Abstract:

Cancer of the prostate is a relatively common and virulent form of cancer from which nearly 17 percent of North American men will suffer and with a cure rate of only 5 percent. Due to the speedy progression of prostate cancer, early diagnostic measures are highly desired to inform and guide treatment decisions. The current study conducts an assay validation analysis comparing the gold standard, PSA, to circulating tumor cells (CTC) and circulating endothelial cells (CEC) as predictors of survival.

Data come from 231 men aged 45 to 92 who have been diagnosed with metastatic prostate cancer. Most of the patients have had their prostates removed, and have received varying treatments prior to entry in the current study. Relevant covariates include age, changes in CEC, changes in CTC and ECOG status, which is a measure of patient mobility and quality of life. CEC, CTC and PSA are measured as many follow-ups as possible, but the completeness of these data vary from patient to patient. The outcome of interest is death from any cause, however most deaths are cancer-related. Cox proportional hazards models were used for primary analysis.

I found that separate Cox proportional hazards models similarly specified using PSA and using CTC yielded nearly identical results suggesting that CTC can perform at least as well as PSA as a predictor of survival. Additional results using CEC counts have implications for treatment decision-making. Patients with an elevated baseline CEC may already be too ill to benefit from chemotherapy and may prefer less invasive, comfort-focused therapies with fewer side effects over aggressive, survival-focused therapies like chemo.

Introduction:

Cancer of the prostate is a relatively common and virulent form of cancer from which nearly 17 percent of North American men will suffer and with a cure rate of only 5 percent. Due to the speedy progression of prostate cancer, early diagnostic measures are highly desired to inform and guide treatment decisions. Historically, oncologists have used levels of prostate specific antigen (PSA) as a proxy for the severity of the cancer and as a predictor of patient survival, but recent literature suggests that PSA may not perform as well as once thought. The current study conducts an assay validation analysis comparing the gold standard, PSA, to circulating tumor cells (CTC) and circulating endothelial cells (CEC) as predictors of survival.

Circulating Tumor Cells:

Circulating tumor cells enter the blood and can be indicative of cancer severity. Typically, when a tumor is localized, the CTC count will range from 0 to 10. CTC values greater than 10 suggest metastatic cancer. Previous literature suggests that circulating tumor cell counts are associated with shorter survival.

Circulating Endothelial Cells:

Circulating endothelial cells have been used extensively in measuring vascular disease and morbidity and are only recently being considered as a potential predictor of cancer survival. One important characteristic of CECs is that they are very unspecific, meaning that they can be caused by

---

many factors and must be phenotyped to determine their root. It is unclear in the context of this study whether the CECs are being generated due to cancer related tissue morbidity, poor vascular health due to age, or some unobserved cause.

**Methods:**

Data come from 231 men aged 45 to 92 who have been diagnosed with metastatic prostate cancer. Most of the patients have had their prostates removed, and have received varying treatments prior to entry in the current study. Information pertaining to previous treatments is not available, but can be thought to vary substantially as these men are not from a homogenous geographic location. Once enrolled in the study, patients all received like-treatments consisting primarily of chemotherapy. Relevant covariates include age, changes in CEC, changes in CTC, and ECOG status, which is a measure of patient mobility and quality of life. CEC, CTC and PSA are measured at as many follow-ups as possible, but the completeness of these data varies from patient to patient. The outcome of interest is death from any cause, however most deaths are cancer-related.

For these analyses I created a set of laboratory covariates that consist of the mean change over all observed time-points. For instance, a subject may have three PSA measurements including baseline, so their aggregate PSA measure would be 

\[
\frac{(t_1 - t_0) + (t_2 - t_1) + (t_3 - t_2)}{3},
\]

where \(t\) is the PSA value at time 0 through 3. When the mean change over time is negative, it suggests that the patient is improving or that the treatment is working. Conversely when the mean change is positive, it suggests that the patient is getting worse or that the treatment is not working. This mean change measurement is meant to capture the average trajectory of the patient’s measures over time. This method may perform poorly when patients have trajectories which start and end at the same values. These subjects will have mean changes near zero, but may this still serve as an adequate measure of trajectory.

I generated three additional dummy variables indicating whether a subject’s baseline laboratory measures were above the corresponding median values. These variables were meant to account for large observed changes that were merely an artifact of having extreme starting values and not of being sensitive to the treatment. These can be thought of as prognosis effects, or baseline effects.

Preliminary data analysis included evaluating Kaplan-Meier survival curves for differential survival trends when stratifying by particular covariates.
Cox proportional hazards models were used for primary analysis. Cox models have several assumptions, the most important of which are proportional hazards and stable contributions of main effect terms over time. These assumptions were tested using plots of scaled Schoenfeld residuals against time. When proportional hazards hold, we expect the slope of the Schoenfeld residuals to be zero. Substantial deviations from zero indicate that the model assumptions are not met. Chi-squared tests of transformed survival time versus the residuals were also conducted to assess model assumptions.

All analyses were done in the statistical program R version 2.5-1 using Mac OS X version 10.4.11. The packages used include 'survival,' and 'gplots.'

**Results:**

Table 1 below displays the range and quartiles for key variables. Most subjects were around the age of 70, and very few had massive reductions in mobility and quality of life (ECOG = 2). It is interesting to note that, in aggregate, PSA decreases following treatment. Also of interest is the observation that CEC increases in aggregate following treatment, which is a counterintuitive result.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>1st Quantile</th>
<th>Median</th>
<th>3rd Quantile</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Years)</td>
<td>46</td>
<td>63</td>
<td>70</td>
<td>77</td>
<td>92</td>
</tr>
<tr>
<td>ECOG Status</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Baseline PSA</td>
<td>1.90</td>
<td>52.50</td>
<td>144</td>
<td>485</td>
<td>17800</td>
</tr>
<tr>
<td>Baseline CTC</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>30</td>
<td>5925</td>
</tr>
<tr>
<td>Baseline CEC</td>
<td>2</td>
<td>17.25</td>
<td>32</td>
<td>87.75</td>
<td>1102</td>
</tr>
<tr>
<td>2 - 5 Week PSA</td>
<td>0.30</td>
<td>39</td>
<td>146</td>
<td>421</td>
<td>17420</td>
</tr>
<tr>
<td>2 - 5 Week CTC</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10.50</td>
<td>767</td>
</tr>
<tr>
<td>2 - 5 Week CEC</td>
<td>3</td>
<td>17.25</td>
<td>32</td>
<td>87.75</td>
<td>1102</td>
</tr>
</tbody>
</table>

Figure 1 contains Kaplan-Meier survival curves stratified by quartiles of the baseline laboratory measures.
Figure 1. Kaplan-Meier plots of survivorship stratified by quartiles of baseline laboratory measures. The crosses on the curves denote censored observations. PSA and CTC quartiles separate logically, with subjects in the first quartile surviving longest, and subjects in the forth quartile having the poorest survival. There is not clear separation in the CEC quartiles.

The survival trends for PSA and CTC are logical, with subjects in the first quartile (lowest values) having the greatest survival, and those in higher quartiles having poorer survival. The trend is muddled in the CEC plot, with no clear separation. These plots suggest that CEC may not serve as a good predictor in the Cox models.

Since we are interested in evaluating the usefulness of CTC and CEC counts in predicting survival, I created a “gold standard” model using changes in PSA. Comparison to this model will determine the value of the two new potential biomarkers.

All Cox models were stratified by ECOG status and a dummy variable indicating whether the patient is younger or older than the median age at baseline. Stratifying in a Cox model is akin to controlling for a variable in traditional linear regression. In this case, I stratified by age because of serious deviations from
model assumptions when age was modeled as a covariate (See diagnostics in appendix figure 4). The coarse categorization of age was chosen to insure that there were a sufficient number of events in each bin.

Table 2. Cox model results, Stratified by ECOG Status and Binary Age

<table>
<thead>
<tr>
<th>Variables</th>
<th>Gold Standard Model</th>
<th>CTC Model</th>
<th>CEC Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exp(coef) p-value</td>
<td>95% CI</td>
<td>exp(coef) p-value</td>
</tr>
<tr>
<td>Mean Change in PSA</td>
<td>1.42 0.012 [1.10, 1.90]</td>
<td>---</td>
<td>2.68 9.3 x 10-04 [1.49, 4.84]</td>
</tr>
<tr>
<td>Mean Change in CTC</td>
<td>--- --- ---</td>
<td>---</td>
<td>3.44 8.0 x 10-07 [2.10, 5.62]</td>
</tr>
<tr>
<td>Mean Change in CEC</td>
<td>--- --- ---</td>
<td>---</td>
<td>--- --- ---</td>
</tr>
<tr>
<td>High PSA?</td>
<td>1.76 0.013 [1.12, 2.75]</td>
<td>---</td>
<td>3.44 8.0 x 10-07 [2.10, 5.62]</td>
</tr>
<tr>
<td>High CTC?</td>
<td>--- --- ---</td>
<td>---</td>
<td>3.44 8.0 x 10-07 [2.10, 5.62]</td>
</tr>
<tr>
<td>High CEC?</td>
<td>--- --- ---</td>
<td>---</td>
<td>3.44 8.0 x 10-07 [2.10, 5.62]</td>
</tr>
<tr>
<td>Mean Change * High PSA</td>
<td>0.71 0.019 [0.54, 0.95]</td>
<td>---</td>
<td>6.81 1.1 x 10-07 [0.57, 2.11]</td>
</tr>
<tr>
<td>Mean Change * High CTC</td>
<td>--- --- ---</td>
<td>---</td>
<td>3.44 8.0 x 10-07 [2.10, 5.62]</td>
</tr>
<tr>
<td>Mean Change * High CEC</td>
<td>--- --- ---</td>
<td>---</td>
<td>3.44 8.0 x 10-07 [2.10, 5.62]</td>
</tr>
<tr>
<td>R² / Total Possible R²</td>
<td>0.057 / 0.945</td>
<td>0.157 / 0.938</td>
<td>0.032 / 0.926</td>
</tr>
</tbody>
</table>

For PSA, the hazard ratio corresponds to an increase in 50 units.
For CTC, the hazard ratio corresponds to an increase in 15 units.
For CEC, the hazard ratio corresponds to an increase in 15 units.

Table 2 summarizes the results from the Cox models. Diagnostics for each model can be found in appendix figures 5 through 7. Survival and cumulative hazard plots can be found in appendix figures 8 and 9. The comparison model (PSA) showed a significant hazardous effect of the mean change in PSA, in the starting magnitude of PSA, and in the interaction of the two. Interestingly, the interaction term was protective, which actually makes good sense: subjects who started at an extreme level will be most sensitive to the treatment. That is, they will see the greatest reduction in PSA over time, and their hazards will begin to approach those of the patients with more moderate starting values (probably leveling off at some point). The CTC model yields very similar results to those of the PSA model, which provides good evidence that CTC may perform at least equally well as PSA in predicting survival. The protective effect of the interaction of high CTC values and the mean change in CTC was again observed, likely for the same reason. Additionally, the R² statistic was marginally increased for the CTC model, however neither of the models appears adequate.

Perhaps the most interesting result was the apparent protective effect of CEC counts even though the coefficients were not statistically significant. As mentioned in the introduction, CEC counts are very unspecific and may represent vascular morbidity unrelated to the cancer. This idea can actually explain all the effects in the CEC model. Chemotherapy is very taxing on the body, and works by destroying cancer cells via large, controlled doses of radiation. Inevitably, the therapy will damage and destroy healthy cells
surrounding cancerous cells. The destruction of healthy cells could increase the CEC count in the blood, and indicate sensitivity to the treatment, which explains the protective effect of the mean change in CEC. Recall that the dummy variable denotes the baseline level of CEC (whether above or below the median). Subjects starting treatment with elevated CEC counts were likely already very ill and in poor vascular health. These subjects would probably be adversely affected by the treatment rather than helped.

Conclusions:

In summary, I found that similarly specified Cox proportional hazards models using PSA and CTC separately yield nearly identical results suggesting that CTC can perform at least as well as PSA as a predictor of survival. Additional results using CEC counts have implications for treatment decision-making. Patients with an elevated baseline CEC may already be too ill to benefit from chemotherapy and may prefer less invasive, comfort-focused therapies with fewer side effects over aggressive, survival-focused therapies like chemo.

A major shortcoming of this study is the lack of information regarding previous treatments and therapies, and also geographic locations. Some of the subjects received treatments outside of the United States and differences in national healthcare infrastructures would influence the treatments received.

A second shortcoming is the missing follow-up data. Obviously, subjects who die will be missing data, but many subjects who did not die during the study also had missing data, which reduces the precision of the model estimates and may bias the results if they are not missing at random. During any collection of data from patients who are extremely ill, there is good reason to believe that missingness may be due to factors related to the outcome. Perhaps many of the subjects were simply too ill to make it to the clinic, or the treatment made them so ill that they dropped out. All of these factors should be considered in future analyses.

CTC appears to do at least as well as PSA in predicting patient survival in these data, however the absolute fit of these models still remains very poor. A logical next step would be to evaluate the functional form of these predictors, perhaps using smoothing splines.

References:


Appendix A:

**Figure 1.** Trajectories of changes in PSA over all the time points. The mean change variables were averaged over times such as these. A value of zero denoted no change in the measurements. A negative value indicated improvement and a positive value indicated worsening over time.
Figure 2. Changes in CTC over time
Figure 3. Survivorship stratified by ECOG status.
Figure 4. Sample model diagnostics when age is modeled as a covariate. These plots depict the Schoenfeld residuals versus time. The slope of the residuals should be zero, and centered around zero when the proportional hazards assumptions are met. Notice the departure from proportionality for age (upper left).
Figure 5. Schoenfeld residuals for the PSA model.
**Figure 6.** Schoenfeld residuals for CTC model

**Figure 7.** Schoenfeld residuals for CEC model.
Figure 8. Survival curves for each of the final models, stratified by age and ECOG status. The legend in the bottom right panel is used for all three plots.
**Figure 9.** Cumulative hazards for all three final models, stratified by age and ECOG status. The legend in the bottom right panel is used for all three plots.