Laboratory and Biometric Predictors of Cancer-Related Mortality in an Insured Population

James Palmier, MD, MPH; Brian J. Lanzrath
Laboratory and Biometric Predictors of Cancer-Related Mortality in an Insured Population

James Palmier, MD, MPH; Brian J. Lanzrath

Objectives.—Identification of statistically significant laboratory and biometric predictors of cancer-related mortality among insured individuals.

Background.—Numerous clinical studies have identified correlations between various laboratory results and physical measurements and cancer risk, often of a univariate nature. A study of life insurance claims has permitted a broad multivariate analysis of laboratory and biometric risk factors for cancer mortality in an insured population.

Methods.—Of the applicants with complete laboratory and physical measurement profiles, 1.25 million were available and followed for an average of 4.7 years. Dates of 518 life insurance claims resulting from cancer deaths were recorded, and the resulting data set analyzed by multivariate Cox Proportional Hazards regression to identify statistically significant predictors of cancer-related mortality among insured individuals.

Results.—Among demographic variables, cancer deaths were found to be strongly associated with age and tobacco use, but not with gender. Among serum and urine analytes, liver function tests (principally GGT and ALP), hypocholesterolemia, proteinuria, and low fructosamine were found to be independently predictive of cancer mortality. Among physical measurement variables, there was a positive relationship between cancer mortality risk and height and relatively weak relationships with pulse and blood pressure. Weight and body mass index (BMI) were not statistically significant covariates.

Conclusions.—The findings highlight the potential value of laboratory analytes and biometric measurements to cancer risk assessment including low to low-normal values in analytes (particularly cholesterol and fructosamine) whose diagnostic value in clinical practice and underwriting may be advantageous.

Cancer is the second-leading cause of death in the United States, narrowly trailing heart disease. The clinical and etiological heterogeneity of cancer, however, have naturally encouraged a comparable emphasis on specificity in the identification of risk factors, tending at times to de-emphasize the role of more global predictors of risk of the kind which are much more prominent in the cardiovascular literature (eg, serum...
Lipids and the metabolic syndrome. Past studies have suggested significant correlations between hypocholesterolemia, enzymatic liver function tests, proteinuria, serum fructosamine, serum protein, and height/weight, and various specific cancers on at least a univariate basis, although more comprehensive multivariate assessments are rare, likely due in part to the logistical difficulties of administering a uniform multi-analyte laboratory panel to large, healthy populations.

Direct and specific biochemical markers of cancer are rarely employed in the life insurance underwriting process, especially in general laboratory screening of applicants. Instead, cancer risk is most often assessed through clinical or family history, and more weakly by the presumed correlation between other underwriting standards (such as those developed for liver or cardiovascular disease) and cancer.

This study draws upon laboratory results and physical measurements collected from a large group of life insurance applicants, and subsequent deaths claims from this population, to identify individual laboratory analytes and physical measurements predictive of cancer mortality in insured individuals.

**METHODS**

Between October 2001 and June 2010, 1.25 million life insurance applicants tested by ExamOne, Inc. (Lenexa, KS) in various periods (varying by claim source) were matched to claims records provided by 3 individual carriers and to the Social Security Death Master File (SSDMF). All applicants were subject to additional underwriting (eg, medical and family history) prior to issue; deaths might fail to be reflected in claims through declination, non-completion, or lapse. The median age of applicants was 40; 54.9% were male, and 9.8% were tobacco users (as defined by a urine cotinine concentration of 0.5 µg/ml or higher). A total of 1901 death claims originating from tested individuals were identified, of which 518 listed some form of cancer as the cause of death. The mean follow-up period was 4.7 years. The complete panel of physical measurements and laboratory results available for all tested individuals (and evaluated for inclusion in the final regression model) is provided in Table 1.

With the exception of those variables evaluated in a strictly Boolean (P/N) sense (urine cotinine, urine leukocytes, urine hemoglobin, urine glucose, and a sex indicator dummy variable), each raw result was used to derive 4 daughter variables, defined in part by each applicant’s membership in 1 of 10 demographic classes (males and females of the age ranges 18–29, 30–39, 40–49, 50–59, and 60–79):

1. A low-outlier indicator (low), set to 1 when a result fell below the first demographic-specific percentile of the parent variable’s distribution, and 0 otherwise.

2. A high-outlier indicator (high) indicator, set to 1 when a result fell above either:
   a. The 95th percentile (for Gamma-glutamyl transferase (GGT), Alanine aminotransferase (ALT), Triglycerides, Urine Protein, and the Urine Protein/Urine Creatinine ratio) or
   b. The 99th percentile (all other variables).
   ...and 0 otherwise. Percentiles were defined relative to each applicant’s demographic group (above). high and low indicator variables were designed to capture the effects of outlier values without causing them to be reflected in the better-populated regions of the distribution. The 95th percentile was chosen over the 99th for variables where the latter was at least twice the former.

3. A truncated linear term (trunc) set to the parent value when both the low- and high- indicators were 0, and the outlier cutoff when either was equal to 1.

4. A second order term (quad), defined as the square of the difference between the
truncated linear term and the demographic-specific median of the parent. Higher order polynomial terms are intended to reflect non-linear (e.g., 'J-shaped', 'U-shaped') relationships between the underlying variable and risk. Third order terms were evaluated on a limited basis, and found to offer little marginal contribution.

Predictive relationships were identified by Cox proportional hazards regression in SAS 9.1.3 SP 4; right censorship times were defined as the earlier of the end of the period for which claims were provided (varying by source), or date of death as determined through claims or the SSDMF (in the case of non-cancer deaths). The final model was selected by the stepwise elimination of high p-value variables. At each step the least statistically significant variable in the working model was removed, and the coefficients of the remaining variables recomputed. This process both simplified the final model, and enabled a more precise quantification of the effects of relevant variables through the elimination of confounding factors.

RESULTS

Parameter estimates, standard errors, hazard ratios, and p-values for all final model variables are provided in Table 2. Although not reported in the table, the linear component of pulse approached the cutoff for statistical significance (p = .0561) with a point estimate for the hazard ratio of 1.01 (implying that each 1 beat/min increase in pulse increases cancer mortality risk by 1%). The global Wald Chi-Square statistic was 786.0834 (df = 21), implying an aggregate model p-value of less than $10^{-100}$. As the interpretation of coefficients for many model variables is non-intuitive, Figures 1–3 provide graphical representations of the marginal risk functions for GGT, urine protein and creatinine, as well as a cross-section of the lipid function at a serum triglyceride concentration of 148.5 (the median value for males 40–49). For consistency, all figures have been calculated for males 40–49, but the risk surfaces were virtually isomorphic across demographic groups, with only the axis values changing to reflect the differing medians and high and low cutoffs for each
analytical stratum. Median GGT values were lower in females 18–29, for instance, so Figure 1 would be compressed horizontally for that group.

Gender, which is among the largest and most robust predictors of all-cause mortality, is not a statistically significant cancer death risk covariate in this population (p = .6015). Cotinine (tobacco-use) positivity is observed to increase the likelihood of cancer-related mortality by 69% (CI: 34%–112%) relative to cotinine-negative individuals. Age increased cancer risk by 8.5% per year of age, a faster rate than is observed for all-cause SSDMF mortality in this population (7.5%, p for H0: βAge – Cancer-related = βAge – All Cause is .027), implying that the risk of cancer-related mortality increases with age in not only an absolute, but also a relative sense, after controlling for other factors. Increasing height also increased cancer risk by 5.4% (2.2%–8.7%) per inch (2.54 cm) of height, a tendency which has been reported elsewhere.6

The liver function tests (LFTs) alkaline phosphatase (ALP) and GGT were positively correlated with cancer mortality over most of their modeled ranges, albeit with a moderate quadratic component in the latter case (flattening the risk curve at higher GGT values). ALT levels were negatively correlated with mortality, with a 1.6% reduction in risk per unit increase. The lipid risk function surface (Figure 3) identifies two distinct regions of elevated mortality – a zone of moderate risk associated with elevated total cholesterol (TC) and below-median high-density lipoprotein (HDL) levels, and an area of more extreme elevations for cholesterol levels below the first percentile, when accompanied by relatively low HDL concentrations. The observed discontinuity of the urine protein/urine creatinine risk function at ratios above the 95th percentile is reflected in Figure 2, although the general tendency of cancer morality rates to increase with both the protein/creatinine ratio and total urine protein remain in effect on either side of this

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Hazard Ratio</th>
<th>p=</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>0.08139</td>
<td>0.00358</td>
<td>1.085</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UCOT</td>
<td>0.52250</td>
<td>0.11688</td>
<td>1.686</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>HEIGHT_trunc</td>
<td>0.05258</td>
<td>0.01563</td>
<td>1.054</td>
<td>0.0008</td>
</tr>
<tr>
<td>ALP_trunc</td>
<td>0.00743</td>
<td>0.00213</td>
<td>1.007</td>
<td>0.0005</td>
</tr>
<tr>
<td>GGT_trunc</td>
<td>0.01881</td>
<td>0.00548</td>
<td>1.019</td>
<td>0.0006</td>
</tr>
<tr>
<td>GGT_quad</td>
<td>−0.0002169</td>
<td>0.0001082</td>
<td>1.000</td>
<td>0.0451</td>
</tr>
<tr>
<td>ALT_trunc</td>
<td>−0.01589</td>
<td>0.00533</td>
<td>0.984</td>
<td>0.0029</td>
</tr>
<tr>
<td>CHOL_low</td>
<td>0.68528</td>
<td>0.32593</td>
<td>1.984</td>
<td>0.0355</td>
</tr>
<tr>
<td>TRIG_trunc</td>
<td>−0.00287</td>
<td>0.00105</td>
<td>0.997</td>
<td>0.0065</td>
</tr>
<tr>
<td>CHOL_HDL_trunc</td>
<td>0.38932</td>
<td>0.12864</td>
<td>1.476</td>
<td>0.0025</td>
</tr>
<tr>
<td>LDL_HDL_low</td>
<td>−2.41596</td>
<td>1.02258</td>
<td>0.089</td>
<td>0.0181</td>
</tr>
<tr>
<td>LDL_HDL_trunc</td>
<td>−0.46357</td>
<td>0.14257</td>
<td>0.629</td>
<td>0.0011</td>
</tr>
<tr>
<td>FRUC_trunc</td>
<td>−1.25481</td>
<td>0.24410</td>
<td>0.285</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UPROT_trunc</td>
<td>0.04589</td>
<td>0.01134</td>
<td>1.047</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>UCREAT_low</td>
<td>0.79003</td>
<td>0.36307</td>
<td>2.203</td>
<td>0.0296</td>
</tr>
<tr>
<td>UCREAT_trunc</td>
<td>−0.00351</td>
<td>0.00108</td>
<td>0.997</td>
<td>0.0011</td>
</tr>
<tr>
<td>UCREAT_quad</td>
<td>0.0000133</td>
<td>6.38671E-6</td>
<td>1.000</td>
<td>0.0368</td>
</tr>
<tr>
<td>UPROT_UCREAT_high</td>
<td>−0.62780</td>
<td>0.22413</td>
<td>0.534</td>
<td>0.0051</td>
</tr>
</tbody>
</table>

Parameter coefficients represent the natural logarithm of the increase in cancer mortality risk resulting from a one unit change in each variable; hazard ratios give the non-log effects. Thus, the change in risk from a 6-month increase in age is e^0.08139*6/12 = 1.042, or 4.2%.
Figure 1. GGT and cancer mortality risk in males 40–49 marginal multivariate contribution.

Figure 2. Urine protein and urine creatinine and cancer mortality risk in males 40–49 (marginal multivariate contribution).
boundary. The effect of decreasing serum fructosamine is highly significant; over the middle 98% of the population range, a 0.1 mmol/L decrease in this value is associated with a 13.4% (8.1%–19.0%) increase in the risk of cancer-related death.

**DISCUSSION**

The nature of the data available in this study – particularly the source of the dependent variable (time to cancer-related life insurance claims) – constrains the range of general cancer-mortality markers observable through our methodology. Not all applicants for whom data were available will have been issued policies (a necessary precondition for any later claims), and carrier underwriting almost certainly assured that those declined coverage were of disproportionately high risk. At 4.7 years, the mean follow-up time was relatively limited, although the earliest applicants from certain carriers were observed for over 8 years. The risk factors identified in this study will therefore tend to be those that rarely prompt a decline of coverage in the course of conventional underwriting, and yet which meaningfully affect cancer mortality risk well within a one decade period.

Finally, there is the possibility of some degree of omitted variable bias stemming from the lack of serum total protein or serum albumin variables (which were not available for one large carrier) in our model development process. Should this effect exist, it is likely the largest for serum fructosamine, which is the included variable most strongly correlated with the conventional protein panel, and therefore most likely to become a statistical proxy for it. Limited regressions on the subset of study applicants which included the complete protein panel exhibit only weak effects for the omitted variables, however, mitigating this risk. Other published studies incorporating serum total protein and albumin have noted similar relationships in some populations.²

The absence of any identifiable relationship between weight or body mass index (BMI) and cancer mortality is at variance with certain other findings,¹ but most of the effects observed in this study have been noted (often in univariate studies) in other published research. The identification of height, independent of weight, as a significant risk factor...
for a leading cause of death in the insured population suggests that BMI in isolation may not be an acceptably comprehensive metric of build in the underwriting process. The strong predictive value of low serum fructosamine in this study implies that this analyte may be of value in roles beyond its traditional function as a marker of diabetes risk.

CONCLUSIONS

Height, liver function tests (GGT, ALP and ALT), lipid panel, serum fructosamine, urine protein and urine creatinine are statistically significant independent predictors of cancer-related death claims in the insured population, after controlling for age, gender, and tobacco use status. This study’s findings may enhance future research and analysis of many conventional underwriting standards which tend to underweight the mortality risk of nominally high-normal GGT and ALP levels, and of low- to low-normal total cholesterol and serum fructosamine values. Additionally, the potential value of low serum fructosamine as a cancer risk marker may enhance that analyte’s current role as an indicator of diabetes (when elevated).

REFERENCES