THE SIMILARITIES BETWEEN LIFE TABLE ANALYSIS AND MULTIVARIATE COX MODELS

Doug Ingle, FALU, FLMI
Vice President, Underwriting Research
Hannover Life Reassurance Company of America
Waukesha, WI
doug.ingle@hlramerica.com

Executive Summary *This article investigates the relationship between clinical medical research, along with the statistical techniques applied to those articles, and compares those results to results achieved conducting a mortality study using life table techniques and actuarial mortality tables. It includes a description of the relationship between a hazard ratio and a mortality ratio and demonstrates why results should be viewed carefully to identify the confounding variables that have been adjusted for in the research produced. The conclusion shows the close correlation between results using Cox multivariate regression when adjusting for age, sex and smoking, compared to actuarial life table methodologies. This will be done using publicly available data from the NHANES III study.*

Introduction

Actuarial mortality tables are periodically produced by the Society of Actuaries reflecting contemporary industry average mortality. This article will use the 2001 tables referred to as the 2001 VBT (Valuation Basic Tables). There are a series of tables created. Each table contains mortality by age and by duration since policy was issued. There are separate tables for gender (male, female) and smoking status (non-smoker, smoker, composite). Thus six separate tables are created to account for these factors.

Underwriting research often applies results from clinical medical studies published in reputable medical journals such as the *New England Journal of Medicine*, *Journal of the American Medical Association* and others. These articles generally describe results using statistical methods such as univariate and multivariate regression equation output that identify the amount of risk that should be associated with a variable under investigation.

This paper will describe how to bridge the gap between these divergent perspectives. By using publicly available data from NHANES III, this paper will show how journal articles might analyze the risk of mortality associated with total cholesterol by applying single and multivariate views of the risk. Results will also be calculated using traditional analysis of the risk applied to an underlying expected mortality referent, the 2001 VBT. When comparing results between these two techniques, they are very similar.

Mortality Associated with Total Cholesterol

Many studies have analyzed the correlation between total cholesterol and all-cause mortality. By using publicly available data, this article can be reproduced by anyone interested in testing the results described here. This article uses the NHANES III database–a publicly available non-random sample of the United States population, where information was originally collected on individuals between 1988 and 1994. The adult portion of this cohort was followed to December 31, 2006, where month and year of death were recorded.

To be included in this research, both a total cholesterol reading is needed as well as follow-up for risk of death. Each NHANES participant is assigned a unique identifier called a sequence number. This number is used to merge databases of information held on the same individual. One of the datasets is referred to as the NHANES III Laboratory Data File, which includes results for total cholesterol. That was merged with mortality outcomes found in the NHANES Linked Mortality Public-Use File. To be eligible for this study, the individual must have a cholesterol value and been

a member of the mortality follow-up dataset. This serves to create a subset of individuals available for consideration.

The serum total cholesterol measured in mg/dl was collected for 23,561 individuals. There were 20,024 individuals included in the mortality follow-up dataset. There were 17,094 individuals with both total cholesterol measurements and membership in the mortality follow-up registry. There were 4,257 deaths recorded in this subset population.

Cholesterol, Mortality and Hazard Ratios

A fundamental question to be answered is: Is there a relationship between mortality and cholesterol reading? This question will be used to help explain what a hazard ratio is. Hazard ratios are often presented in clinical literature. A hazard can be thought of as a risk. A hazard can be a hazard of anything the researcher wants to study. In this instance the outcome of interest is the risk of death or hazard of death. Because it is a ratio (hazard ratio), it is a comparative statistic, in this case, comparing risk of death to a referent population's risk of death. For insurance purposes we often compare mortality to standard (non-rated) risks and refer to this as a mortality ratio or a relative risk. For clinical articles, often a subset of all the members of the study acts as the referent population for which hazard of other groups can be compared.

In the application of life table methodology as taught within the insurance industry, a mortality rate is constructed by determining the number of individuals in a risk class who die, divided by the total number of person-years represented by that group as a whole. Person-years may also be referred to as exposure years (exp_yrs). For example, a person exposed to the risk of death for 3 years would provide 3 person-years of exposure.

Cholesterol will be categorically defined (using ranges of cholesterol values to define a group) to allow for the possibility that mortality isn't a simple linearly increasing pattern. An example of a linear pattern would be an increase in risk of death for each higher range of cholesterol. If this simple pattern suffices, nothing has been lost in conducting the analysis in this manner. If risk doesn't constantly increase, that finding will show up in the results. The shape of the curve will come in to play later in the article.

Table 1

Chol Group	Number	exp_yrs Sum	deaths Sum
59-140	987	13,385	156
141-160	1,836	25,515	281
161-180	2,675	36,256	518
181-200	3,171	42,873	686
201-220	2,884	38,366	750
221-250	3,158	40,813	979
251-275	1,288	16,360	452
276-300	659	8,309	254
301-325	248	3,038	95
326- UP	188	2,286	86

Individuals are placed into cholesterol groups (Chol Group) based on their measured cholesterol in mg/dl as defined in the table above (Table 1). The total number of exposure years are summed as well as the number of deaths recorded for each cholesterol group.

These data are used to produce mortality rates by dividing the deaths by the sum of the exposure years (exp_yrs), then multiplying by 1,000 to produce deaths per one thousand person-years of exposure (d/1000). See Table 2 below.

Table 2

Chol Group	Number	exp_yrs Sum	deaths Sum	Mortality Rate (d/1000)
59-140	987	13,385	156	11.65
141-160	1,836	25,515	281	11.01
161-180	2,675	36,256	518	14.29
181-200	3,171	42,873	686	16.00
201-220	2,884	38,366	750	19.55
221-250	3,158	40,813	979	23.99
251-275	1,288	16,360	452	27.63
276-300	659	8,309	254	30.57
301-325	248	3,038	95	31.27
326- UP	188	2,286	86	37.62

One of the cohorts is used to provide a referent population. Total cholesterol in the 181 to 200 mg/dl range is used as the referent. This category's death rate is 16 deaths per 1,000 person-years. The mortality rate from the other categories is compared to this. For example, the mortality rate for cholesterol

readings of 59 to 140 mg/dl is 11.65 deaths per thousand person-years. That is compared to (divided by) the referent group's mortality rate of 16 deaths per thousand person-years to produce a relative risk or mortality ratio of 11.65/16.0=73%, which can be found in the Mortality Ratio column of Table 3. This method is carried through to all the other cells. Note that, by definition, the mortality for cholesterols of 181-200 is 100%.

Table 3

Chol Group	Number	exp_yrs Sum	Deaths Sum	Mortality Rate (d/1000)	Mortality Ratio
59-140	987	13,385	156	11.65	73%
141-160	1,836	25,515	281	11.01	69%
161-180	2,675	36,256	518	14.29	89%
181-200	3,171	42,873	686	16.00	100%
201-220	2,884	38,366	750	19.55	122%
221-250	3,158	40,813	979	23.99	150%
251-275	1,288	16,360	452	27.63	173%
276-300	659	8,309	254	30.57	191%
301-325	248	3,038	95	31.27	195%
326- UP	188	2,286	86	37.62	235%

Results for mortality ratios by cholesterol group are plotted and displayed in Graph 1. Other than the first group, this shows a consistently increasing pattern for mortality, implying that as cholesterol increases, risk of death increases in an almost linear fashion. This will be challenged later in the article.

How do these mortality ratios or relative risk ratios compare to results that would be achieved should findings be reported using the Cox Proportional Hazard Model? The Cox Model is a popular regression model applied in clinical studies because it models outcomes particularly well when looking at time to failure, which is what life insurance is at a fundamental level. Table 4 below displays Cox Regression results using the same categories and same referent to see what hazards are associated with risk relative to the same referent. All statistical calculations were done using SAS version 9.3 (copyright 2002-2010, SAS Institute Inc., Cary NC).

Table 4 compares results achieved through calculated relative risks to results for hazard ratios using the Cox Model. The results that compare the previously described mortality ratios and Cox hazard ratios are expanded to 3 decimal places so the finer differences between the hazard ratio and mortality ratios can be identified. Therefore a univariate hazard ratio is very similar to a mortality ratio ex-

Graph 1

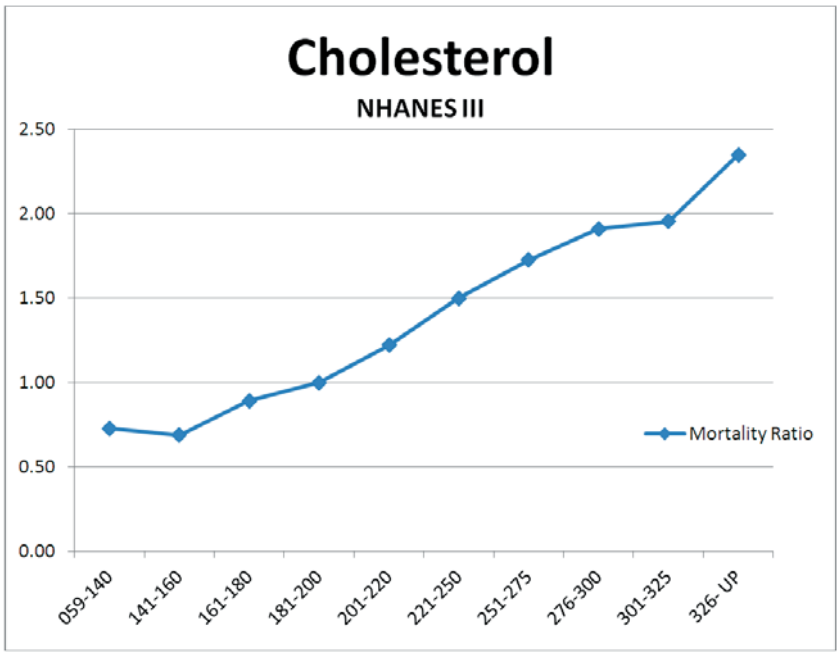


Table 4

Cholesterol Groupings	Mortality Ratio	Hazard Ratio
059-140	0.728	0.728
141-160	0.688	0.687
161-180	0.893	0.893
181-200	1.000	1.000
201-220	1.222	1.223
221-250	1.499	1.503
251-275	1.727	1.733
276-300	1.910	1.917
301-325	1.954	1.965
326- UP	2.351	2.367

pressed as deaths/exposure years divided by the referent group deaths/exposure years.

Age

Isn't it possible that, on average, cholesterol increases with age such that the rise in mortality is only caused by the artifact that cholesterol is just a proxy for age (meaning that, as age increases so does average cholesterol)? If so, then all that is truly being displayed is a pattern of increasing mortality due to age. Table 5 examines the relationship between a given age range and average cholesterol associated with the individuals in that age range.

Table 5

Age Group	Cholesterol Average
00-19	168.4
20-29	182.6
30-39	194.6
40-49	207.5
50-59	221.7
60-69	225.6
70-UP	218.3

It is apparent that as age increases, so does average cholesterol up to age 70. The univariate study previously presented can't determine whether or not the hypothesis that the difference in mortality is primarily due to age and not cholesterol range is true or not. What is needed is a multivariate view of risk. This concept is at the heart of the Framingham study as well. At first the Framingham researchers thought there might be one plausible predictor for risk of heart disease. They quickly became aware that risk for heart disease was multifactorial.

This calls for more complex equations, referred to as multiple variable equations, used to answer the question both for Framingham as well as here. By introducing other confounding variables into the equation, the independent risk associated with each variable can be determined. This is akin to the concept learned in basic algebra when attempting to solve problems using two equations with two unknowns. The outcome, the mortality rate, is already known. The variables that contribute to that finding (age and cholesterol) are also known. The equation is used to solve for the relative importance of each factor in predicting outcome. By introducing multiple factors into the equation, the goal of teasing out the independent contribution to risk of death for cholesterol alone–after adjusting for the confounding influence of age–can be achieved.

Those having read any medical research or reviewed journal publications on underwriting topics will recognize this concept of adjusting for confounding variables reported in the literature. This is done mathematically by including those variables (referred to as covariates) into the equation, thus the concept and application of multiple variable regression.

Then why not adjust for all the factors that are included in an industry mortality table, such as those found and published by the Society of Actuaries (SOA)? These industry tables display mortality by age, gender and smoking status. It holds that this research should adjust for these same factors as well.

Two Ways to Investigate the Same Question

Now that it has been shown that a hazard ratio is comparable to a relative risk ratio or mortality ratio, the next step is to see what happens to the cholesterol question when adjustments are made for age, sex and smoking. By introducing those variables into a multivariate Cox Model, the resulting hazard ratios can be compared to results found when using an industry mortality table and traditional actuarial life table techniques are employed.

Using the same NHANES III data and merging age, sex and smoking to each member of the pool allows for an actuarial life table "actual to expected" calculation to be performed. Smoking is defined by a "yes" answer to the question from the Household Adult Data File that asks, "Do you smoke cigarettes now?" The appropriate age, sex and smoking status mortality rates contained in the 2001 VBT age last birthday (ALB) tables are used to define the expected mortality anticipated. These expected mortality rates, sometimes referred to as qx's, are appended to each individual in the dataset based on age, sex, smoking status and duration. Duration in the VBT table is incremented by year. For those who die, the qx incorporates the full year of exposure in the year of death. For survivors, exposure year stops at the end of the exposure period. The individual mortality rates from the SOA tables are summed (added together) for each individual based on the factors described above. These summed qx's produce an actuarially expected mortality rate that can be compared to the observed mortality rate.

This analysis compares a non-underwritten population mortality rate to an insurance industry table. Insurance industry tables are impacted by underwriting. Therefore the population mortality rates are considerably higher than mortality rates found in the insurance industry tables. The mortality rate for cholesterols of 181-200 mg/dl produces mortality that is 172% of the VBT. Findings are standardized

Table 6

Cholesterol Group	N	exp_yrs Sum	deaths Sum	2001 VBT Expected	deaths/2001 VBT A/E	Standard-dized A/E	Cox HR
059-140	987	13,385	156	58.4	267%	155%	155%
141-160	1,836	25,515	281	148	190%	110%	113%
161-180	2,675	36,256	518	248.6	208%	121%	120%
181-200	3,171	42,873	686	397.7	172%	100%	100%
201-220	2,884	38,366	750	451.3	166%	96%	95%
221-250	3,158	40,813	979	573.3	171%	99%	96%
251-275	1,288	16,360	452	256.3	176%	102%	99%
276-300	659	8,309	254	142.4	178%	103%	101%
301-325	248	3,038	95	52.1	182%	106%	103%
326- UP	188	2,286	86	37.5	229%	133%	130%

by using this percentage of the VBT to represent expected or referent mortality. Stated another way, all the mortality ratios are divided by 172%.

Results in Table 6 compare the hazard rates from the Cox Model to the mortality rates from the standardized 2001 VBT. For some additional explanation for the headings:

- 2001 VBT Expected: The sum of qx's from the 2001 VBT appended to each individual in the dataset and these individual qx's are further summed by cholesterol group.
- Deaths/2001 VBT A/E: The ratio of the actual death rate observed to the expected death rate based on the summed qx's from the 2001 VBT before any other adjustments.
- Standardized A/E: This divides the A/E results from the deaths/2001 A/E by 172% to standardize the data.
- Cox HR: These are the results from a Cox Proportional Hazards regression that show the hazard ratios for each cholesterol group relative to the hazard associated with the referent group whose cholesterol readings are 181-200 mg/dl. This equation includes the categorical cholesterol ranges along with covariates age, gender and smoking. (Hazard ratios for these covariates are: age 1.093, gender 1.396 hazard for male gender relative to female, tobacco 1.730 for cigarette smoking relative to all others.) In other words, results are adjusted for age, gender and smoking.

By looking at the last two columns of Table 6, it can be seen that the hazard ratios from an age, gender and smoking adjusted Cox Model (Cox HR) are very close to results achieved using traditional mortality study techniques when applying the actuarial calculations to the data (Standardized A/E).

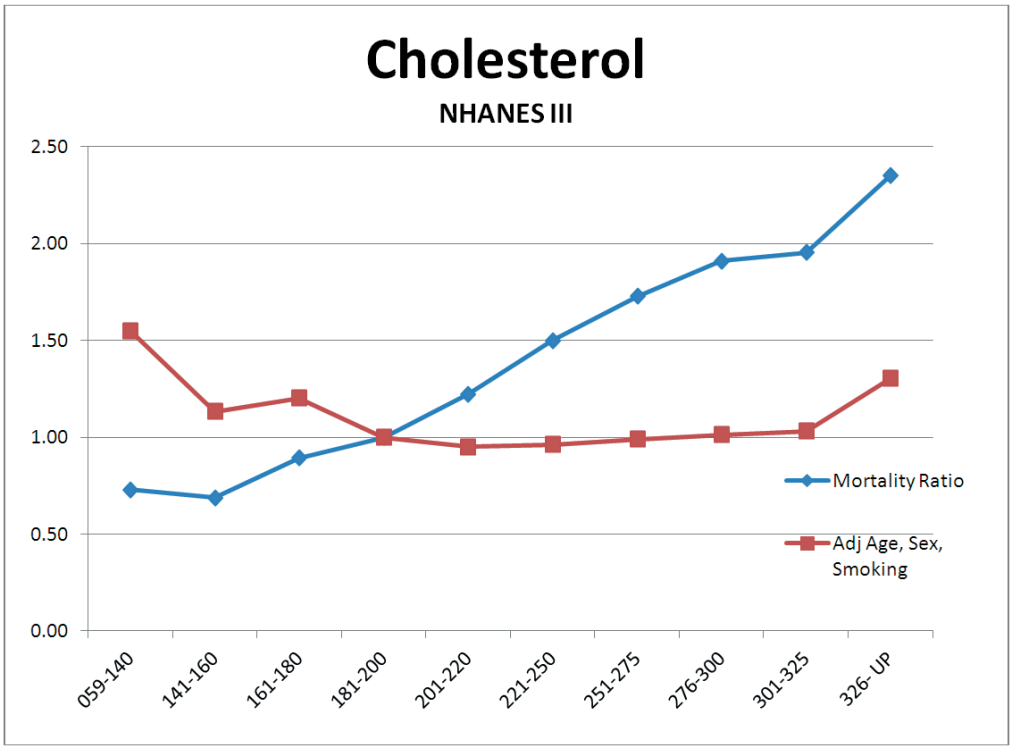
After adjusting for age, sex and smoking, the mortality curve for the independent effect of cholesterol shifts considerably. The change in the shape of the curve can be traced to the fact that the confounding variables age, sex and smoking have been introduced into the Cox equation, or adjusted for by use of an actuarial mortality table that takes into consideration age, sex and smoking status. By viewing Graph 2 (next page) it can be seen that the independent influence of cholesterol on mortality takes on a new U-shaped pattern (red line) similar to results described elsewhere. Note the hazard ratios after adjusting for age, sex and smoking and contrast them with the hazard ratios found when only adjusting for cholesterol. The two lines are plotted on Graph 2 to show how the curve has changed.

Conclusion

The first goal of this article was to clarify the close relationship between hazard ratios and mortality ratios. The second goal of this article was to show the close correlation between mortality results for cholesterol when using Cox multivariate regression adjusting for age, sex and smoking, and findings using traditional actuarial calculations. This also sheds light on the importance of adjusting for these other variables when considering the independent influence of cholesterol on mortality.

Other variables such as low albumin, underweight or cancer history could have been considered for this article as well. In fact, there are numerous combinations of potential confounding variables and these

Graph 2



variables provide ammunition for reaching different conclusions. Perhaps this illustrates why outcomes from one clinical study compared to another may suggest conclusions that don't agree. But that wasn't the purpose for this article. The main purpose for this article was to shed light on the relationship between multivariate regression and traditional actuarial mortality findings. This shows, when using the same variables, the results are very similar. The second goal of this paper was to describe the relationships between hazard ratios and mortality ratios.

References

1 http://soa.org/Research/Experience-Study/Ind-Life/Tables/final-report-life-insurance-valuation.aspx.
2 Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. www.cdc.gov/nchs/nhanes/nh3data.htm.
3 Wesley D, Cox HF. *J Ins Med.* 2011;42(2-4):62-75.
4 Cox DR. Regression models and life tables. *J R Stat Soc B.* 1972;34:187–220.
5 www.ncbi.nlm.nih.gov/pmc/articles/PMC1065454/pdf/hsre-search00098-0068.pdf.
6 Wesley D, Cox HF. *J Ins Med.* 2011;42(2-4):62-75.

About the Author

Doug Ingle, FALU, FLMI, is Vice President of Underwriting Research at Hannover Life Reassurance Company of America. He has over 37 years of underwriting experience split between direct writers and reinsurers. Doug completed Beginning and Advanced Life Table Methodology, co-founded a program used for Preferred Risk underwriting and pricing, has written a text on interpreting resting EKGs for underwriters, and authored two chapters of the ALU curriculum. He is a current member of the ILEC and UCT, is a current member and former chair of both the MMLC and UESC, and has been published in *ON THE RISK*, *Journal of Insurance Medicine*, *Contingencies* and *Best's Review*. Doug has spoken at many regional and national SOA, AAIM and AHOU meetings.