outcomes. We distinguish between static and dynamic prediction models. In static prediction modeling, variables collected at baseline typically are used in building models. On the other hand, dynamic predictive models leverage the longitudinal data of covariates collected during treatment or follow-up and hence provide accurate predictions of patients’ prognoses. To date, most risk assessment models in oncology have been based on static models. In this article, we cover topics related to the analysis of prognostic factors, centering on factors that are both relevant at the time of diagnosis or initial treatment and during treatment. We describe the types of risk prediction and then provide a brief description of the penalized regression methods. We then review the state-of-the-art methods for dynamic prediction and compare the strengths and limitations of these methods. Although static models will continue to play an important role in oncology, developing and validating dynamic models of clinical outcomes need to take a higher priority. A framework for developing and validating dynamic tools in oncology seems to still be needed. One of the limitations in oncology that may constrain modelers is the lack of access to longitudinal biomarker data. It is highly recommended that the next generation of risk assessments consider longitudinal biomarker data and outcomes so that prediction can be continually updated.

INTRODUCTION

The identification of prognostic factors and building of risk assessment prognostic models will continue to play a major role in 21st century medicine in patient management and decision making. Prognostic factors in oncology associate host and tumor variables with clinical outcomes independent of treatment. Gospodarowicz et al classified factors as either tumor related, host, or environmental. Tumor-related factors are variables related to the tumor and reveal the tumor biology and pathology, such as size of tumor, lymph node involvement, presence of metastasis, and molecular markers (overexpression of PTEN gene, presence of androgen receptor variant AR-V7). Host factors are associated with patient characteristics, such as age and comorbidities. Finally, environmental factors are external to the patient, such as access to health care. Prognostic models are increasingly used in the design, conduct, and analysis of clinical trials. For example, in several trials of prostate cancer, randomization was stratified by the predicted survival probability determined by the prognostic model of overall survival. In the TAILORx trial, Oncotype DX, a 21-gene score that predicts the likelihood of recurrence, is used to classify women with breast cancer by their risk group of recurrence. Prognostic factors also have been used for enriching the patient population in trials with targeted therapies. For example, in the ToGA trial, patients with human epidermal growth factor receptor 2–positive gastric cancer were randomly assigned to either trastuzumab plus chemotherapy or chemotherapy alone. In this article, we cover topics that are related to the analysis of prognostic factors, centering on factors that are relevant both at the time of diagnosis or initial treatment and during treatment or follow-up. We use the terms prognostic models, risk models, and risk assessments interchangeably. This article is organized in the following way. First, we describe the type of risk prediction and then provide a brief description of the penalized regression methods. Second, we review the state-of-the-art methods for dynamic prediction and compare the strengths and limitations of these methods. Third, we offer a concise discussion of validation and metrics for assessing models. Finally, we present recommendations for the next generation of risk assessment methods to be built in modern oncology.
CONTEXT SUMMARY

Key Objective
Building prognostic models will continue to play a role in 21st century medicine in patient management. We make a distinction between static and dynamic predictive models and provide a review of the state-of-the-art methods for dynamic predictive models to promote them for future use.

Knowledge Generated
To date, most risk assessment models in oncology have been based on static prognostic models. An understanding of the longitudinal relationship between host and tumor-related factors and their impact on clinical outcomes is critical. Regardless of whether static or dynamic modeling is the primary objective, we envision that this review will encourage investigators to take risk assessment as a discipline by itself.

Relevance
We expect to see an upsurge in dynamic risk assessments, and as such, it is highly recommended that the next generation of models consider the longitudinal data and outcomes so that predictions are updated.

TYPES OF RISK PREDICTIONS
Investigators are interested in examining the relationship between host and tumor-related factors in predicting clinical outcomes (Fig 1A). We distinguish between static and dynamic prediction. In static prediction modeling, variables collected at baseline typically are used in building models. For example, prostate-specific antigen (PSA) measurements at baseline are used for prediction of recurrence. On the other hand, dynamic predictive models explicitly leverage the longitudinal data of covariates that are collected during treatment or follow-up. In patients with advanced cancer, the disease has substantially evolved

![Diagram A](image)

**FIG 1.** (A) Relationship between host and tumor-related factors and clinical outcomes. Modified from Halabi and Owzar.2 (B) Dynamic risk prediction framework. LDH, lactate dehydrogenase; PSA, prostate-specific antigen.
and has a heterogeneous presentation within the patient. The inter- and intrapatient variability should be taken into account in statistical modeling. Dynamic prediction incorporates time-dependent covariates so that risk prediction would be continually updated with new observations to reflect the patient’s prognosis.

We define terminology that is typically used in dynamic risk prediction. The term landmark time is defined as a current time point at which we have data (host, tumor-related variables, and outcomes). The term horizon time is defined as a future time point at which we want to predict a time-to-event outcome, such as overall survival. Dynamic predictive models essentially capture the historical information of the longitudinal measurements from the study baseline to the landmark time (t) such that the risk at horizon time (u) can be accurately predicted (Fig 1B).

IDENTIFICATION OF PROGNOSTIC FACTORS

Several popular strategies exist for identifying prognostic factors in static risk assessment. Standard variable selection approaches, such as forward selection, backward selection, and so forth, with logistic regression for binary end points, and proportional hazards regression for time-to-event end points, have been applied. Criticism for the stepwise variable selection has been well documented. Several popular strategies exist for identifying prognostic factors, and outcomes). The term horizon time is defined as a future time point at which we want to predict a time-to-event outcome, such as overall survival. Dynamic predictive models essentially capture the historical information of the longitudinal measurements from the study baseline to the landmark time (t) such that the risk at horizon time (u) can be accurately predicted (Fig 1B).

Identification of Prognostic Factors

Several popular strategies exist for identifying prognostic factors in static risk assessment. Standard variable selection approaches, such as forward selection, backward selection, and so forth, with logistic regression for binary end points, and proportional hazards regression for time-to-event end points, have been applied. Criticism for the stepwise variable selection has been well documented. Of note, classification trees, such as recursive partitioning for both binary and time-to-event end points, frequently have been used. We concentrate on penalized methods that fit and shrink $p$ predictors and in doing so, reduce the variance of the coefficient estimates. Thus, these methods would improve the accuracy of the model. The least absolute shrinkage and selection operator (LASSO) and adaptive LASSO (ALASSO) have been widely used to develop prognostic models of clinical outcomes. We will briefly describe ridge regression and penalized methods. Ridge regression minimizes the residual sum of squares function, but a caveat is that it does not reduce all the coefficients exactly to 0.

Let $y_i$ be the response, $x_{ij}$ the $j$th covariate value ($j = 1, 2, \ldots, p$) corresponding to the $i$th individual, $b_j$ the regression coefficient $j$th covariate, and $\lambda$ a tuning parameter. Similar to ridge regression, the first term in LASSO is the residual sum of squares, and LASSO minimizes this function subject to the $l_1$ penalty (Eq 1):

$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|.$$  

A large tuning parameter causes the coefficient estimates to be equal to 0, thus the LASSO will have the sparsity property. LASSO is an improvement over ridge regression, although it has the main limitation of tending to select too many unimportant variables, and it performs poorly in situations when $p > n$. ALASSO has been proposed as an improvement over LASSO to overcome the limitation of LASSO. ALASSO minimizes this function (Eq 2):

$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|.$$  

ALASSO uses a weighted penalty term in the $l_1$ penalty where $w = (w_1, w_2, \ldots, w_p)$ is the weight vector. If $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator (eg, $\hat{\beta}$ (OLS)) of $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$, then an appropriate choice of the weight $w$ is $1 / \sqrt{\hat{\beta}}$. ALASSO is considered to be an improvement over LASSO because it has consistent variable selection as well as lower prediction error. Consequently, ALASSO tends to select fewer nonzero coefficients than the LASSO despite having a smaller prediction error. The ALASSO enjoys the oracle property.

Elastic net regression uses a combination of $l_1$ penalty and ridge $l_2$ penalty and is a compromise between LASSO and ridge regression. Furthermore, one of its main advantages when $p > n$ is that it retains more than $n$ variables in the model. Hastie et al provided a thorough comparison of these shrinkage techniques.

We applied LASSO and ALASSO from CALGB 90401, a phase III clinical trial in advanced prostate cancer, with the overall goal of building a model of overall survival. We had 22 variables that were common between CALGB 90401 and the Enthuse trial (external data set). Because of missing data in the covariates, we used methods to impute them as described by White and Royston. The regression’s estimates from the Cox proportional hazards model, LASSO, and ALASSO, are listed in Table 1. We considered LASSO and ALASSO and applied both the Akaike information criterion and the Bayesian information criterion to choose the optimal model of overall survival. LASSO and ALASSO selected eight and nine variables, respectively (Table 1). We determined the ALASSO model to be the final optimal model since it included the site of metastases for bone disease. Figure 2 shows the solution path for ALASSO, and we observe that lactate dehydrogenase (LDH) greater than the upper limit of normal and Eastern Cooperative Oncology Group (ECOG) performance status were selected early in the $l_1$ path compared to the other variables. This is followed by visceral disease, alkaline phosphatase, albumin, hemoglobin, pain, bone metastases, and then PSA (the Bayesian information criterion stopped at PSA). The final model selected the following prognostic factors: LDH greater than the upper limit of normal, ECOG performance status, metastatic site (presence of visceral disease, presence of bone metastases), PSA, alkaline phosphatase, albumin, hemoglobin, and analgesic opioid use. We have focused on variable selection when the number of predictors is small relative to the sample size. There are two main challenges in identifying potential prognostic factors.
TABLE 1. Identified Prognostic Factors by LASSO and ALASSO in the Training Set

<table>
<thead>
<tr>
<th>Description of Variable (variable name)</th>
<th>Cox Model, $\hat{\beta}$</th>
<th>LASSO, $\hat{\beta}$</th>
<th>ALASSO, $\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site of metastases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone (DS2)</td>
<td>0.234</td>
<td>—</td>
<td>0.058</td>
</tr>
<tr>
<td>Visceral (DS3)</td>
<td>0.400</td>
<td>0.056</td>
<td>0.293</td>
</tr>
<tr>
<td>Liver (LIVER)</td>
<td>0.013</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lung (LUNG)</td>
<td>0.199</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Opioid analgesic use (PAIN)</td>
<td>0.136</td>
<td>0.077</td>
<td>0.088</td>
</tr>
<tr>
<td>Age in years (AGE)</td>
<td>−0.003</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Body mass index (BMI)</td>
<td>−0.021</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>White race (Caucasian)</td>
<td>0.034</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ECOG performance status (ECOG)</td>
<td>0.278</td>
<td>0.190</td>
<td>0.305</td>
</tr>
<tr>
<td>Comorbidity (Comorb)</td>
<td>0.070</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gleason score (GLEAST)</td>
<td>0.052</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prior treatment with radiotherapy (Radio)</td>
<td>0.105</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LDH ≥ ULN (LDH.Hight)</td>
<td>0.325</td>
<td>0.203</td>
<td>0.335</td>
</tr>
<tr>
<td>Albumin (ALB)</td>
<td>−0.133</td>
<td>−0.080</td>
<td>−0.122</td>
</tr>
<tr>
<td>Bilirubin (BIL)</td>
<td>−0.017</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Hemoglobin (HGB)</td>
<td>−0.094</td>
<td>−0.085</td>
<td>−0.065</td>
</tr>
<tr>
<td>Platelets (PLT)</td>
<td>0.000</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>WBCs (WBC)</td>
<td>0.055</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Alkaline phosphatase (ALKPHOS)</td>
<td>0.145</td>
<td>0.138</td>
<td>0.145</td>
</tr>
<tr>
<td>Aspartate aminotransferase (AST)</td>
<td>0.025</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prostate-specific antigen (PSA)</td>
<td>0.063</td>
<td>0.026</td>
<td>0.015</td>
</tr>
<tr>
<td>Testosterone (TESTO)</td>
<td>−0.088</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Training C-index</td>
<td>0.662</td>
<td>0.660</td>
<td>—</td>
</tr>
<tr>
<td>Integrated time AUC</td>
<td>0.742</td>
<td>0.740</td>
<td>—</td>
</tr>
</tbody>
</table>

Abbreviations: ALASSO, adaptive least absolute shrinkage and selection operator; AUC, area under the curve; $\hat{\beta}$, estimated regression coefficient; LASSO, least absolute shrinkage and selection operator; LDH, lactate dehydrogenase; ULN, upper limit of normal.

in high-dimensional space: computational intensity and a high false discovery rate.38,39 Of note, several prescreening methods are useful in identifying prognostic features both in the large $p$, small $n$ problem and in ultra-high-dimensional space.31,32,39-41

The concept of variable selection is more challenging in building dynamic models because the main goal is to identify important factors that are related to the longitudinal process and the outcomes. In recent years, a few statistical studies extended the penalized method for the joint modeling of longitudinal data and survival outcomes.42,43

The general idea is to postulate the joint likelihood linking the two submodels through latent random variables and to add shrinkage operators to select fixed and random effects. He et al.185 proposed a variable selection method for joint modeling with a univariable longitudinal outcome, and Chen and Wang184 extended the framework to incorporate multiple longitudinal outcomes. While these methods have not been implemented in oncology, the statistical development paves the way for dynamic risk prediction.

Heterogeneity of treatment effect (HTE) is another important area to consider when building prognostic models. HTE is the nonrandom, explainable variability in the direction and magnitude of treatment effects for individuals within a population.44 There are different sources for HTE, and HTE may arise from an underlying causal mechanism, artifacts, measurements, or methods.45,46 One main goal of HTE analyses is to predict whether a patient might benefit from a treatment. Traditionally, the Cox regression method has been used to identify subgroups of patients who may benefit from treatment.14 Recursive partitioning also has been used to identify a subgroup of patients.47 While these classification tree methods have several advantages, they can create complicated structures and produce overfitting.19,41,47 Other methods, such as permutation methods, SIDES (subgroup identification based on differential effect search), doubly robust augmented inverse probability–weighted estimator, and virtual twins, have been developed to take HTE into consideration.45,48-52

The personalized prediction can be adequately addressed by the dynamic prediction framework, although it is an area for future research.

ESTIMATING PATIENT-SPECIFIC OUTCOME PREDICTION AND CONSTRUCTING RISK GROUPS

Once the final model is chosen, the next step is estimating patient-specific outcome prediction. The estimated survival function at time $t$ is (Eq 3)

$$\hat{S}(t) = (\hat{S}_0(t))^{\exp R},$$  \hspace{1cm} (3)

where $R$ is the estimated linear predictor or risk score for the $i$th individual ($R_i = \mathbf{x}_i^T \hat{\beta} = \sum_{j=1}^p \hat{\beta}_j x_{ij}$), the baseline survival function is $\hat{S}_0(t) = \exp(-H_0(t))$, and the baseline cumulative hazard function is $H_0(t)$. When we turned to our prognostic model of overall survival in prostate cancer, we computed a risk score from the estimated regression coefficients and the predicted survival at 24 months using the baseline cumulative hazard. We present the profiles of two patients with different baseline prognostic factors and their predicted overall survival at 24 months5 (Table 2). We observe that patient 2 had a worse predicted survival probability at 24 months than patient 1.

Another important task in static predictive modeling is to construct prognostic risk groups, which can be formed on the basis of their quantiles from the estimated linear predictor. In our overall survival model, we constructed two and three prognostic risk groups and determined the cut points from the training set on the basis of quantiles (33rd, 50th, and 67th percentiles).5 While demonstrating that the overall survival curves differ across the three risk groups is appropriate, this approach is not sufficient.53 The optimal
strategy would be to compute a measure of discriminative ability of the model.

**METHODS FOR DYNAMIC MODELING**

Let $T_i$ denote the true failure time and $f_i(t)$ a set of longitudinal measurements at some time points up to landmark time $t$. We are interested in predicting the probability that a new patient $i^*$ is event free at least up to horizon time $u > t$ given survival up to $t$. The conditional probability is defined as (Eq 4)

$$\pi_i(u|t) = Pr(T_i^* \geq u| T_i > t, f_i(t), D_t),$$

where $D_t$ denotes a sample from the target population and on which the prediction is based. This formulation enables a dynamic updating scheme. Indeed, if a new

---

**FIG 2.** Solution path for adaptive least absolute shrinkage and selection operator. ALB, albumin; ALKPHOS, alkaline phosphatase; BMI, body mass index; comorb, comorbidity; DS2, bone metastases; DS3, visceral metastases; ECOG, Eastern Cooperative Oncology Group performance status; GLEAST, Gleason score; HGB, hemoglobin; LDH.High, lactate dehydrogenase greater than or equal to the upper limit of normal; PAIN, opioid analgesic use; PSA, prostate-specific antigen; Radio, prior treatment with radiotherapy; TESTO, testosterone. Reproduced with permission.96

---

**TABLE 2. Profiles of Patient-Specific Predicted Probabilities**

<table>
<thead>
<tr>
<th>Patient No.</th>
<th>Disease Site</th>
<th>Opiate Use</th>
<th>ECOG</th>
<th>LDH &gt; ULN</th>
<th>ALB (g/dL)</th>
<th>HGB (g/dL)</th>
<th>ALK (U/L)</th>
<th>PSA (ng/mL)</th>
<th>Predicted Probability at 24 Months (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bone</td>
<td>Yes</td>
<td>1</td>
<td>No</td>
<td>4</td>
<td>14</td>
<td>130</td>
<td>90</td>
<td>0.47 (0.42 to 0.52)</td>
</tr>
<tr>
<td>2</td>
<td>Visceral</td>
<td>Yes</td>
<td>1</td>
<td>Yes</td>
<td>3</td>
<td>10</td>
<td>90</td>
<td>75</td>
<td>0.28 (0.19 to 0.38)</td>
</tr>
</tbody>
</table>

NOTE. From Halabi et al.5

Abbreviations: ALB, albumin; ALK, alkaline phosphatase; ECOG, Eastern Cooperative Oncology Group performance status; HGB, hemoglobin; LDH, lactate dehydrogenase; PSA, prostate-specific antigen; ULN, upper limit of normal.

*Computed using the First-Line Metastatic Castrate-Resistant Prostate Cancer Patients risk classification tool.97
There are two general dynamic risk prediction frameworks: joint modeling and landmark analysis. Joint modeling comprises two linked submodels, one for the longitudinal process and one for the time-to-event data, and both depend on a common set of latent random variables. In particular, the longitudinal data usually are modeled by a linear mixed-effects model. The time-to-event data are modeled by the proportional hazards model, with true longitudinal processes as time-varying covariates. The Cox regression coefficient quantifies the association between the latent longitudinal process and the hazard rate at time \( t \). The longitudinal process and the event time process are assumed to be independent given the latent random effects, and the joint likelihood can be derived. The model can be estimated either using a frequentist approach that attains maximum likelihood through an expectation-maximization algorithm or a Bayesian approach that uses Markov chain Monte Carlo to obtain posterior integration. In contrast, landmarking circumvents the aforementioned model assumptions and computational complexity. Joint modeling and landmarking approaches differ in three aspects: model assumptions, information used, and computational complexity. Joint modeling models the dual aspects: model assumptions, information used, and computational complexity. In particular, these methods have been incorporated with joint modeling and are exploited to construct dynamic prediction models.

**EXAMPLES**

A few examples exist in cancer where longitudinal data were modeled with clinical outcomes. We have previously explored whether PSA decline at different landmark times is prognostic for overall survival in patients with advanced prostate cancer. Fontein et al developed a dynamic model for predicting overall survival in patients with breast cancer. The authors validated the overall survival model using a dynamic cross-validated C-index and reported C-indices of 0.72, 0.76, and 0.79 at 1, 2, and 3 years, respectively. Suresh et al and Ferrer et al compared the two approaches in the case of model misspecification, and they aimed for predicting competing risks of prostate cancer from PSA history. Dynamic risk prediction (joint modeling) relies on model assumptions, and hence, its performance suffers from model misspecification. Functional data analysis, which is a nonparametric framework, has received considerable attention in medical studies because it is a flexible tool for modeling nonlinear longitudinal processes. In particular, these methods have been incorporated with joint modeling and are exploited to construct dynamic prediction models.
trajectories (Fig 3A) and risk of functional disability (Fig 3B). Patient 169, who had more severe disease with earlier development of functional disability, and patient 718, who had less severe disease, were selected to illustrate the patient-specific predictions at clinically relevant future time points conditional on their available longitudinal measurements. The predicted Unified PD Rating Scale trajectories were biased, with wide uncertainty bands when only baseline measurements were used. Although dynamic prediction for longitudinal trajectories is an advantage of the

![Graphs showing predicted trajectories and failure probabilities over time for different visits.](image)

FIG 3. (A) Predicted Unified Parkinson Disease Rating Scale (UPDRS) trajectories and (B) predicted conditional failure probabilities for patient 169 (top rows) and patient 718 (bottom rows). Solid lines are the means of 2,000 Markov chain Monte Carlo samples. Dashed lines are the 2.5 and 97.5 percentiles of the 2,000 Markov chain Monte Carlo samples. The vertical lines represent the landmark time. Reproduced with permission.87
joint modeling, our major interest was to predict the probability of functional disability after visits at time $t$ given the patients’ longitudinal profiles and event-free status up to time $t$. We found a similar pattern in that the risk predictions with higher accuracy were achieved on the basis of more longitudinal observations. For example, according to the longitudinal profiles of the first 12 months, the predicted probabilities in the next 3, 6, 9, and 12 months were 0.21, 0.46, 0.78, and 0.97, respectively, for patient 169 (Fig 3B, last plot, top row) and 0.02, 0.06, 0.13, and 0.30, respectively, for patient 718 (Fig 3B, last plot, bottom row). Therefore, patient 169, who had a higher risk of functional disability in the next few months, might need attentive medical intervention to control disease progression. Meanwhile, the Brier’s scores were 0.216 and 0.108, respectively, which implied a better prediction in terms of calibration given more follow-up data.

**VALIDATION AND ASSESSMENT OF PROGNOSTIC MODELS**

The primary goal of a risk assessment is to provide accurate outcome prediction in new patients. Overfitting remains one of the main challenges in model building. Overfitting occurs when a high predictive accuracy is estimated from a model that has been applied to the training set but has low accuracy when assessed in an independent data set. A good example of overfitting is provided in Halabi and Owzar.

The validation of a prognostic model is considered a critical step after a risk assessment model has been built. There are two types of validation: external and internal. External validation, where the frozen model from the training data is applied to an independent data set, is the most rigorous approach. However, investigators often may not have access to the external data set. Of note, other types of resampling methods, such as cross-validation, bootstrapping, and bootstrapping using 0.632+, are considered appropriate approaches to model validation.

Assessment of the model’s performance usually is conducted by examining the calibration and discriminative ability of the model. Calibration signifies the extent of the match between the predicted and observed outcome. Often, investigators plot the predicted versus the observed outcome. The model would be calibrated if the data fall on a 45° line. Using data from two phase III clinical trials, we evaluated the overall survival model for calibration at different landmarks. Figure 4A shows that the predicted survival probabilities at 18 months in the Enthuse 33 trial were close to the proportion of patients who survived 18 months. On the other hand, Figure 4B demonstrates that the model was not well gauged because the observed-predicted data points did not fall on the 45° line. The first two points (circles) show that the model overpredicted the proportion of patients who survived 12 months, whereas the third and fifth data points show that the model underpredicted the proportion of patients who survived 12 months.

Discrimination describes the ability of a prognostic model to distinguish between patients with and without the outcome of interest. Several metrics are used to report the performance of a model. A widely used measure is concordance (C-index), which is the agreement between observed outcomes and prediction. Another widely used measure of predictive accuracy is the time-dependent area under the receiver operating characteristic curve (tAUROC), which can be combined to form an integrated AUROC for the whole range of the study. Circling back to our prognostic model in prostate cancer, we evaluated the performance of the model by implementing the integrated measure for the tAUROC by Uno et al, which was 0.73 (95% CI, 0.70 to 0.73) and 0.76 (95% CI, 0.72 to 0.76) in the testing and validation sets, respectively.

Dynamic models also need to be assessed for their discriminative ability and calibration. These measures can evaluate the performance of the model at various time points of the prediction. The tAUROC and Brier’s score are widely used for dynamic prediction validation. In the DATATOP study, we applied fivefold cross-validation to evaluate the predictive performance of our framework. Conditional on the longitudinal history up to month 3 and month 12, our model yielded tAUROCs of 0.744 and 0.766, respectively, for correctly assigning a higher risk of functional disability by month 15 to patients with more severe disease.

Criteria for evaluating risk assessments have been published by the Precision Medicine Core of the American Joint Commission on Cancer, the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis, and the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies. Investigators are encouraged to follow these guidelines as more rigorous tools of clinical outcomes would be developed in oncology. These models are anticipated to be implemented in both the design and the conduct of future trials.

Although static models will continue to play an important role in oncology, developing and validating dynamic models of clinical outcomes need to take a higher priority. A framework for developing and validating dynamic tools in oncology seems to be needed. One of the limitations is that modelers may be constrained by the lack of access to the longitudinal biomarker data; therefore, the next generation of risk assessments are highly recommended to take into consideration the longitudinal biomarker data and outcomes so that predictions are updated.

In summary, risk assessment will remain an important research task in precision oncology. We advocate for good clinical practice in risk assessment studies and recommend that investigators design these studies prospectively to obtain accurate individual outcome prediction and prognostic risk group classification. Prognostic studies should begin by asking fundamental questions that are...
pertinent to patient outcomes, define the primary end point a priori, justify the sample size, and describe the appropriate methods for variable selection and model assessment. Lastly, they should be validated using external data sets if available.

An understanding of the longitudinal relationship between host and tumor-related factors and their impact on clinical outcomes is critical. We expect to see an upsurge in dynamic risk assessments in oncology, and as such, the American Joint Committee on Cancer and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines should be extended to dynamic predictive modeling. Regardless of whether static or dynamic modeling is the primary objective, we envision that this review will bridge gaps in knowledge and motivate investigators to take risk assessment as a discipline by itself. Funding opportunities with the primary goal of building and validating high-quality prognostic models will be critical for personalized predictions.

FIG 4. Calibration of the overall survival model for observed and predicted survival probability at (A) 18 months and (B) 12 months.

Developing and Validating Risk Assessment Models

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST
The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Susan Halabi
Employment: ASCO TAPUR
Consulting or Advisory Role: Eisai, Ferring Pharmaceuticals, Bayer AG

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT
This article draws from and elaborates on the work of Halabi and Owzar,2 Kelly et al,41 and Halabi and Michiels.96

1Duke University Medical Center, Durham, NC

CORRESPONDING AUTHOR
Susan Halabi, PhD, Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Rd, Ste 11088, Durham, NC 27705; Twitter: @DukeCancer; e-mail: susan.halabi@duke.edu.

SUPPORT
Supported in part by US Army Medical Research Awards W81XWH-15-1-0467 and W81XWH-18-1-0278 and the Prostate Cancer Foundation. S.L.’s and C.L.’s research was partly supported by the National Institutes of Health (R01NS091307).

AUTHOR CONTRIBUTIONS
Conception and design: All authors
Collection and assembly of data: All authors
Data analysis and interpretation: All authors
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors