An online tool for recurrence and survival prognostication for patients with newly-diagnosed oropharyngeal cancer

Creator: Dr. Lauren J Beesley
Department of Biostatistics, University of Michigan
Contact: lbeesley@umich.edu

Last updated August 21, 2020

The goal of this application (`http://shiny.sph.umich.edu/Oropharynx_Calc/`) is to provide predictions of likely outcome events for patients newly-diagnosed with oropharyngeal cancer. In particular, we consider outcomes of death, distant metastasis, and locoregional recurrence. For a particular patient, we can imagine possible outcome events that can occur after treatment. At a particular time $t > 0$ after treatment, we can view patients as being in different states corresponding to combinations of no events, locoregional recurrence, distant metastasis, and death. We use results from a statistical model to provide estimates of the proportion of patients with given baseline characteristics that will have had different combinations of outcome events at time t after treatment. We call these probabilities "state occupancy probabilities." Below, we provide some brief descriptions of the model used to estimate these probabilities and how they are calculated and interpreted.

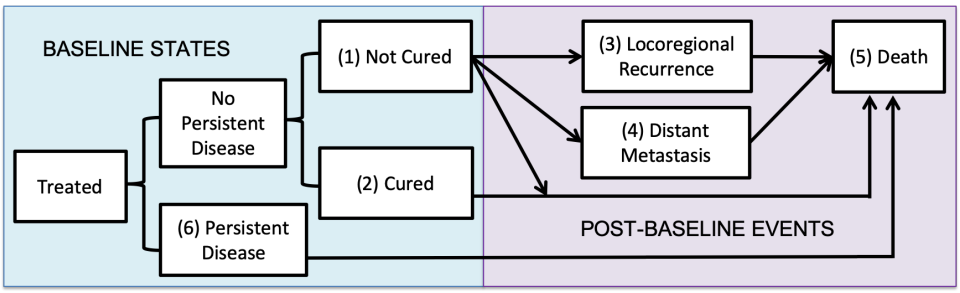
1 The Model

At the time of treatment, we imagine patients can be in one of three baseline states. Firstly, treatment may have erradicated all cells of the primary cancer. We will call these patients cured of disease, and assume they will never experience a primary recurrence (even with infinite follow-up). Secondly, treatment may have reduced the number of primary cancer cells so that they are not observable on scans, but given long enough follow-up they will eventually multiply and become observable, called recurrence. Thirdly, the primary cancer may always be visible after treatment, called persistent disease. These three groups of patients are expected to have very different prognosis and different rates of locoregional recurrence, distant metastasis, and death.

In **Figure 1**, we provide a conceptual diagram of the assumed model structure. Brackets correspond to logistic regression models related to the probability of being in each baseline state. Solid arrows correspond to proportional hazards regression models corresponding to the rate of movement between states. These various logistic regression models and proportional hazards regression models together make up the multistate model of interest. We can

imagine different patient characteristics may drive each one of these models. For example, age may be important for modeling death from other causes but may be less related to the probability of having persistent disease after treatment.

Figure 1: Conceptual diagram of model structure



We fit this model using Bayesian methods to a dataset of N=840 patients treated for oropharyngeal cancer at The University of Michigan. The resulting posterior means for the model parameters are then used to obtain individualized predictions as described in the next section.

2 Using model to predict likely outcomes

State occupancy probabilities represent the expected proportion of patients (with specified characteristics) who will be in different possible outcome states at any follow-up time t. For example, we might be interested in the proportion of 100 patients with fixed characteristics that will die without prior recurrence or persistence by 5 years. We will not provide the formulas for these predictions here. Instead, we just list the possible states a patient may be in at a given time t under our model:

(1) alive with persistent disease
(2) died with persistent disease
(3) alive without persistent disease or any recurrence
(4) alive without persistent disease but with prior locoregional recurrence (with or without subsequent metastasis)
(5) alive without persistent disease but with prior metastasis (with or without subsequent locoregional recurrence)
(6) died without persistent disease but with prior locoregional recurrence (with or without subsequent metastasis)
(7) died without persistent disease but with prior metastasis (with or without subsequent locoregional recurrence)
(8) died without persistent disease or any prior recurrence

Using our model, we obtain the proportion of patients with given characteristics (e.g. age, p16 positivity) that will be in each of the above states at time t. These proportions

will sum to 1. To calculate overall survival, we can add the proportions (1), (3), (4), and (5). Similarly, the probability of being alive with persistent disease or a prior recurrence is obtained by adding proportions (1), (4), and (5). We use different combinations of proportions 1-8 to obtain the probabilities presented in this web tool. These proportions can be calculated for a fixed time t, and they can also be calculated across a range of values for t, which provides a curve of likely outcomes over time.

3 Entering patient characteristics and plot options

Figure 2 shows the fields to enter to obtain the predictions. The first set of fields corresponds to patient characteristics for which likely outcome states are to be estimated. The second set of fields corresponds to whether additional PET and CT scan results are available. If checked, the user is prompted to enter metabolic tumor volume (from PET scan) and whether there was extracapsular extension observed on the CT scan. The third set of fields controls the level of detail in the plot results (right) and the time horizon we want to make predictions for (left). The default time horizon is 60 months, meaning we are predicting the likelihood of different outcomes for a patient with the entered characteristics 5 years after treatment.

Figure 2: Web tool options

4 Understanding output

Figure 3 provides example output for a given set of inputs. Predictions are provided at and up to 60 months post-baseline. The user can input different time horizons using the plot options above. Additionally, the user can control the degree of detail/granularity in the provided predictions using the plot options above. **Figure 3** shows predictions with the least level of detail, which corresponds to the most merging of proportions 1-8 discussed previously.

Subplot 1 provides a visualization of the proportion of 100 patients that would fall into each of three outcome states at 60 months post-treatment. These three states correspond to patients being alive without recurrence or persistence, alive with persistent disease or prior recurrence, and dead. The probability for the first state corresponds to proportion (3) above. The probability for the second state corresponds to the sum of proportions (1), (4), and (5) above. The probability for the third state corresponds to the sum of proportions (2), (6), (7), and (8) above.

Subplot 2 provides the proportions at and up to t=60 months. This plot can be read by fixing a point along the x-axis and viewing the proportion of the y-axis corresponding to each of the three states. These correspond to the predicted state occupancy probabilities for that time t, and these probabilities sum to 1. Subplot 3 provides legend information along with the predicted proportion of patients in the different plotted states at the specified time horizon (in this case, 60 months). Subplot 4 provides additional legend information to help users relate the plotted states to the multistate model structure in **Figure 1**.

Figure 4 shows predictions for the same set of patient characteristics but with more detail. In this plot, we show proportions (1) - (8) directly. When we merge states corresponding to death from different causes (e.g. death without prior recurrence, death with prior locoregional recurrence or metastasis, and death among persistent patients), we recover the same proportion of patients predicted to have died by 60 months from **Figure 3**.

Figure 3: Example output

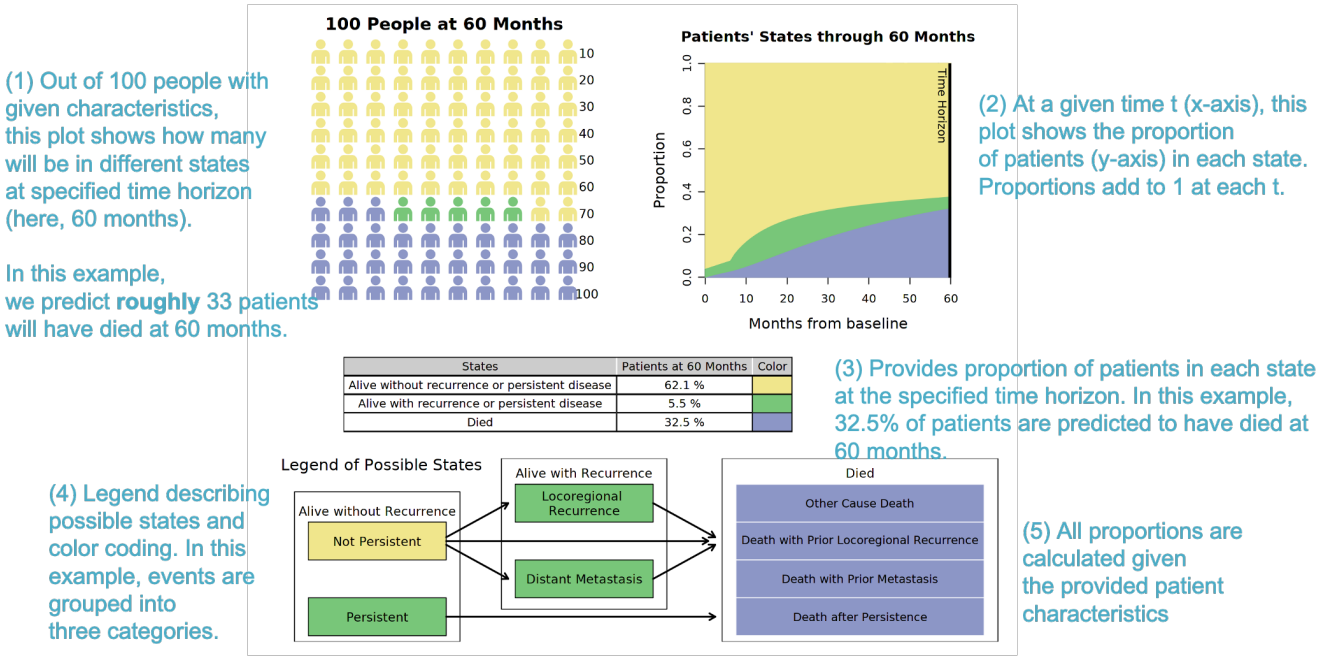


Figure 4: Example output with greater detail