Validation of a Transfusion Prediction Model in Head and Neck Cancer Surgery

Nadia L. Krupp, BS; Gregory Weinstein, MD; Ara Chalian, MD; Jesse A. Berlin, ScD; Patricia Wolf, BS; Randal S. Weber, MD

Background: Allogeneic transfusions are necessary in 14% to 80% of patients undergoing major head and neck cancer surgery. Defining the risk for receiving allogeneic transfusion allows for informed decisions regarding appropriateness of type and crossmatch, preoperative autologous blood donation, and priming with erythropoietin. Based on logistic regression analysis of transfusion risk factors in 438 patients, we developed a transfusion prediction risk assessment (TPRA) model to determine the need for transfusion based on the preoperative hemoglobin value, tumor stage, and need for flap reconstruction.

Objective: To examine the utility of this TPRA model in clinical practice by assessing the performance of the model in a validation set of patients.

Methods: Between 1996 and 1999, 125 consecutive patients entered into a clinical care pathway underwent major surgical procedures. The ability of the model to discriminate between patients requiring and those not requiring transfusion was assessed using the area under the receiver operating characteristic curve. The agreement between actual and predicted risks was tested using the goodness-of-fit statistic.

Results: The overall transfusion rate was 25%. A 1-U transfusion was required in 7 patients, and multiple units were necessary for 24 patients. Flap reconstruction was required in 63 patients, 44 patients had preoperative anemia by normative values, and 64 had T3/T4 tumors. Among the low-risk non-T3/T4 patients whose preoperative hemoglobin level was normal, the actual/predicted transfusion rate without flap reconstruction was 10%/2%. For high-risk patients with T3/T4 tumors, anemia, and flap reconstruction, the actual/predicted transfusion rate was 43%/65%. The area under the receiver operating characteristic curve was 0.72. The goodness-of-fit statistic indicated lack of fit of the original model, but a recalibrated model fit the observed data well.

Conclusions: In general, the TPRA model identifies patients at low or high risk for allogeneic transfusion and provides guidelines for preoperative counseling regarding the risk of receiving a transfusion. Knowledge of a patient’s risk can help direct cost-effective utilization of type and crossmatch, preoperative autologous blood donation, and preoperative priming with erythropoietin.


Although allogeneic transfusion has become increasingly safe, the risks of transmissible disease and transfusion reaction have not been entirely eliminated. Strategies have been sought for the minimization of exposure to allogeneic blood, without compromising patient safety. To this end, there has been growing interest in alternatives to allogeneic transfusion. Options include type and crossmatch (T&C), preoperative autologous blood donation (PABD), and priming with erythropoietin. When the risk of perioperative transfusion is 15% or higher, preoperative planning is necessary to ensure that the patient’s blood requirements are met through one or more of these methods.

Optimal utilization of these techniques requires identification of patients who would benefit most from their use. Over 50% of donated autologous blood units collected are not transfused, and blood collected from cancer patients cannot be transfused into another patient because of the risk of cancer cells surviving and proliferating in the recipient. Erythropoietin is costly, and it is therefore impractical for use in most patients. In an effort to minimize waste and costs, Dulguerov et al stratified patients into low-, moderate-, and high-risk groups based on the type of procedure planned. However, as the only variable was the proposed pro-
The transfusion prediction and risk assessment (TPRA) model was developed with the intent of providing a meaningful range of risk that a particular patient would require perioperative transfusion. It was based on prospective data for 438 patients who underwent surgery for head and neck cancer over a 4-year period. Various preoperative patient factors were evaluated. For each variable, a binary designation was adopted to simplify analysis. For example, stage was broken down into advanced stage (T3 or T4 = 1) vs less advanced stage (T1, T2, not staged, or recurrent disease = 0). Logistic regression was used to analyze each variable with respect to transfusion rate. Tumor stage, need for a flap, and preoperative hemoglobin level were the 3 strongest independent variables that emerged from univariate and multivariate analysis.

The model generated from multivariate analysis was as follows: probability = \( \frac{\exp(u)}{1+\exp(u)} \), where \( u = -3.989 + 2.120(\text{flap}) + 1.710(\text{preoperative hemoglobin}) + 0.775(\text{stage}) \). The variables took the following values: flap (0 = no, 1 = yes), preoperative hemoglobin level (0 = normal or above normal, 1 = below normal), stage (0 = not T3/T4, 1 = T3/T4). The \( P \) value for the goodness-of-fit statistic was .78, indicating a good fit of the model to the test set.

The model was used to generate predicted transfusion rates for the 8 risk categories (Table 1), ranging from 2% to 65%. These estimates enable a more informed decision regarding the need for having blood cross-matched preoperatively, and the utility of performing PABD. Ultimately, the goal is to minimize exposure to allogeneic blood while containing costs and waste of blood units.

The purpose of this study was to determine the accuracy and discrimination of the TPRA model by analyzing transfusion requirements in a validation set of patients.

### METHODS

#### VALIDATION SET PATIENT DATA ACQUISITION

This study protocol was reviewed and approved by the University of Pennsylvania institutional review board. All patients at the Hospital of the University of Pennsylvania undergoing the following procedures, concurrent with tracheostomy, were placed on the head and neck cancer pathway: head and neck free flap reconstruction, total or partial laryngectomy, major intraoral resection, composite resection, or neck dissection. The following data were obtained from the head and neck cancer pathway patients treated between 1996 and 1999: primary site, TNM stage, use of a flap, sex, age, prior radiotherapy, prior chemotherapy, and surgical procedure. We also collected data on the number of units transfused up to 30 days postoperatively and preoperative hemoglobin values.

Of the 125 consecutively eligible pathway patients, 3 patients were excluded: 2 because of missing preoperative hemoglobin values and 1 because of excessive heparin administration leading to hemorrhage and an 8-U transfusion requirement. Thus, 122 patients form the validation set for this study.

### DEFINITIONS AND GROUPS

The identical variables used to develop the TPRA model were examined in the validation set. Perioperative transfusion was defined as a transfusion required intraoperatively or within 30 days postoperatively. Hemoglobin groups were based on the following normal ranges: male, 14.0 to 18.0 g/dL; female, 12.0 to 16.0 g/dL.

Statistical significance was defined as \( P < .05 \). Data were organized into the same binary system that was used for the test set. Age and neck stage did not have binary designations and, therefore, retained the group designations used for the test set.

### DATA ANALYSIS

Each variable was analyzed by univariate analysis. Transfusion rates were calculated for each of the 8 risk groups designated in the model, defined by preoperative hemoglobin value, the use of a flap, and primary tumor stage. The coefficients from the original model were used to calculate the predicted probabilities for each patient. The transfusion rates for the validation set were compared with rates predicted by the model.

To evaluate the fit of the model to the test set and validation set, a \( \chi^2 \) goodness-of-fit statistic was calculated based on the observed vs expected probabilities (and the corresponding cell frequencies). For the 8 combinations of patient and surgical characteristics, the resulting \( \chi^2 \) statistic (with 6 df) assesses the calibration of the model, ie, how close the predicted probabilities from the model are to the observed risks of transfusion. A \( \chi^2 P < .05 \) represented a poor fit of the model to the data.

To measure the ability of the model to discriminate between low- and high-risk patients, the logistic regression results were used to generate the area under the receiver operating characteristic (ROC) curve. The ROC areas were calculated for the test set and validation set. An area of greater than 0.75 was considered representative of very good discrimination, and an area of 0.5 indicated that the test was equivalent to random guessing.

The original model was recalibrated to accommodate the different (higher) observed frequency of transfusions in the validation data (23%) as compared with the original development data (12%). The method used was described by DeLong and colleagues and applied to coronary bypass data. This technique involves “modeling” the original risk scores. Specifically, the method fits a secondary logistic regression model in the validation sample using the risk score calculated from the original model as the single predictor variable in the new model. This generates a new prediction equation, of the form “\( y = a + bx \)” that requires only knowledge of the predicted probability derived from the original model, where \( y \) is the recalibrated risk score (equivalent to “\( u \)” in the equation above) and \( x \) is the original probability. The new intercept term (\( a \)) corrects for differences in baseline risk between the 2 populations, while the new coefficient (\( b \)), a constant by which each of the original model coefficients is multiplied, recalibrates the original risk score it-
The overall transfusion rate was 31 (25%) of 122 patients. Two units were required in 18 of 32 patients transfused, and 1 U was required in 7 patients. The number of patients requiring 3, 4, and 5 U were 3, 2, and 1, respectively. The most common primary site was the oral cavity (34.4%), followed by the larynx (27.0%). The most common tumor stage was T4 (42.6%), followed by T2 (24.6%), and the most common neck stage was N0 (48.4%), followed by N2b (18.0%).

Transfusion rates were determined for each of the 12 preoperative variables. By univariate analysis, only the use of a flap was statistically significant (P < .001) (Table 2). Tumor stage approached significance (P = .08) and preoperative hemoglobin value, treated as “normal or above normal” vs “below normal,” was not a statistically significant variable (P = .32). However, the mean hemoglobin values for the transfused and nontransfused groups were 13.1 g/dL and 14.2 g/dL, respectively (P = .002). Transfusion rates for each of the 8 risk groups in our patient population are shown in Table 3. The transfusion rates were lowest in groups with normal preoperative hemoglobin values that did not require flap reconstruction, with a rate of 7% for T3/T4 tumors and 10% for non-T3/T4 tumors. The groups with T3/T4 tumors that required flap reconstruction had the highest rate of transfusion, with a rate of 49% for those with normal preoperative hemoglobin values and 43% for those with preoperative anemia.

The transfusion rates of the validation set followed the same general trend as predicted by the TPRA model, with a good correspondence of low-, intermediate-, and high-risk groups (Figure). The goodness-of-fit statistic indicated quite a good fit of the model to the test set data (χ² P = .78), but the closeness of the fit substantially declined when applying the model to the validation set (χ² P = .001). The ROC area also fell slightly from 0.84 (95% confidence interval, 0.77-0.91), indicating a very good fit of the model to the test set data, to 0.72 (95% confidence interval, 0.62-0.81) for the validation set. It should be noted, however, that there is substantial overlap of the confidence intervals around the ROC areas from the 2 data sets.

The prediction equation for the recalibrated model works similarly to the equation given in the introduction. Specifically, the recalibrated version of the equation was: Revised u = u* = -1.793 + 2.785(predicted probability from original model). Then the revised probability is calculated, as above, as exp(u*) /[1 + exp(u*)].

The recalibrated model predicted higher transfusion rates in the low-risk groups and lower rates in the high-risk groups compared with the original model, though the same overall trend in transfusion risk remained (Table 3).
The fit of the model improved dramatically with recalibration \(\chi^2 P = .27\) and the ROC area remained at 0.72 since the recalibration process did not affect the shape of the ROC curve. The group with T3/T4 tumors, flap reconstruction, and normal hemoglobin values emerged as an outlier following recalibration (Figure), with a transfusion rate nearly double of that predicted and almost identical to the corresponding group with preoperative anemia.

When a regression model was derived using data from the validation set and the same variables as in the original model, the coefficients all moved closer to zero (Table 4). The coefficient for hemoglobin had the most marked decrease from the original model, approaching zero, while it had been a strong predictor of transfusion rate, the transfusion rates trended similarly across the 8 risk categories. In general, one expects some deterioration in the performance of a model when comparing the fit in the test set with the fit in a validation set, despite a decrease in goodness of fit, the ROC area was well preserved, thus indicating a retained ability to distinguish between high- and low-risk patients. Since the test set and validation set data were obtained from 2 different institutions, this speaks to the generalizability of the model to other institutions.

It should be noted that in everyday practice, quoting a range of risk is more clinically relevant than a precise percentage. Therefore, in our practice and based on these data, we stratified patients into 3 larger risk groups (<15%, 15%-24%, and ≥25%). Either the original or recalibrated models could be used to determine placement of a patient in a risk group. For either model, the high-risk group included (1) patients with preoperative anemia who would require a flap and (2) patients with normal preoperative hemoglobin values and T3/T4 tumors who would require flap reconstruction. Patients with preoperative anemia, T3/T4 tumors without flap reconstruction remained in the intermediate-risk group before and after recalibration. The remaining groups were originally in the low-risk group and crossed into the intermediate-risk group with recalibration.

### VALIDATION OF TPRA MODEL

The overall transfusion rate was 25% for the validation set, compared with 12% for the test set. The reason for the more than doubled rate of transfusions in the validation set is unclear. It is most likely reflective of the fact that the 2 studies were performed at different institutions. As a result, the patient population, surgical technique, and triggers for transfusion were slightly different between the 2 data sets. In addition, transfusion practices of 1995 and 1999 may have changed on a more global level. Improved techniques in screening blood products for transmissible disease may have led to more liberal transfusion practices toward the end of the decade.

The 3 variables used in the model were the only ones that displayed significance on univariate analysis, though to varying degrees. The use of a flap was highly significant, tumor stage approached significance, and while the dichotomous version of the hemoglobin variable was not significant on univariate analysis, the mean values were statistically different for the transfused and nontransfused groups. The absence of other statistically significant variables supports the inclusion of use of a flap, tumor stage, and preoperative hemoglobin value as the only relevant variables in the model.

Despite the obvious difference in the overall transfusion rate, the transfusion rates trended similarly across the 8 risk categories. In general, one expects some deterioration in the performance of a model when comparing the fit in the test set with the fit in a validation set, and such was the case for the TPRA model. However, despite a decrease in goodness of fit, the ROC area was well preserved, thus indicating a retained ability to distinguish between high- and low-risk patients. Since the test set and validation set data were obtained from 2 different institutions, this speaks to the generalizability of the model to other institutions.

---

**Table 3. Actual Transfusion Rates Seen in Validation Set and Rates Predicted by Recalibrated Model***

<table>
<thead>
<tr>
<th>Preoperative Hemoglobin</th>
<th>Not T3/T4</th>
<th>T3/T4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Flap</td>
<td>Flap</td>
</tr>
<tr>
<td>Normal</td>
<td>0.10 (3/29)</td>
<td>0.15 (3/12)</td>
</tr>
<tr>
<td>&lt;Normal</td>
<td>0.11 (1/9)</td>
<td>0.18 (3/8)</td>
</tr>
</tbody>
</table>

*Actual transfusion rates are shown in plain text with numbers transfused/total number in group in parentheses. Rates predicted by recalibrated model are shown in italics.

**Table 4. Comparison of Model Coefficients for Test Set and Validation Set**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient From Original Model</th>
<th>Coefficient Based on Validation Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Normal hemoglobin</td>
<td>1.710</td>
<td>0.174</td>
</tr>
<tr>
<td>Stage T3/T4</td>
<td>0.775</td>
<td>0.389</td>
</tr>
<tr>
<td>Flap</td>
<td>2.120</td>
<td>1.609</td>
</tr>
<tr>
<td>Constant</td>
<td>−3.989</td>
<td>−2.378</td>
</tr>
</tbody>
</table>

©2003 American Medical Association. All rights reserved.

Downloaded From: https://archotol.jamanetwork.com/ by a Non-Human Traffic (NHT) User on 04/30/2019

RECALIBRATION FOR OVERALL TRANSFUSION RATE

With recalibration, the greatest changes in predicted transfusion rate occurred in lower-risk groups. Recalibration reversed the underprediction observed in low-risk groups using the original model and resulted in a slight overprediction in these groups. The overprediction observed in high-risk groups was almost completely eliminated by recalibration. Statistical goodness of fit of the model also improved with recalibration, suggesting that the overall transfusion rate was largely responsible for the decline in fit of the original model from the test set to the validation set.

The recalibration process was necessary to investigate the effect that the overall transfusion rate had on the distribution of risk across the 8 groups and to enable accurate statistical analysis of the strength of the model. However, the ability of the model to discriminate between low- and high-risk patients is unchanged by recalibration. Since the clinical usefulness of this model rests on this ability, and not a precise estimate of transfusion risk, recalibration to correct for transfusion rate at another institution would likely add little clinically useful information.

PATIENT COUNSELING

By using either the original or recalibrated model, designations of low (<15%), intermediate (15%-24%), or high (≥25%) risk may be made for each patient type. Which model is used may be determined by deciding which overall transfusion rate (12% for the original, 25% for the recalibrated) is most similar to the practice in question. For the purposes of discussion here, actual transfusion rates for our validation set are used, since they are presumed to be most representative of our institution. Importantly, the actual rates can be tabulated for ease of use, eliminating the need to perform the arithmetic manipulations.

By referring to Table 3, a patient at our institution with a normal preoperative hemoglobin value scheduled to undergo resection of a T2 laryngeal tumor without flap reconstruction would have a transfusion risk of approximately 10%, placing him or her in the lowest risk group. For patients in the lowest risk group for requiring allogeneic transfusion, a preoperative type and screen (T&S) may be all that is required preoperatively. The T&S would rule out any major incompatibilities and should blood become necessary, a T&C can be completed in 20 to 30 minutes.

For patients at intermediate risk for requiring allogeneic transfusion, a T&C may be more appropriate because it reserves in advance the requested number of units for a specific patient. Since the cross-matched blood is held for a designated period of time and cannot be administered to another patient, using T&C for low-risk patients increases the likelihood that some units may expire and need to be discarded. Furthermore, routine use of T&C requires the blood bank to keep inventories high and increases costs associated with perioperative transfusion management. The TPRA model can effectively identify patients who require T&E vs T&C, and therefore help to minimize waste and costs associated with routine T&C.

By contrast, a patient with a low preoperative hemoglobin value who is scheduled to undergo resection of a T2 squamous cell carcinoma with flap reconstruction would have a risk of approximately 38% of requiring a transfusion, placing him or her in the highest risk group. This patient may be counseled to consider PABD or directed donation. In addition to the absence of risk of transmissible disease, PABD has been shown to decrease the risk of cancer recurrence, though this remains a controversial topic. Of course, each patient situation is unique, and this model would best be used as a tool to better educate patients as to their risk, and thus facilitate an informed decision. It is not intended to strictly dictate the use of transfusion options for every patient. Some patients will not be comfortable with a 5% to 10% risk of receiving allogeneic transfusion, and PABD may be a better option for these patients to ensure peace of mind.

PREOPERATIVE ANEMIA AND PRIMING WITH ERYTHROPOIETIN

The variable preoperative hemoglobin level showed a decrease in predictive power far greater than the other 2 variables, and in high-risk groups appeared to have little or no effect on transfusion risk. The reason for this is unclear. The increasing use of erythropoietin in recent years is one possible explanation for the loss of importance of preoperative hemoglobin levels in the validation set. In some patients undergoing head and neck cancer surgery, the use of erythropoietin has been reported to have eliminated the need for allogeneic transfusion altogether. Any effect that erythropoietin use may have had on our results could not be evaluated owing to the incompleteness of available data in both the test set and validation set for this variable.

As the high cost of erythropoietin makes its use in all patients preparing to undergo head and neck surgery impractical, the TPRA model can be used to predict which patients might benefit from its use. Patients with a hemoglobin level below normal and who will require a flap reconstruction regardless of tumor stage have a transfusion risk of approximately 40% to 60%. A risk this high would warrant consideration of PABD, but the subnormal hemoglobin value would preclude its use. Raising the hemoglobin level with erythropoietin may enable these very high-risk patients either to avoid transfusion by correcting their anemia or to undergo PABD. The relatively small number of these patients (29 of 122) means that cost would be contained using this strategy, compared with using erythropoietin for all patients with subnormal hemoglobin levels scheduled to undergo head and neck cancer surgery.

FUTURE STUDY

To circumvent the inherent subjectivity in transfusion practices, future studies may use precise operative estimated blood loss as the outcome of interest instead of transfusion. Determination of estimated blood loss would not be subject to global changes in transfusion prac-
Despite the presence of some discrepancies, the TPRA model retained its ability to discriminate between low-, intermediate-, and high-risk patients when tested in a distinct patient population, in the face of institutional variation. The model is a useful guide for counseling patients regarding the probability for requiring a transfusion, and guidelines derived from this model may facilitate decision making. This model is unique in that it can provide an individual patient with an assessment of their risk of requiring transfusion. In order to provide more accurate estimates of transfusion risk, the TPRA model may be recalibrated to account for institutional differences in overall transfusion rate, though this is cumbersome and unlikely to contribute clinically useful information. Instead, using overall transfusion rate as a guide, either the original model or the recalibrated model can be used to estimate transfusion risk for each of the 8 patient categories. The clinician need not use the complicated mathematical model, as the risk values may be read directly from a table showing the probabilities within the 8 categories. Patients may then be placed into a low-, intermediate-, or high-risk group, as follows:

### Risk Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>T3/T4, flap, any preoperative hemoglobin value; not T3/T4, flap, normal preoperative hemoglobin value</td>
</tr>
<tr>
<td>Intermediate</td>
<td>T3/T4, no flap, preoperative anemia</td>
</tr>
<tr>
<td>Low</td>
<td>All other groups*</td>
</tr>
</tbody>
</table>

*Unless overall transfusion rate approximately 25%, then all intermediate.

The model’s placement of the patient into 1 of these 3 groups will permit the physician and patient to make informed decisions that will minimize exposure to allogeneic transfusion and more cost-effectively pursue appropriate crossmatching of blood units, PABD, and administration of erythropoietin. Flexible guidelines may be similar to the following: PABD, erythropoietin, or directed donation for high-risk patients; T&C for intermediate-risk patients; and T&S for low-risk patients.

Submitted for publication January 21, 2003; final revision received March 24, 2003; accepted April 15, 2003.

Corresponding author and reprints: Randal S. Weber, MD, Department of Head and Neck Surgery, Unit 441, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030 (e-mail: RSWeber@mdanderson.org).

### REFERENCES


