Models to Predict the Risk of Infection After Surgery in Pancreatic Cancer Patients

Jolene Ranek
Brooke Bell
Jeffrey Antunes
Mentor: Dr. Brian Smith
2 glands: exocrine and endocrine
- Enzymes, islets
- Unknown cause for cancer
- Biopsy
- Stages 1-4
  - 1 – cancer is limited to pancreas
  - 2 - spread to lymph nodes
  - 3 - spread to blood vessels
  - 4 - metastasized to nearby organs

“In 2014, it's estimated over 46,000 people in the U.S. will be diagnosed with pancreatic cancer...”
“Cancer of the pancreas barely makes the top 10 most common cancers in the U.S. However, pancreatic cancer's tendency to spread silently before diagnosis makes it the fourth deadliest cancer diagnosis with more than 39,000 people expected to die of the disease in 2014.”

- Asymptomatic until the later stages
- 1-year survival rate is 20%, and the five-year rate is 6%

http://www.pancreatic.org/site/c.htJYJ8MPlwE/b.891917/k.5123/Prognosis_of_Pancreatic_Cancer.htm
Overview of Treatment

- Surgery in combination with chemotherapy and radiation
- Surgery may be done to remove the cancerous tumor or lessen symptoms
  - Whipple - most common operation to remove a cancer of the exocrine pancreas; surgeon removes the head of the pancreas and nearby structures
  - Distal - surgeon removes only the tail of the pancreas and the spleen
- Chemotherapy and radiation are often given together- used to slow growth of pancreatic cancer
- In those cases where resection can be performed, the average survival rate is 18 to 20 months

http://www.pancreatic.org/site/c.htJYJ8MPIwE/b.891917/k.5123/Prognosis_of_Pancreatic_Cancer.htm
Create predictive models that can be used to assess the risk of surgical complications
Overview of Variables

- American College of Surgeons National Surgical Quality Improvement Program (NSQIP)
  - 150+ variables for 23,000+ patients
  - Preoperative risk factors
  - Intraoperative variables
Examined 50+ variables
- Continuous and categorical
- Ex: Age, height, race, prior operation in past 30 days, history of transient ischemic attacks, type of procedure

Examined 3 separate outcomes
- Surgical Site Infections
- Data accounts for infections that occurred within 30 days after surgery
Surgery Outcomes

- Surgical Site Infections (SSI)
  - Superficial Incisional SSI: skin or subcutaneous tissue
  - Deep Incisional SSI: deep soft tissue
  - Organ Space SSI: organs or spaces
Summary Statistics

Distribution of Ages

Distribution of Races

Disseminated Cancer
## Summary Statistics

<table>
<thead>
<tr>
<th>OUTCOME</th>
<th>NO INFECTION</th>
<th>INFECTION</th>
<th>PERCENT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superficial Incisional SSI</td>
<td>21563</td>
<td>1934</td>
<td>8.23</td>
</tr>
<tr>
<td>Deep Incisional SSI</td>
<td>23008</td>
<td>489</td>
<td>2.08</td>
</tr>
<tr>
<td>Organ Space SSI</td>
<td>21076</td>
<td>2421</td>
<td>10.30</td>
</tr>
</tbody>
</table>
Univariate Analysis

- Conducted univariate analysis for each variable to identify the most statistically significant based off of their p-values
- Used these variables as an initial starting point for our multivariate model
Multivariate Analysis

- Built a multivariate regression model to predict outcomes
- Identified a set of variables that carry the most information for predicting the outcome
- Used to make a logistic regression model
Logistic Regression

$Y \sim \text{Binomial } (1, \pi(x))$

\[
\ln \left( \frac{\pi(x)}{1-\pi(x)} \right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n
\]

- $Y$ is coded as 1 for infection and 0 for no infection
- Assumed to have a binomial distribution
- $\pi$ is the probability of the infection modeled as a predictor variable $x$
- $\beta$ is the effect of predictor on infection probability
To construct our models, we took the data and randomly divided the patients into two equal groups

- Training group – the group we used to construct our final models
- Test group – an entirely independent group that we applied our model to assess its predictive performance
- We used the area under the ROC curve to evaluate the predictive performance of the models
Surgery Outcome: Superficial Incisional SSI

http://blog.sisfirst.com/blog/bid/147767/What-s-the-Secret-to-a-High-Performing-Surgical-Department
Step-wise Model Selection

- Applied step-wise model selection on the variables that were statistically significant in the univariate model at the 0.5 alpha level
- Created a model by backward step-wise selection
- Final model had 10 variables
Patient characteristics that influence the risk of infection in superficial tissue include:

- Age
- Weight
- Previous angioplasty (PCI)
- Pre-op bun
- Pre-op serum albumin
- Pre-op phosphatase alkaline
- Pre-op platelet count
- Pre-op normalized PT values
- Radio therapy
- Bleeding disorder
Predictive Power

ROC curve for training set
Area under the curve is 0.5888

ROC curve for test set
Area under the curve is 0.5783
# Odds Ratios

<table>
<thead>
<tr>
<th>Variable</th>
<th>SD * Normal Levels **</th>
<th>Odds Ratios</th>
<th>Lower Bounded 95 % CI</th>
<th>Upper Bounded 95 % CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>13.40 *</td>
<td>1.11</td>
<td>2.73</td>
<td>3.39</td>
</tr>
<tr>
<td>Weight</td>
<td>43.36 *</td>
<td>1.17</td>
<td>2.92</td>
<td>3.55</td>
</tr>
<tr>
<td>Previous angioplasty (PCI)</td>
<td>1</td>
<td>1.32</td>
<td>2.60</td>
<td>5.41</td>
</tr>
<tr>
<td>Pre-op bun</td>
<td>13 **</td>
<td>0.84</td>
<td>1.91</td>
<td>2.79</td>
</tr>
<tr>
<td>Pre-op serum albumin</td>
<td>2 **</td>
<td>0.76</td>
<td>1.56</td>
<td>2.94</td>
</tr>
<tr>
<td>Pre-op phosphatase alkaline</td>
<td>103 **</td>
<td>1.07</td>
<td>2.74</td>
<td>3.08</td>
</tr>
<tr>
<td>Pre-op platelet count</td>
<td>250 **</td>
<td>1.30</td>
<td>2.91</td>
<td>4.66</td>
</tr>
<tr>
<td>Pre-op normalized PT values</td>
<td>0.30 **</td>
<td>0.78</td>
<td>1.73</td>
<td>2.75</td>
</tr>
<tr>
<td>Radio therapy - No</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Radio therapy - Yes</td>
<td>1</td>
<td>1.57</td>
<td>2.99</td>
<td>7.74</td>
</tr>
<tr>
<td>Bleeding Disorder - No</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bleeding disorder - Yes</td>
<td>1</td>
<td>1.52</td>
<td>2.89</td>
<td>7.29</td>
</tr>
</tbody>
</table>
Surgery Outcome: 
Organ Space SSI
Significant Variables

- Initial model had 32 variables and after implementing stepwise model selection, final model had 15 variables
  - Weight
  - Sex
  - CPT
  - Dyspnea
  - Pre – operative serum albumin
  - Pre – operative alkaline phosphatase
  - Pre – operative hematocrit
  - Disseminated cancer
  - Weight loss
  - Bleeding disorder
  - Emergency case
  - Height
  - Diabetes
  - Previous cardiac surgery
  - Pre – operative serum sodium
Predictive Power

ROC curve for training set

Area under the curve (AUC) is 0.6195

ROC curve for test set

Area under the curve (AUC) is 0.5836
# Odds Ratios

<table>
<thead>
<tr>
<th>Variable</th>
<th>SD* Normal levels**</th>
<th>Odds Ratio</th>
<th>Lower Bounded 95% CI</th>
<th>Upper Bounded 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyspnea Moderate Exertion</td>
<td>1</td>
<td>2.73</td>
<td>5.36</td>
<td>43.59</td>
</tr>
<tr>
<td>Dyspnea - No</td>
<td>1</td>
<td>1.70</td>
<td>1.96</td>
<td>15.43</td>
</tr>
<tr>
<td>Dyspnea - At Rest</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pre-operative serum albumin</td>
<td>2**</td>
<td>0.62</td>
<td>1.46</td>
<td>2.38</td>
</tr>
<tr>
<td>Disseminated cancer - Yes</td>
<td>1</td>
<td>1.67</td>
<td>3.90</td>
<td>7.17</td>
</tr>
<tr>
<td>Disseminated cancer - No</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Previous cardiac surgery - Yes</td>
<td>1</td>
<td>0.66</td>
<td>1.32</td>
<td>2.85</td>
</tr>
<tr>
<td>Previous cardiac surgery - No</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Surgery Outcome: Deep Incisional SSI
## Significant Variables

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>0.0174</td>
</tr>
<tr>
<td>Weight</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sex</td>
<td>0.0107</td>
</tr>
<tr>
<td>Ethnicity Hispanic</td>
<td>0.0322</td>
</tr>
<tr>
<td>CPT code</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Pre-operative serum albumin</td>
<td>0.0029</td>
</tr>
<tr>
<td>Pre-operative bilirubin</td>
<td>0.0213</td>
</tr>
<tr>
<td>Pre-operative PT</td>
<td>0.0396</td>
</tr>
<tr>
<td>Probability of morbidity</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>ASA classification</td>
<td>0.0146</td>
</tr>
<tr>
<td>Bleeding disorder</td>
<td>0.0217</td>
</tr>
</tbody>
</table>
Model Selection

- Backward selection
  - Alpha level = 0.05
  - Began with all variables of interest
  - Dropped least significant variable
  - Continued process until all remaining variables were statistically significant at $\alpha = 0.10$

- Final model included 6 variables
  - Weight, Hispanic ethnicity, type of surgery (CPT), pre-operative serum albumin, pre-operative PT, bleeding disorders
Predictive Power

Training set of data

Area under the curve (AUC) = 0.6618

Test set of data

Area under the curve (AUC) = 0.5814
## Odds Ratios

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>Lower Bounded 95% CI</th>
<th>Upper Bounded 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>1.32</td>
<td>1.12</td>
<td>1.18</td>
</tr>
<tr>
<td>Hispanic Ethnicity – Not Reported</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hispanic Ethnicity - No</td>
<td>1.61</td>
<td>1.05</td>
<td>2.49</td>
</tr>
<tr>
<td>Ethnicity Hispanic - Yes</td>
<td>1.25</td>
<td>0.37</td>
<td>4.19</td>
</tr>
<tr>
<td>CPT - Other</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CPT - Distal</td>
<td>1.87</td>
<td>0.55</td>
<td>6.39</td>
</tr>
<tr>
<td>CPT - Whipple</td>
<td>3.57</td>
<td>0.49</td>
<td>25.84</td>
</tr>
<tr>
<td>CPT - Total</td>
<td>2.47</td>
<td>0.22</td>
<td>27.64</td>
</tr>
<tr>
<td>Pre-operative serum albumin</td>
<td>0.74</td>
<td>0.57</td>
<td>0.96</td>
</tr>
<tr>
<td>Pre-operative PT</td>
<td>0.89</td>
<td>0.81</td>
<td>0.98</td>
</tr>
<tr>
<td>Bleeding disorders - Yes</td>
<td>2.29</td>
<td>1.13</td>
<td>4.64</td>
</tr>
<tr>
<td>Bleeding disorders - No</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Conclusion

- A good predictive model has an AUC of at least 0.75 or higher
  - Shows that our models poorly predict the risk of infection
  - Although these models do not predict risk well, the results can still be used for future research
  - Possibility to combine all forms of infection and create a model based off of that
Thank You
Acknowledgements

Dr. Brian Smith
Dr. Gideon Zamba
Terry Kirk
Joe Moen & Melissa Pugh
John VanBuren
Biostatistics Graduate Students
Biostatistics Faculty and Staff
University of Iowa College of Public Health
National Institute of Health
Questions