JOURNAL OF INSURANCE MEDICINE

J Insur Med 2012;43:169–177

MORTALITY

Scoring Life Insurance Applicants' Laboratory Results, Blood Pressure and Build to Predict All-Cause Mortality Risk

Michael Fulks, MD; Robert L. Stout, PhD; Vera F. Dolan, MSPH

Objective.—Evaluate the degree of medium to longer term mortality prediction possible from a scoring system covering all laboratory testing used for life insurance applicants, as well as blood pressure and build measurements.

Method.—Using the results of testing for life insurance applicants who reported a Social Security number in conjunction with the Social Security Death Master File, the mortality associated with each test result was defined by age and sex. The individual mortality scores for each test were combined for each individual and a composite mortality risk score was developed. This score was then tested against the insurance applicant dataset to evaluate its ability to discriminate risk across age and sex.

Results.—The composite risk score was highly predictive of all-cause mortality risk in a linear manner from the best to worst quintile of scores in a nearly identical fashion for each sex and decade of age.

Conclusion.—Laboratory studies, blood pressure and build from life insurance applicants can be used to create scoring that predicts all-cause mortality across age and sex. Such an approach may hold promise for preventative health screening as well.

Address for Correspondence: Vera F. Dolan, MSPH, 601 North State Street, Suite 2000, Ukiah, California 95482; ph: 707-463-3200; Fax: 707-463-3209; e-mail: dolanvp@consultancy.com.

Correspondent: Vera F. Dolan, MSPH, Senior Research Scientist, Clinical Reference Laboratory, Ukiah, CA

Key words: Blood pressure, build, laboratory tests, life insurance, mortality, scoring.

Received: September 8, 2011

Accepted: January 24, 2012

INTRODUCTION

Health screening for well adults is still largely based on a disease model. Clinicians, researchers and test vendors have looked for test results that are associated with common diseases or the future risk of disease, and created screening profiles including those tests. The more specific a test is for a disease, the greater has been the acceptance. For example, more attention is often paid to ALT (SGPT) results rather than AST (SGOT) results since ALT elevations are more specific to liver diseases while AST may be elevated by a broader range of conditions with (up to recently) uncertain risk.[1]

Laboratories and physicians also most commonly call values within 2 standard deviations of the mean value (≈95% of values) as ''normal,'' based on a reference group that often includes all tested adults or patients. However, the real statistical reference range for many tests often varies dramatically by age and sex. For example, GGT (a liver enzyme) is typically listed as having a reference range up to 65 IU/L. This

actually includes 98.8% of 29- to 30-year-old female but only 91.5% of 64- to 65-year-old male insurance applicants whose samples were recently tested by Clinical Reference Laboratory (CRL).

Morbid events and disease-specific mortality are usually the measured endpoints for prevention or treatment programs, although the importance of all-cause mortality is also increasingly recognized and reported. In contrast to what has been reported in clinical studies, determining all-cause mortality risk is clearly a major financial and research focus of the life insurance industry. The underwriters who evaluate such risks for each applicant and medical directors who support that effort are, however, still strongly tied to a disease-centric screening model.

Blood and urine samples are obtained from applicants and usually submitted to 1 of 3 industry-wide laboratories (including CRL) for serum chemistries, hepatitis, HIV and diabetes screening, urinalysis, and limited drug testing. Limited examinations are also typically conducted including blood pressure (BP), height and weight. Test results are often translated into risk assessments using insurer-specific proprietary tables based on either the degree of variance from universal statistical normal ranges or other widely recognized cut-off values such as total cholesterol >240 mg/dL. Standards for BP and height-weight risk are taken from a variety of clinical and epidemiological studies on general populations, which are supplemented by additional limited industry data.

With the development of various ''preferred'' risk categories (with lower premiums) for life insurance carved out of the very broad ''standard'' risk class which encompasses most applicants, risk assessment shifted from disease-finding to identifying applicants with optimal health. To an extent, the same shift in focus can be seen in preventative medicine. It requires identifying individual and combinations of factors associated with reduced or improved longevity, rather than screening for particular diseases.

The mortality risk factors of age, sex and smoking status are excluded from underwriting risk assessment since these are already priced into the underlying base premium rates. This leaves lipids, BP, height-weight, family history and medication use (less commonly included) as the main risk discriminators in use by life insurance underwriters today for applicants otherwise qualifying for preferred rating classes. Personal medical history, motor vehicle history, and other laboratory abnormalities thought to be associated with increased mortality risk are more likely to prevent entry into any preferred category rather than discriminate risk between these categories.

We can find no published research to guide the use of multiple laboratory studies (or those tests plus BP and build) in predicting medium-term to long-term all-cause mortality risk in a general population stratified by age, sex and smoking status. Nor is there any published work to act as a guide in an even lower-risk population such as life insurance applicants (or comparable individuals typically seen for health maintenance examinations).

Considerable research is available that looks at cardiovascular (CV) events and, less commonly, CV or all-cause mortality in conjunction with a limited selection of tests and ages.[2,3] Limited research is also available for all-cause mortality at older ages using a limited numbers of tests in a small sample.[4] Horne, et al evaluated the use of the complete blood count along with sodium, potassium, chloride, bicarbonate, BUN, creatinine, glucose, calcium and age in predicting all-cause mortality at 30 days, 1 year and 5 years in a healthcare population of hospitalized patients and outpatients.[5] That test panel plus age was also tested against an NHANES III sample representing a more general population, and still appeared predictive of all-cause mortality.

What we lack is data evaluating a broad range of tests, BP and build in combination

(independent of age, sex and smoking status) for predicting 5 to 10 year all-cause mortality in individuals presenting for general health or wellness evaluations. This requires a very large population at least as healthy as NHANES that is tested and followed for mortality, as well as a methodology to create a risk prediction tool effective for individuals as well as groups. Data on individual life insurance applicants, who are typically age 20 to 80 and middle socio-economic class or higher without obvious increased risks such as congestive heart failure or stage 4 cancer, can potentially provide this.

METHODS

Applicants for individual life insurance in the US above minimum policy limits must provide blood and urine samples in addition to having BP, height and weight measured. The body fluid samples are almost all processed at 1 of 3 industry laboratories (including CRL), and the results provided to and paid for by the insurer(s) requesting the testing. Beginning in the early 1990s and commonly done after 2002, not only were samples processed, but the laboratories also began to record BP, height and weight from the examination and forward this electronically to the requesting insurer.

Over the past 5 years, the authors have been involved in research on the mid- and long-term relative mortality associated with single tests or physiologic groupings of tests (such as liver function tests or kidney function tests) along with examination measurements. This is done on an age- and sex-specific basis using subsets of our database of 24 million insurance applicants with test results using the Social Security Death Master File (DMF) to determine vital status.[1,6–8]

In addition to our published research on some tests, almost all tests utilized by the life insurance industry have been evaluated by us as to their predictive ability for all-cause mortality; this information has been provided privately to laboratory clients. More recently, we also began to examine the impact of combinations of results and measurements to improve the prediction of age-, sex- and smoking-specific relative mortality. The database for this effort included 2,010,877 insurance applicants applying for insurance from 1993 to 2006, with follow-up for vital status by the DMF in May of 2010. There were 25,483 deaths and mean duration was 6.89 years. All included applicants had a BP, height and weight along with routine laboratory studies. Some applicants had additional testing as requested by individual insurers.

The methodology for establishing relative risk for each test (or physiologic grouping of tests) is described in detail in previous published studies.[1,6–8] Usually this is done by establishing the risk for the middle 50% of test values on an age- and sex-specific basis, and comparing mortality for these results to bands of values outside this range. This approach was chosen for most tests because statistical ranges and association with mortality vary substantially by age and sex, but mortality across the middle 50% of values, though age- and sex-specific, otherwise varies little. This allows determination of test results associated with higher mortality *and* for those associated with lower mortality across a much wider range of values than is possible by using 2 standard deviations (95%) as the reference group. For some tests and measurements, this methodology was adjusted so that the lowest values were considered as the reference band and assigned a relative risk of one (for example: urine protein/creatinine ratio). For some other tests and measurements, a broader band of values was assigned a risk of one and compared to narrower bands inside this range and to those outside it (for example: BP, where the reference band was all those with BP $<140/90$).

A tabular approach was used to assign relative risk for each band of test values for groups by age and sex. The excess or reduced relative risk was obtained directly

from the data, which often generated non-linear curves that varied by age and sex instead of creating an algorithm modeled on the data. With our large dataset, it was usually possible to establish the relative mortality risk associated with each test, with the exclusion of obvious confounding factors. For example, the relative risk associated with various levels of urine protein/creatinine (p/c) ratio was determined after excluding cases with an admitted diabetic history, laboratory test results indicative of diabetes, hematuria, and/or a reduced estimated glomerular filtration rate (eGFR).[8] A similar approach was used in evaluating the risk associated with each liver enzyme in isolation including AST, ALT, alkaline phosphatase and GGT.[1] Tests which independently showed discrimination of relative mortality according to test value or level were selected for inclusion in the overall scoring.

The excess or reduced risk (expressed as debits or credits) associated with each included test result relative to the age- and sex-specific reference band was summed. That sum was then adjusted upward or downward slightly by age and sex so that the same composite score identified the same percentage of insurance applicants by sex and by decade of age. This composite score (including the demographic adjustment) was then tested separately against the 1993 to 2006 applicant database to determine its effectiveness in risk prediction by age and sex.

The test components selected for inclusion in the composite score reflect the testing typically done on insurance applicants either primarily or on a reflexive basis if other tests are abnormal (for example: HCV testing if ALT is elevated). Although the basic test panel is similar across insurers, use of certain tests such as hemoglobin and NT-ProBNP is not. Reflex guidelines for testing with HbA1c (vs only the more common screening test of fructosamine, a glycated protein) and hepatitis screening also vary by client. Any test that was performed for insurance testing and (based on our earlier analyses) appeared to independently predict risk, and not to measure the same factor twice (for example: LDL and total cholesterol) was included in the composite score. Tests were not excluded if they were less predictive (typically less commonly abnormal) based on multivariate analysis of the entire population, yet were clearly predictive for those individuals with abnormal values or if they impacted even a limited range of applicants according to their age and/or sex.

Tests we included in the composite score were: systolic and diastolic BP, pulse, height and weight converted to body mass index (BMI), albumin, globulin, hemoglobin (rarely available), fructosamine or HbA1c (when the latter is available), eGFR (using the Rule equation based on creatinine), alkaline phosphatase, GGT, AST, bilirubin, cholesterol/HDL ratio, total cholesterol (scored for low values), HDL (scored for high values), hepatitis screening, NT-ProBNP (rarely available), PSA (typically on males age 50+), urine p/c ratio and albumin, urine RBC count, HIV, and drug testing (typically only cocaine). Any test noted is included in the score if available or given a neutral ''0'' when not available.

Age and sex are not included in the composite score, and neither is use of tobacco products, all of which are already evaluated and priced into underlying base premium rates (or serve to identify an appropriate reference population in a clinical setting). When tobacco users are compared to those not using tobacco, the relative tobacco-specific mortality risk associated with various scores and the distribution of laboratory values is nearly identical (data not shown). Our scoring and this scoring analysis combine smokers (roughly 10% of applicants) with non-smokers.

Analysis was performed with IBM SPSS version 19.

RESULTS

Figures 1–4 compare relative mortality risks for each sex and decade of age for each

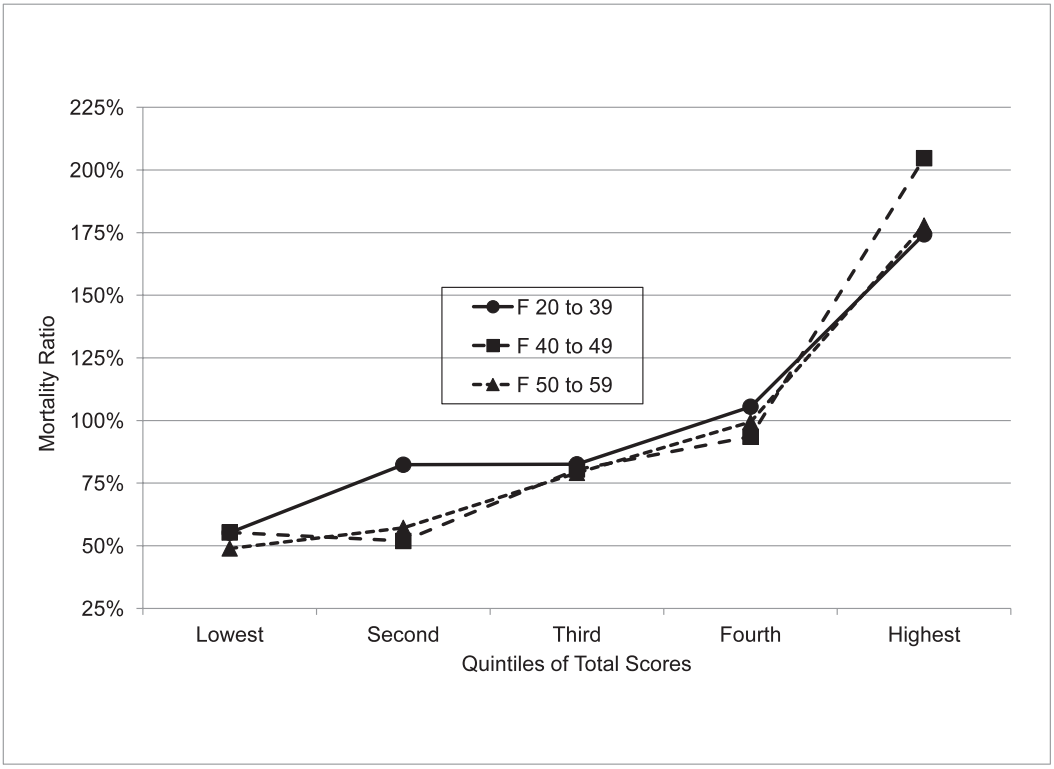

Figure 1. *Relative risk by quintile of composite score, females age 20 to 59.*

quintile of female (Figures 1 and 2) or male (Figures 3 and 4) applicants based on composite risk scores. These composite scores include the demographic adjustment for age and sex, so scoring generates nearly identical distributions across age and sex. Relative mortality curves are shown for each decade of age (age 20 to 39 combined) for each quintile of score relative to all scores for that sex and decade of age. These figures demonstrate a nearly identical linear trend in relative risk across age and sex for the best 4 quintiles. Risk in the quintile with the highest (least favorable) scores shows a steepening slope compared to the lower quintile scores.

For insurance applicants, this fifth quintile typically includes those with findings such as elevated HbA1c, hepatitis positivity, other markedly abnormal findings, or combinations of abnormalities that (if evaluated using age- and sex-specific statistical or risk-based normal ranges) would currently be identified as needing further underwriting evaluation. The best 4 quintiles are composed of applicants falling into a broad ''healthy pool,'' and who are typically eligible for the standard and preferred classes. Within these best 4 quintiles, however, the score (based on laboratory testing, BP and build only) is still capable of two-fold risk discrimination regardless of age or sex.

Linearity and concordance between decades of age is high, except for women age 20 to 39 in the second quintile. This deviation may be in part random error, but other analysis (not shown) suggests our scoring may function less perfectly for this group and needs further refinement, especially regarding the impact of pregnancy on test values. Pregnancy status is not available from our current data.

Because duration since testing varied considerably by applicant, and because even within each decade of age there could have been an association of younger lives to more favorable scores, the analysis in Figures 1–4 was replicated using a Cox regression split by sex and age groups 20 to 59 and 60 to 89, with score quintile and age as covariates. We

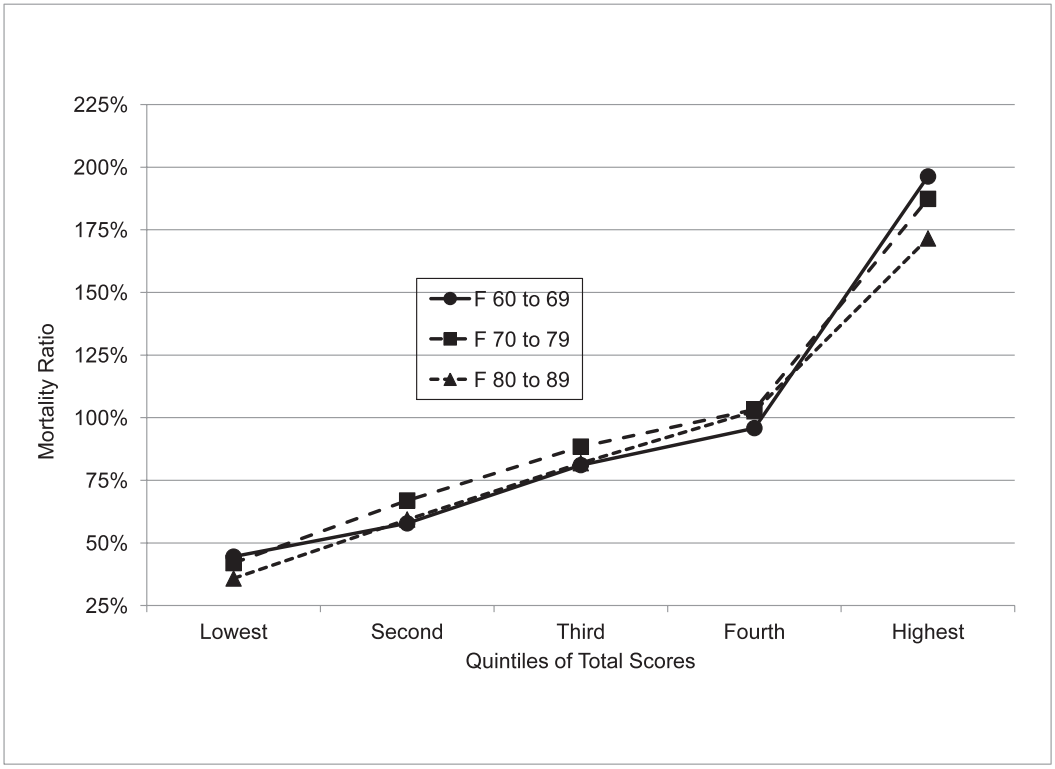

Figure 2. *Relative risk by quintile of composite score, females age 60 to 89.*

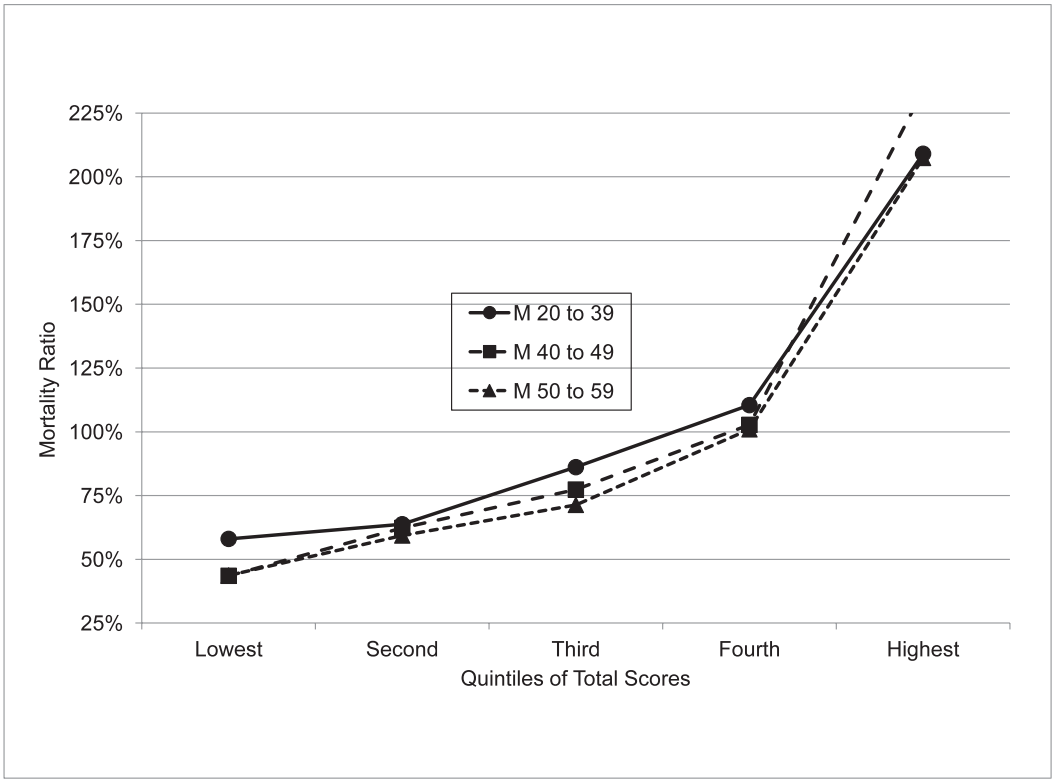

Figure 3. *Relative risk by quintile of composite score, males age 20 to 59.*

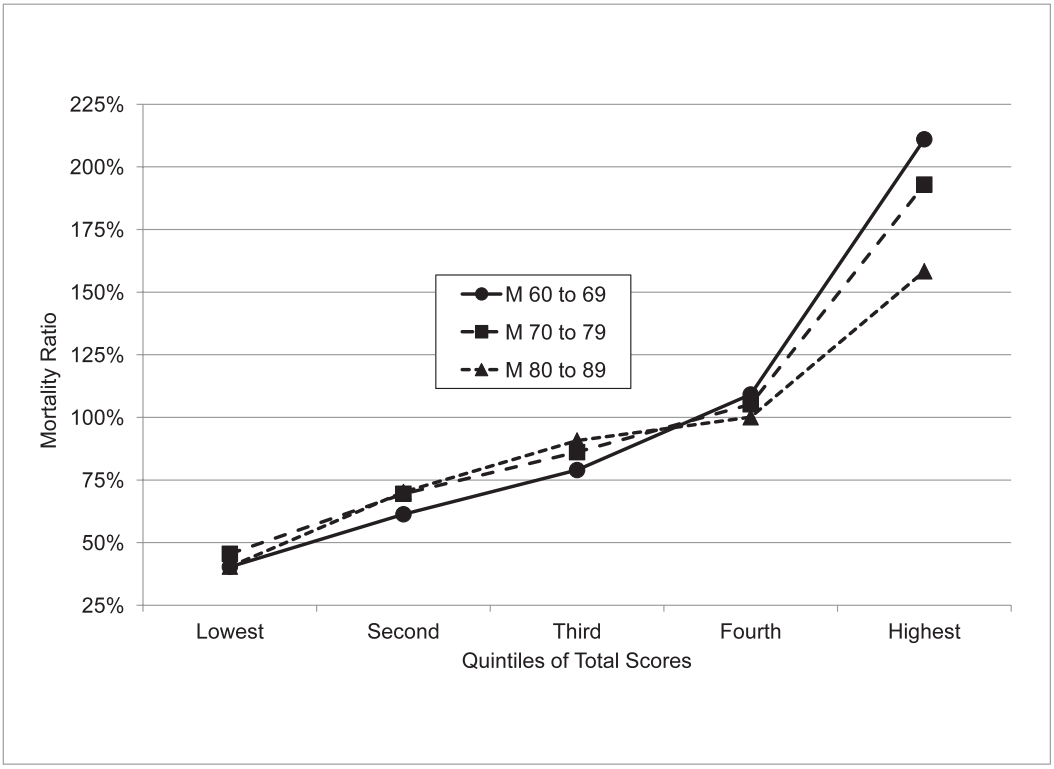

Figure 4. *Relative risk by quintile of composite score, males age 60 to 89.*

compared each of the 5 quintiles to mortality relative to all quintiles combined. The results appear in Tables 1–4. The mortality (hazard) ratios closely approximate those shown in Figures 1–4, and are significantly different from 1 except for the fourth quintile, which nearly matches the mortality of the entire group. Ninety-five percent confidence intervals show very limited overlap between quintiles. While this analysis allows for inclusion of age as a variable, and identifies both statistical significance and 95% confidence intervals, any lack of linearity or discordance by decade of age (apparent in the figures) may be hidden.

DISCUSSION

The methodology used to create a composite score is critical to making accurate all-cause mortality predictions using laboratory testing, BP and build across a wide range of sex, age and risk levels. Obviously, there is collinearity between many of the variables; sometimes two or more test results may measure almost the same thing. Correct

Table 1. Cox Regression Results for Age and Score Quintile, Females Age 20 to 59

			95% CI for MR	
	Sig.	MR	Lower	Upper
Age (years)	.000	1.088	1.086	1.091
Lowest Quintile	.000	.561	.513	.613
Second	.000	.649	.593	.711
Third	.000	.814	.754	.878
Fourth	.208	.954	.888	1.026
Highest Quintile	.000	1.755	1.667	1.847

Table 2. Cox Regression Results for Age and Score Quintile, Females Age 60 to 89

			95% CI for MR	
	Sig.	MR	Lower	Upper
Age (years)	.000	1.099	1.095	1.102
Lowest Quintile	.000	.476	.415	.546
Second	.000	.622	.566	.684
Third	.000	.813	.730	.905
Fourth	.211	.952	.882	1.028
Highest Quintile	.000	1.821	1.712	1.936

Table 3. Cox Regression Results for Age and Score Quintile, Males Age 20 to 59

			95% CI for MR	
	Sig.	MR	Lower	Upper
Age (years)	.000	1.072	1.071	1.074
Lowest Quintile	.000	.525	.494	.557
Second	.000	.638	.600	.678
Third	.000	.773	.736	.812
Fourth	.588	.987	.943	1.034
Highest Quintile	.000	1.966	1.899	2.034

Table 4. Cox Regression Results for Age and Score Quintile, Males Age 60 to 89

			95% CI for MR	
	Sig.	MR	Lower	Upper
Age (years)	.000	1.092	1.089	1.095
Lowest Quintile	.000	.463	.421	.510
Second	.000	.658	.617	.702
Third	.000	.809	.748	.875
Fourth	.422	1.023	.968	1.081
Highest Quintile	.000	1.900	1.814	1.990

attribution of risk by mathematical modeling alone is impossible. For example, only total cholesterol or LDL should be included in a multivariate model, not both. Another example is that the overlapping impact of two kidney function tests (such as urine protein/creatinine ratio and eGFR) must be considered prior to adding both into a multivariate analysis.

A second issue is that a multivariate analysis including or stratifying all ages and both sexes will establish the risk of components over the entire population. Many specific test results such as a GGT of 65 IU/L, the commonly used upper limit of normal, carry dramatically different relative risks by age and sex. Any model not based on analyses distinct by sex and multiple bands of age may not accurately identify that risk for the majority of individual subjects. Merely stratifying the data while performing a Cox analysis, which splits the groups during the analysis but still generates a single averaged hazard ratio shared across age and sex, will not accomplish this.

A third issue is the allocation of risk across test findings. We believe that more effective risk scoring will result if diabetics are excluded from any multivariate analysis when determining the risk of proteinuria, and if HCV positives are excluded when scoring liver enzymes. This is because the goal is determining the risk associated with those serum and urine findings in the absence of an obvious cause.

Our approach to creating component test scores and combined risk score may not prove to be the optimal one, but the limitations noted above were considered both when assigning risk to each test value and creating a composite risk score. Whether evaluating insurance applicants or individuals for health maintenance review, an accurate individual risk assessment is the goal. Our approach also provides the relative risk level on an age-, sex- and smoking-specific basis for each test, for BP and for build as well as the composite score. This allows identification of particular areas of risk on which the insurer or health provider can focus.

Our study has limitations. Because of the roughly 2-day delay between sample collection and processing associated with insurance testing, hematology studies beyond a hemoglobin test are not routinely attempted. Other authors have found them useful in risk prediction.[4] However, the examination, serum and other whole blood analysis (HbA1c) for insurance applicants is extensive and includes tests on a reflexive basis such as HCV, HBV and CDT (for heavy chronic alcohol use), which are not usually performed as screening tests in clinical medicine, as well as including BP and BMI. Another limitation is the lack of cause-specific mortality data to better connect certain tests with particular risks, which would aid any preventative effort. Hopefully, future epidemiologic studies can provide this.

Both the lack of evidence of reduction in mortality or morbidity by use of laboratory panels and the relatively high cost of such testing have led experts to not encourage its use for preventative screening.[9,10] However, our age- and sex-specific risk-based approach suggests a possibly larger role for such testing in preventative medicine, as well as more effective use for insurance risk assessment. Testing costs are still a major clinical limitation. However, in the insurance setting, the entire panel of testing including a representative portion of reflex testing typically costs the insurer less than $25, including overnight transport of the sample to the laboratory. If this could be replicated in the clinical setting, such testing may be practical and could better focus limited resources on those patients who might benefit from further evaluation and treatment. However, the extent to which including such testing in a preventative program would reduce risk rather than merely identifying it (as we have demonstrated), has not been evaluated in a primary care setting and is currently unknown.

CONCLUSION

In this population of life insurance applicants, our approach of using a single risk score created on an age- and sex-specific basis using a broad range of laboratory tests along with BP and build was highly predictive of all-cause mortality risk across age and sex for all quintiles of scores.

REFERENCES

1. Fulks M, Stout RL, Dolan VF. Using liver enzymes as screening tests to predict mortality risk. *J Insur Med*. 2008;40:191–203.
2. D'Agostino RB, Ramachandran SV, Pencina MJ, et al. General cardiovascular risk profile for use in primary care, the Framingham heart study. *Circulation*. 2008;117:743–753.
3. Aktas MK, Ozduran V, Pothier CE, et al. Global risk scores and exercise testing for predicting all-cause mortality in a preventive medicine program. *JAMA*. 2004;292:1462–1468.
4. Lan TY, Chie HC, Chang HY, et al. Clinical and laboratory predictors of all-cause mortality in older population. *Arch Gerontol Geriatr*. 2007;45:327–334.
5. Horne BD, May HT, Muhlestein JB, et al. Exceptional mortality prediction by risk scores from common laboratory tests. *Am J Med*. 2009; 122:550–558.
6. Fulks M, Stout RL, Dolan VF. Association of cholesterol, LDL, HDL, cholesterol/HDL and triglyceride with all-cause mortality in life insurance applicants. *J Insur Med*. 2009;41:244–253.
7. Fulks M, Stout RL, Dolan VF. Albumin and all-cause mortality risk in insurance applicants. *J Insur Med*. 2010;42:11–17.
8. Fulks M, Stout RL, Dolan VF. Urine protein/creatinine ratio as a mortality risk predictor in non-diabetics with normal renal function. *J Insur Med*. 2012;43:76–83.
9. USPSTF Guidelines. http://www.ahrq.gov/clinic/prevenix.htm.
10. Canadian Task Force on Preventative Health Care. Accessed at: http://www.canadiantaskforce.ca/index.html.