**Objective:** To determine which of the 5 preoperative variables, age at diagnosis (age), level of serum acid phosphates (acid), Xray reading (Xray), pathology reading (grade), and tumor stage (stage), are predictive of nodal involvement.

**Methods:** Logistic regression models were fitted to the supplied data via stepwise selection. Interactions involving binary variables were included in model selection but interactions involving numeric variables were not considered. The significance of each individual variable was assessed by Wald test and Likelihood ratio test. Competing models were compared to the model selected from a stepwise procedure using criteria, such as AIC. The Hosmer-Lemeshow test, deviance, Pearson chi-squared, and analysis of patient residuals were used to assess whether the selected model adequately fits the data.

**Results:** There were significant differences in the preoperative variables between the patients with and without nodal involvement except for age (p=0.31) and acid (p=0.08)(Table 1). The patients with serious case in grade, Xray, and stage were more likely to have nodal involvement. The resultant model selected from the stepwise procedure showed that Xray and stage remained an independent predictor of nodal involvement (Table 2). The greatest effect size on nodal involvement was Xray. For the patients who had a positive or more serious finding in their Xray result, their odds of having nodal involvement was 8.33 times the odds for those patients who had a negative or less serious finding in their Xray result. Another significant factor was tumor stage. The odds of having nodal involvement in patients who were in more serious tumor stage was 4.90 times the odds for those patients who were in less serious stage. Table 3 showed the effect size of the 3 predictors not selected by the stepwise procedure on nodal involvement when the association was assessed individually as well as when the association was assessed with Xray and stage included in the model. The magnitude of the association between grade and nodal involvement was attenuated when Xray and stage included in the model.

**Conclusions:** Patients with serious finding in Xray reading and tumor stage were more likely to have nodal involvement compared to patients with less serious finding in Xray reading and tumor stage.

### Table 1. Associations between Preoperative Variables and Nodal Involvement in 53 Patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Nodal Involvement</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Event (n=20)</td>
<td>No Event (n=33)</td>
<td>p Value</td>
<td></td>
</tr>
<tr>
<td>Age at diagnosis (age)</td>
<td>59 (45-68)</td>
<td>61 (49-68)</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Level of serum acid phosphates (acid)</td>
<td>74 (48-136)</td>
<td>55 (40-187)</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Xray reading (Xray)</td>
<td>11 (73.3)</td>
<td>4 (26.7)</td>
<td>0.0008</td>
<td></td>
</tr>
<tr>
<td>Pathology reading (grade)</td>
<td>11 (55.0)</td>
<td>9 (45.0)</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>Tumor stage (stage)</td>
<td>15 (55.6)</td>
<td>12 (44.4)</td>
<td>0.006</td>
<td></td>
</tr>
</tbody>
</table>

Data presented as number of patients (%) or median (range)

Associations between characteristics and event status were tested with the t test for continuous variables (age and acid) and the Chi-square test for discrete variables (Xray, grade, and stage)

### Table 2. Association of Nodal Involvement with Selected Predictors from Stepwise Regression

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Event</th>
<th>No Event</th>
<th>OR</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xray reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serious case</td>
<td>11</td>
<td>4</td>
<td>8.33</td>
<td>1.93-35.99</td>
<td>0.0045</td>
</tr>
<tr>
<td>Less serious case</td>
<td>9</td>
<td>29</td>
<td>ref</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serious case</td>
<td>15</td>
<td>12</td>
<td>4.90</td>
<td>1.24-19.30</td>
<td>0.0233</td>
</tr>
<tr>
<td>Less serious case</td>
<td>5</td>
<td>21</td>
<td>ref</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^1Number of patients had nodal involvement found at surgery.

OR = odds ratio; CI = confidence interval; ref = reference group

### Table 3. Association of Nodal Involvement for Predictors not selected in Stepwise Regression

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Unadjusted OR</th>
<th>95% CI</th>
<th>p Value</th>
<th>Adjusted with Xray and stage OR</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.95</td>
<td>0.87-1.04</td>
<td>0.300</td>
<td>0.94</td>
<td>0.85-1.05</td>
<td>0.267</td>
</tr>
</tbody>
</table>
### Model Selection

Model selection was performed by PROC LOGISTIC via stepwise with an entry and exit significance level of 0.1. Interactions involving binary variables (Xray, grade, stage) were included in model selection but interactions involving numeric variables (age, acid) were not considered. Binary variables were treated as dummy variables with the less serious case used as reference. Xray was selected by the stepwise algorithm followed by stage. Acid was entered into the model and finally removed when Xray and stage were in the model. The p value for age and for grade did not reach the entry significance level. Xray and stage remained independent predictors of nodal involvement. Two additional candidate models were further evaluated and compared with the selected model from the stepwise procedure. Model 1 was selected from the stepwise procedure which included 2 predictors: Xray and stage. Model 2 included 3 predictors: Xray, stage, and acid (the next most significant candidate variable for inclusion in the stepwise procedure). Model 3 was selected from the backward procedure which included Xray, stage, acid, grade, and the interaction term of grade and stage.

### Criteria considered in selecting the resultant model

- The AIC for Model 1 and Model 2 were 59.35, and 58.66 respectively, and were practically the same.
- Global test statistics showed when Model 1 was compared with the null model, the likelihood ratio statistic was 16.90 with 2 df, p value was 0.0002 (results can be found under "Testing Global Null Hypothesis" in appendix). When Model 2 was compared with null model, the likelihood ratio statistic was 19.59 with 3 df, p value was 0.0002. The result suggested that the both models were useful in predicting nodal involvement.
- When acid was included in the model with Xray and stage, the Wald test p value was 0.103. Comparing the models with and without acid, the likelihood ratio statistic was 2.69 with 1 df, p value was 0.1008. This result indicated that the addition of acid to the model did not significantly improve prediction of nodal involvement.

### Rationales for Selecting Model 1

Model 3 suggested the patients should be evaluated according to their grade and stage status because a significant interaction was found between grade and stage. However, given the small sample size (50% of the cells had only 5 or less patients if patients indeed were stratified according to grade and stage status), the ML estimate from Model 3 may be quite biased. Given this uncertainty, Model 3 was not preferred and it should be further validated in a larger data set. Model 1 is preferred because it has the minimal number of variables which is more likely to be numerically stable and is more easily generalized. Moreover, there is insufficient evidence to show Model 2 was significantly better than Model 1. Thus, Model 1 was selected.

### Fitted Model and Parameter Estimates

\[
\text{logit}(\hat{\pi}(x)) = \log(\hat{\pi}(x)/1-\hat{\pi}(x)) = -2.045 + 2.12 \text{Xray} + 1.59 \text{stage}
\]

where \(\hat{\pi}(x)\) is the probability of having nodal involvement in patients at surgery. This model satisfies homogeneous association. The effect size of Xray on nodal involvement is the same at each category of the tumor stage. Likewise, the effect size of tumor stage on nodal involvement is the same at each category of Xray.

Effect of Xray on the probability of having nodal involvement:

\[
\text{OR} = 8.33 \text{ (Wald 95\%CI 1.93 - 35.99, } p=0.0045). \text{ The odds of having nodal involvement in patients who had a positive or more serious state in their Xray was 8.33 times the odds of those patients who had a negative or less serious state in their Xray. Another way to interpret the effect of Xray on nodal involvement is that the odds of having nodal involvement for patients with a serious case in Xray were estimated to be between 1.93 and 35.99 times the odds of those with less serious case in Xray. With a low p value and the CI for the OR excluding unity, we conclude that there is a significant association between Xray and the nodal involvement in patients at surgery.}
\]

Effect of stage on the probability of having nodal involvement:
OR=4.90 (Wald 95%CI 1.24 - 19.30, p=0.0233). The odds of having nodal involvement in patients who were in serious stage was 4.90 times the odds of those patients who were in less serious stage. The 95% CI can be interpreted as follows. The odds of finding nodal involvement for patients who were in serious stage were estimated to be between 1.24 and 19.30 times the odds of those with less serious stage.

Effect of age on the probability of having nodal involvement:
Age was not significantly associated with nodal involvement when the association was assessed individually (OR=0.95, 95%CI: 0.87-1.04, p=0.30, Table 3) as well as when the association was assessed together with Xray and stage (adjusted OR=0.94, 95%CI: 0.85-1.05, p=0.267, Table 3)

Effect of acid on the probability of having nodal involvement:
Acid was not significantly associated with nodal involvement when the association was assessed individually (OR=1.02, 95%CI: 0.99-1.05, p=0.105, Table 3) as well as when the association was assessed together with Xray and stage (adjusted OR=1.02, 95%CI: 0.99-1.05, p=0.103, Table 3)

Effect of grade on the probability of having nodal involvement:
Although grade was significantly associated with nodal involvement when the association was assessed individually (OR=3.26, 95%CI: 1.01-10.48, p=0.047, Table 3), the magnitude of the association between grade and nodal involvement was attenuated when Xray and stage included in the model (OR=1.73, 95%CI: 0.42-7.14, p=0.0=447, Table 3),

Model Checking
- The Hosmer-Lemeshow statistic equal 10.90, with df=9 (p=0.28). This result indicated there is not evidence of an inadequate fit.
- The deviance $G^2 = 53.35$ and Pearson chi-squared=54.02, both with df = 50 have p-values of 0.35 and 0.43, respectively, and both show that we failed to reject Ho and so that the fit was reasonable.
- Figure 1 illustrated that the residuals of each patients was within ±3 although there were 3 mild outliers. Thus, no strong evidence of outlier based on the fitted model.

Discussion
Based on the 5 preoperative variables from 53 prostate cancer patients, the global test statistic showed that a model, which was built by stepwise regression including Xray and stage, was significantly associated with nodal involvement (p=0.0002). Patients with serious finding in Xray reading and tumor stage were more likely to have nodal involvement compared to patients with less serious finding in Xray reading and tumor stage. Deviance, Hosmer-Lemeshow, and Pearson chi-squared test all supported the appropriateness of the model. No outlier was identified when the residuals of each patient were examined.
Result of Stepwise Selection

Result of Backward Elimination

Result of the Association between Xray and Stage and Nodal Involvement Using PROC GENMOD

Result of the Association between Xray and Stage and Nodal Involvement Using PROC LOGISTIC

AIC, Deviance, Global Test Statistic and Pearson Chi-Square Test Result for a Model with Xray and Stage
AIC, Global Test Statistic Result for a Model with Xray, Acid, and Stage

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>72.252</td>
<td>58.660</td>
<td>Likelihood Ratio</td>
<td>19.5925</td>
<td>3</td>
<td>0.0002</td>
</tr>
<tr>
<td>SC</td>
<td>74.222</td>
<td>66.541</td>
<td>Score</td>
<td>17.6961</td>
<td>3</td>
<td>0.0005</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>70.282</td>
<td>50.660</td>
<td>Wald</td>
<td>12.4569</td>
<td>3</td>
<td>0.0060</td>
</tr>
</tbody>
</table>

Goodness of Fit: Hosmer-Lemeshow Test

Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7977</td>
<td>2</td>
<td>0.6711</td>
</tr>
</tbody>
</table>
**SAS CODE**

```sas
data cancer;
input xray grade stage age acid nodes;
datalines;
0 1 64 40 0
0 0 1 63 40 0
1 0 0 65 46 0
proc print data=cancer;
run;

**Stepwise and backward selection;**
proc logistic data=cancer descending;
model nodes=age acid xray|grade|stage/link=logit selection=stepwise slentry=0.1 slstay=0.1;
run;
proc logistic data=cancer descending;
model nodes=age acid xray|grade|stage/link=logit selection=backward slstay=0.1;
run;

**Compare and confirm the estimates of the resultant model between PROC LOGISTIC and PROC GENMOD;**
**Obtain the deviance, Likelihood ratio and Wald statistic;**
proc logistic data=cancer descending;
model nodes=xray stage/link=logit;
run;
proc genmod data=cancer descending;
model nodes=xray stage/link=logit dist=binomial;
run;

**Result for table 1 and 2;**
proc freq data=cancer;
tables nodes*grade nodes*xray nodes*stage /chisq;
run;
proc ttest data=cancer;
class nodes;
var acid age;
run;
proc univariate data=cancer;
class nodes;
var acid age;
run;

**Model comparsion;**
proc logistic data=cancer descending;
model nodes= xray stage/link=logit;
run;
proc logistic data=cancer descending;
model nodes= xray stage acid/link=logit;
run;

R code to generate plot:
res=plot(Patient, res, pch=16, ylab="Residual",xlab="Patients", cex.lab=1.5, ylim=c(-4,4),
xlim=c(0,54), col="darkred")
abline(h=0,lty="dashed",col="darkblue")
abline(h=3,lty="dashed",col="darkred")
abline(h=-3,lty="dashed",col="darkred")
```

---

Statistical Analysis Report of Nodal Involvement

Carmen Tong