OTR JOURNAL OF THE ACADEMY OF LIFE UNDERWRITING

A COMPREHENSIVE MULTIVARIATE APPROACH TO THE STRATIFICATION OF APPLICANT-LEVEL ALL-CAUSE MORTALITY RISK

Brian Lanzrath
brian.j.lanzrath@questdiagnostics.com

Betsy Sears
betsy.r.sears@examone.com
Exam*One*
Lenexa, KS

Troy L. Hartman
troy.l.hartman@examone.com

Executive Summary *Information technology has made underwriting much more efficient and it has evolved to the point where patterns in seemingly disparate medical data now show a clearer portrait of mortality risk. This article describes the risk stratification potential of an applicant-level mortality model derived from an analysis of 144 laboratory and physical measurement variables, in which each variable is fully controlled for the remaining 143. Based on these variables, individualized risk scores can be developed for each insurance applicant, identifying high risk individuals who were previously in preferred insurance pools and candidates who may turn out to have lower-than-expected mortality risk. These tailored risk scores can set the foundation for an era of personalized life insurance policy pricing.*

Introduction

"Science is built of facts the way a house is built of bricks: but an accumulation of facts is no more science than a pile of bricks is a house."

– Henri Poincare, circa 1890

Consider the following case: A 52-year-old woman applies for $150,000 of term life insurance. She denies tobacco use or any significant medical history. Upon testing, she is negative for HIV, cocaine and cotinine. Her laboratory results and physical measurements are as follows:

BPSYST	100	SGLUC	83	UpH	5.65
BPDIAS	72	AST	10	UPROT	13
PULSE	72	ALT	4	UCREAT	237.9
ALB	3.6	SPROT	6.1	UGLUC	Neg
ALP	47	TRIG	141	UPROT/CREAT	.055
BILI	0.3	GLOB	2.5	ULEUK	Neg
BUN	13	HDL	65	UHEMO	Neg
CHOL	230	**TC/HDL**	3.54	**BMI**	23.2
SCREAT	0.6	LDL	137	FRUC	1.4
GGT	55	LDL/HDL	2.1		

What would be the appropriate premium rating for this individual? Select, preferred or super-preferred? Read on to learn this applicant's true mortality risk. The life insurance risk assessment literature has been inundated recently with analyses of the mortality implications of individual laboratory tests. At their most comprehensive, these studies assess the effects of restricted groups of closely related assays (e.g., total cholesterol and HDL or AST, ALT and GGT). Unfortunately, the usefulness of these studies in applicant-level risk stratification is severely limited by the pervasive cross-correlation of laboratory and physical measurement variables (Fig. 1-next page). This interrelationship invariably afflicts all univariate and limited multivariate analyses with two closely associated shortcomings: *cross-attribution of effects* and *non-additivity of results*. Because BMI, for instance, is strongly correlated with ALB, ALP, blood pressure, pulse, TRIG, the AST/ALT ratio, BILI, GGT and the CHOL/HDL ratio, any analysis of the relationship between BMI and mortality which does not explicitly control for the effects of these variables (and for others with a weaker, but statistically significant relationship to BMI) will implicitly attribute much of the risk of low albumin, hypertension, etc., to BMI *per se*. Conversely, univariate analysis of any of the cor-

related variables will confound the consequences of abnormal BMI with that of the attribute under study. This is *cross-attribution*, the crediting of the effects of correlated variables solely to the test or measurement under consideration. *Non-additivity* follows directly from this effect; it is impossible for an underwriter to apply the conclusions of separate univariate studies of correlated variables (and again, virtually all common laboratory tests and physical measurements are correlated to a significant degree) without partially "double-counting" the effects of each test.

Fig. 1: Coefficients of Correlation Between BMI and Various Physical and Laboratory Metrics: Male Non-smokers 40-49 (N=1048289)

HEIGHT	0.03203	GLUC	0.09905	CHOL	0.05146
WEIGHT	0.87916	TRIG	0.17599	GGT	0.10308
ALB	-0.13267	URNPH	-0.06604	PROT	0.04565
AST	0.08227	GLOB	0.1447	HDL	-0.25608
FRUC	-0.02544	ALB/PROT	-0.17391	TC/HDL	0.23805
ALP	0.10173	AST/ALT	-0.23246	UPROT	0.06772
BPSYST	0.29885	LDL	-0.01389	UCREAT	0.08646
BPDIAS	0.26057	LDL/HDL	0.12993	UPROT/ UCREAT	0.01272
PULSE	0.17424	BILI	-0.11503		
ALT	0.1867	BUN	-0.00577		

For H0: r=0, p<.0001 for all variables

Note on correlation coefficients: A correlation coefficient of zero indicates a complete lack of any relationship between two variables; a value of 1 or -1 indicates a perfect relationship. Absolute values of >0.1 can be considered 'strong', but all of the values presented above are statistically distinct from 0 at a p=.0001 level of significance.

The objective of underwriting is the quantification of the mortality risk of a given *applicant*, not the decontextualized analysis of a set of unrelated *laboratory tests*—particularly not when these analyses are inherently intractable to applicant-level integration, and result in flawed assessments of composite mortality risk. This article describes the risk stratification potential of an applicant-level mortality model derived from a comprehensive multivariate analysis of 144 laboratory and physical measurement variables, in which each variable is fully controlled for the effects of the remaining 143.

Data and Methodology

The multivariate model described here was developed from 5.95 million life insurance applicants for whom a complete standard laboratory and physical measurement profile (Fig. 2) was available. The earliest profiles date to late 2001, when substantial numbers of physical measurements were first captured in a database, operated by a provider of services for insurance companies. The current version includes individuals tested through the end of 2008. This sample size represents less than one-tenth of its 62 million 1992-2010 records. Dates of deaths were obtained from the Social Security Death Master File (SSDMF).

Fig. 2: Standard Laboratory and Physical Measurement Profile

Age	ALB	Diuretic	TRIG
Sex	ALP	FRUC	UCOT
Height	ALT	GGT	UCREAT
Weight	AST	GLUC	UGLUC
BPSYST	BILI	HDL	UHEMO
BPDIAS	BUN	LDL	UPROT
PULSE	CHOL	PROT	URNPH

In order to more precisely model mortality risk by allowing for the very different implications of many variables according to age and gender, the risk assessment system consists of 10 independent stratum models—one for each of five age ranges (18-29, 30-39, 40-49, 50-59 and 60-74) per gender; final rankings were further subdivided by cotinine status (<0.3, ≥0.3 µg/ml). The models were constructed by Cox proportional hazards multivariate regression, and incorporated all of the "raw" variables listed in Fig. 2, as well as a variety of calculated ratios (e.g., TC/HDL, ALB/PROT), and various synthetic variables designed to permit the modeling of non-linear relationships between results/measurements and mortality risk (i.e., the detection of "J-shaped," "U-shaped" and other non-linear relationships). No individual variable was included in a given stratum model unless its p-value within that context was less than 0.05; the global p-value of the weakest complete stratum model (that for females 18-29) was 8.6×10^{-37} (chi-square=224, df=19).

Model Outputs

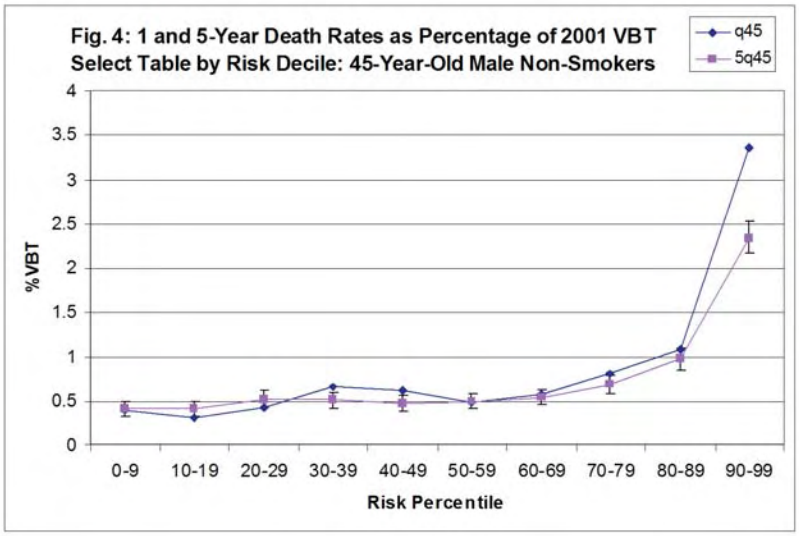
Hazard Score

An applicant's hazard score is simply the ratio of his mortality risk to the median risk in individuals of the applicant's age, sex and cotinine status, multiplied by 100. Thus, a 48-year-old male non-smoker with a hazard score of 153 is 1.53 times as likely to die during any given period as the median 48-year-old male non-smoker.

Percentile Ranking

The final output of this new method is a percentile ranking of the applicant's hazard score relative to the appropriate age/sex/cotinine status peer group. The possible range for a percentile is, of course, 0-99, where a score of 0 indicates that fewer than 1% of life insurance applicants present a lower mortality risk than the individual under consideration, while a 99

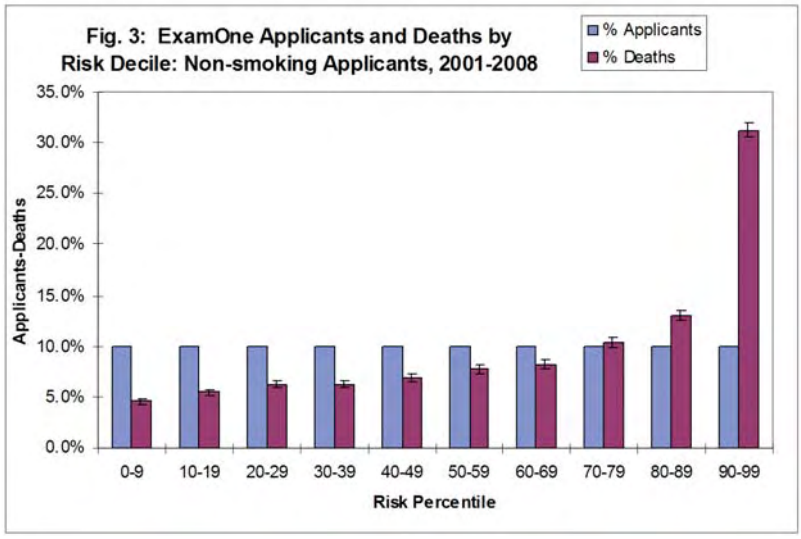
indicates that the applicant is within the most mortality-prone 1% of his demographic peers. The demographic normalization naturally precludes direct comparisons of scores among groups; a 74-year-old male smoker with a score of 20 is far more likely to die during a given period than an 18-year-old female non-smoker with a score of 80, although the older man is a substantially better risk relative to his peer group.



Mortality Stratification and Implications for Underwriting Requirements

Relative Death Rates by Risk Decile

Fig. 3 illustrates the distribution of applicants and deaths by risk decile in cotinine-negative applicants between 2001 and 2008. By design, each decile encompasses 10% of the applicant population, but the distribution of deaths is decidedly skewed. Individuals with scores of 9 or lower were 54.3% (54.0-54.6%) less likely to die during this period, while applicants scoring 90 and above were overrepresented in deaths by a factor of 3.12 (CI: 3.05 – 3.19).



Absolute Death Rates by Risk Decile

After assigning scores to all 5.95 million scorable applicants in the development database, 7-year mortality results were compared to the 2001 Valuation Basic Table (VBT) select death rates for each demographic group. Results for male non-smokers 40-49 are displayed in Fig. 4, but the overall shape of this graph differed very little among strata. Although the slope of this line between 0 and 74 was statistically significant, the overall impression is of comparatively flat, and distinctly low, absolute death rates in scores of roughly 74 and below. Indeed, the mean death rate of this group was less than 60% of the VBT, a level generally considered consistent with a preferred rating by conventional underwriting standards. By definition, of course, 75% of applicants lie within this ≤75 group.

As this stratification system successfully excludes high-risk applicants from its lower ranges, it must disproportionately concentrate them among higher scores. This is in fact the case. Above 75, the mortality risk increases in an approximately exponential fashion, such that the average individual placed in the highest (90-99) decile died at more than 250% of the 2001 VBT Select rate, while the most mortality-prone percentile (the 99th) experienced nearly a 10-fold multiple of the VBT.

Underwriting Requirements and Risk Percentile

Given the extremely favorable mortality rates among scores of 74 and below, it is reasonable to question the need for additional requirements in this group–particularly when the application offers no indication of unusual medical conditions. Fig. 5 (next page) illustrates that, although a majority of all reflex tests (the results of which are not reflected in current algorithms) are ordered for applicants with low scores, in many cases half or more of all positive results originate among the ~25% of specimens with scores of 75 or more. Microalbumin and CDT are particularly extreme cases of this trend; in 2009-2010, 61.6% of positive CDT and nearly three-fourths of positive microalbumin results were attributable to the highest risk quartile. In PSA, by contrast, positive results were distributed almost uniformly across the ranking range, suggesting that this test may add significant predictive value.

Attending physician statements (APSs) are among the

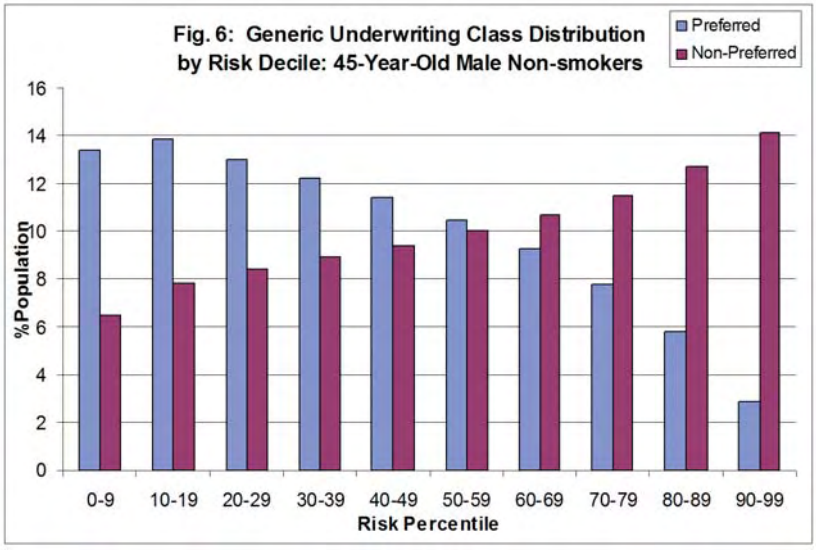
Fig. 5: Reflex Orders and Positives in Scorable Applicants (2009-2010)

Test	Percentile Range	<75	75+
CDT	%Reflexes	56.9%	43.1%
	%Positives	38.4%	61.6%
HAA	%Reflexes	62.5%	37.5%
	%Positives	51.9%	48.1%
HEPC	%Reflexes	75.0%	25.0%
	%Positives	53.5%	46.5%
A1C	%Reflexes	75.3%	24.7%
	%Positives	52.4%	47.6%
PSA	%Reflexes	82.2%	17.8%
	%Positives	80.0%	20.0%
UMALB	%Reflexes	55.4%	44.6%
	%Positives	25.2%	74.8%

most costly and – perhaps more importantly – time consuming requirements likely to arise in the course of the life insurance application process. During 2009 and the first half of 2010, 80.3% of APS orders among scorable applicants were associated with scores of 74 and lower–individuals who could, based on laboratory results alone, be expected to experience preferred levels of mortality. It seems reasonable to conclude that a significant fraction of these requests could have been eliminated had the underwriter possessed such compelling evidence of low underlying risk.

Comparison with Conventional Underwriting Criteria

While underwriting manuals naturally differ somewhat among carriers, prevailing standards are sufficiently similar that it is possible to speak coherently of a set of "generic preferred criteria." As applied to laboratory results in the course of this study, 35-40% of applicants met this fairly comprehensive set of guidelines (among many others, TC≤230 or TC≤250 and TC/HDL≤5, systolic BP<140, 20<BMI<28). While there is clearly a strong relationship between conventional underwriting status and the risk stratification approach, the agreement is far from perfect (Fig. 6). As assessed by this model, more than 30% of applicants currently excluded from preferred pools actually present lower mortality risks than the majority of preferred-qualified individuals. Conversely, more than 25% of preferred applicants actually exhibit higher mortality risk than most non-preferred individuals. Clear, identifiable deficiencies exist in traditional underwriting practices, resulting in the assignment of low-risk individuals to high-risk pools, and vice-versa.



Some of this divergence is attributable to differing assessments of individual risk factors–compared to conventional underwriting standards, this risk stratification method tends, for example, to be less tolerant of elevated liver function tests, and more permissive with respect to BMI–but the large majority stems from the ability to integrate the mortality implications of a profile as a whole. An applicant with multiple test results at the outer margins of the "normal" ranges may in fact represent an extremely elevated risk, while applicants with a single "out-of-range" result, but an otherwise ideal profile, may well rank among the least risky of their peers.

Cryptic Risk

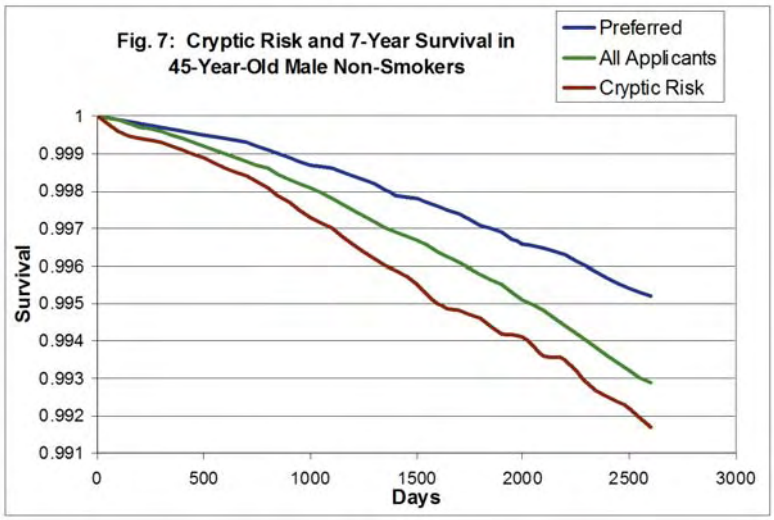
"A very small cause which escapes our notice determines a considerable effect that we cannot fail to see, and then we say that the effect is due to chance."
– Henri Poincare

A "cryptic risk" applicant is defined as an individual who meets all conventional preferred criteria, yet whose profile generates a score of 75 or above. As discussed above, this most frequently results from the interaction of several "borderline normal" (in conventional terms) laboratory results, no one of which would necessarily appear worthy of attention in isolation. Among all applicants analyzed, 5.9% can be classified as cryptic risk. The lowest incidence is among females 60-74 at 3.2%; the highest is in males 18-29 at 8.9%.

The typical cryptic risk applicant experiences a mortality risk of 142% VBT, more than twice the 63% observed of the preferred group as a whole in the data (Fig. 7, next page).

Hidden Healthy

Under the assumption that the life insurance industry's aggregate mortality expectations have been generally correct, the identification of a substantially underpriced sub-population (cryptic



risk applicants) implies the existence of a cross-subsidizing overpriced group–which the model has in fact identified. "Hidden healthy" applicants are those who would be excluded from the generic preferred class on the basis of one or more criteria, but whose scores are less than 75–implying preferred levels of actual mortality risk. This group is substantially larger than the cryptic risk population (encompassing 30-40% of applicants, again varying somewhat by age and sex), although the discrepancy between actual and expected mortality is less extreme, on average, than in cryptic risk cases.

BMI is by a comfortable margin the most common grounds for the denial of preferred classification to the Hidden healthy, followed more distantly by lipids (TC, TC/HDL and LDL) and blood pressure (Fig. 8). The marketing implications of extending preferred rates to so large a population are substantial.

Fig. 8: Cause of Exclusion from Generic Preferred Pool in Males 40-49 with Risk Percentiles Rankings<75 ("Hidden Healthy")

BMI	72.9%
Lipids	33.4%
Blood Pressure	8.2%
Urine Protein	5.3%
Urine Hemoglobin	3.3%
GGT	2.4%

Note: Because applicants can be excluded on multiple criteria, proportions do not sum to 100%

Studies in Fully Underwritten Applications

Multiple carriers are currently participating in studies of this system in fully underwritten policies. The conclusions of one such study (scored applicants =139,486, deaths=716, claims=307) are outlined here. As shown in Fig. 9, only a minority of all known deaths were reflected in claims (the remainder having presumably been declined, not taken out or lapsed prior to death). The disproportionately low death/claims ratio in high-ranked applicants strongly suggests an enhanced propensity to decline these individuals, even under conventional underwriting criteria–as would be expected given the known correlation between the model and current underwriting. However, large numbers of high-scoring applications were fully underwritten, issued and then fairly promptly paid as claims. Among applicants in the 99th percentile alone–a group which, it should be recalled, will reliably exhibit a 10-fold elevation of mortality rates–were 10 paid claims totaling $2.5 million.

Fig. 9: Deaths and Claims in Fully Underwritten Policies by Risk Percentile

Risk Percentile	Deaths	Claims	%Deaths in Claims
0-9	40	18	45.0%
10-19	50	20	40.0%
20-29	48	25	52.1%
30-39	72	38	52.8%
40-49	61	28	45.9%
50-59	42	23	54.8%
60-69	63	30	47.6%
70-79	78	41	52.6%
80-89	87	29	33.3%
90-99	175	55	31.4%
99 Only	42	10	23.8%
Total	716	307	42.9%

Beyond Underwriting Classes

A review of Fig. 4 may offer intimations of a more fundamental transformation of current underwriting models than any discussed in detail here. Given the continuous nature of risk stratification model's results (which can be adapted to stratify applicants into arbitrarily fine grades of risk) and the established relationship between these results and existing mortality tables such as the VBT, there is no conceptual barrier against using this system to generate fully personalized mortality tables for individual applicants, thereby enabling the calculation of applicant-specific premiums to the very penny. While such a radical revision of established practices cannot be expected to gain currency overnight, it does represent the natural culmination of the maturing science of mortality modeling.

The Outcome of the Case:

What was your rating? The mortality risk percentile: 99

That means that this individual had roughly 10 times the mortality risk of the typical 52-year-old female applicant. In this case the applicant died approximately 5 months after the paramedical exam, resulting in a claim.

Summary

In underwriting the life insurance applicant, the underwriter and underwriting managers have a fiduciary responsibility to utilize the best available tools to assess risk. Through its applicant-level approach to risk stratification, the multivariate model overcomes the inherent limitations of univariate and noncomprehensive multivariate analyses (cross-attribution and non-additivity) and accurately assesses the aggregate all-cause mortality risk of individual laboratory and physical measurement profiles. It offers the potential to reduce underwriting requirements, and identify both high-risk individuals currently accepted into preferred pools (cryptic risk), and low-risk applicants currently denied preferred classification (hidden healthy) in a manner which could ultimately extend preferred-level rates to as many as 75% of applicants. In time, and particularly as teleunderwriting and reflex testing are fully incorporated into future models, it will create the possibility of fully individualized policy pricing.

About the Authors

Brian Lanzrath is a scientist in Exam*One*'s Research and Development department, where he has participated in the development of multiple risk assessment and drugs of abuse assays for serum, urine and oral fluid. Since 2009 his primary responsibilities have been in data analysis, including reflex criteria definition and mortality modeling.

Betsy Sears has more than 23 years of experience in the insurance laboratory industry. She works closely with Exam*One*'s laboratory operations and medical staff on customer issues, and supports insurance customers through a variety of presentations on laboratory, medical and risk assessment topics. Her responsibilities also include managing international sales initiatives and overseeing Insurance Client Solutions in the US and Canada.

Troy L. Hartman, Executive Vice President, Exam*One*, earned his bachelor's degree from Baker University in Baldwin, KS, and his master's degree from Central Missouri State University. He joined Exam*One* as Executive Vice President in October 1999, where revenues have grown from $6M annually to over $100M annually. From October 1993 through December 1997, he served as Vice President of Business Development for IMR (Insurance Medical Reporter), a paramedical company. IMR merged with Physical Measurements Inc. (PMI) in December 1997. IMR had annual revenues of $32 million at the time of the merger. In December 1997, he became the Vice President of Sales and Marketing for PSA (Paramedical Services of America), a paramedical company. At the time Mr. Hartman left PSA in October 1999 it had annual revenues of approximately $75 million.