LifeScoreLabs℠

White Paper:

# Development of LifeScore Med360℠

Author: Marc Maier, PhD

February 2018

# Content

# Development of LifeScore Med360

The practice of underwriting for life insurance relies on multiple sources of medical, behavioral, and financial data to produce estimates of mortality and financial risk. Life insurance carriers are uniquely positioned to maintain data assets that provide individual-level profiles spanning these types of personal data attributes. As one of the largest life insurance companies in the U.S., MassMutual has strategically collected and curated a large historical database of its applicants over a 15-year period. This covers nearly one million applications, underwriting decisions, and ground-truth mortality experience.

This massive data asset combined with state-of-the-art data science methodology has produced a powerful framework for predictive modeling, grounded by medical expertise and professional underwriting experience. By applying methods based on statistical and machine learning research, MassMutual has developed a life score (termed LifeScore Med360, or LS360) that is an accurate, point-in-time estimate of individual mortality risk. We demonstrate a strong relationship between the score and relative mortality. Statistical models that accurately estimate risk provide a flexible, yet consistent platform for insurance carriers to offer competitive prices to consumers and generate innovative approaches to issuing life insurance. This white paper provides an overview of (1) the modeling methodology, (2) how to interpret the life score, (3) the performance of the life score and its ability to stratify mortality risk, and (4) the underlying data asset used to build the model.

# Methodology

The fields of statistics, computer science, and more specifically, data science, have developed tools and methods that leverage data sets at scale, with respect to both sample size and variables. With nearly one million records and hundreds of variables, the underwriting data set at MassMutual presents an ideal opportunity for data science to extract value. In particular, we develop models that estimate mortality risk given the medical and behavioral attributes that exist within underwriting data sets.

The vast majority of predictive modeling tasks aim to estimate the probability of some discrete or continuous outcome. For example, spam filters use classifiers that predict the likelihood an e-mail should be marked as spam and financial analysts may develop regression models to predict the future price of a security. In the context of survival modeling, however, the outcome of interest is the duration until the study period ends for each individual and an indicator that specifies if the duration ended in an event. This type of modeling forms the basis of survival analysis, distinct from classification and regression techniques.

In contrast with traditional predictive modeling tasks, the objective of survival analysis is to estimate the survival function. The survival function, defined as S(t) = Pr(T>t), describes the probability that an event, occurring at random variable time T, occurs later than some given time t. Also of primary interest is the hazard rate, $\boldsymbol{\lambda}(t)$, which is the rate of the event at time t conditioned on having survived until time t. In actuarial science, the hazard is often denoted as μ, and describes the mortality rate for a given attained age. Additionally, the cumulative hazard function, typically defined as !(t), is the integral of the hazard up to time t, and is related to the survival function as !(t) = - log S(t). There exist straightforward, nonparametric estimators,

TABLE 1: **Model inputs are captured by lab tests and health questionnaires**

| Data Source | Variables |
|---|---|
| Demographic | Age |
| Demographic | Sex |
| Biophysical | Build (height, weight) |
| Biophysical | Systolic/Diastolic Blood Pressure, Pulse |
| Lab Test | Liver Function (total bilirubin, AST, ALT alkaline phosphatase, GGT) |
| Lab Test | Kidney Function (creatinine, BUN) |
| Lab Test | Blood Protein (albumin, globulin, total protein) |
| Lab Test | Lipids (cholesterol, HDL, LDL, triglycerides) |
| Lab Test | Urine Protein (creatinine, total protein, microalbumin) |
| Lab Test | Blood Sugars (glucose, fructosamine, hemoglobin A1C) |
| Lab Test | Indicator Tests (HIV, HCV, PSA, Cocaine, Nicotine) |
| Personal Health History | Blood Disorder |
| Personal Health History | Cancer |
| Personal Health History | Digestive Condition |
| Personal Health History | Disability Claims |
| Personal Health History | Endocrine Disorder |
| Personal Health History | Heart Condition |
| Personal Health History | Mental Condition |
| Personal Health History | Muscular Disorder |
| Personal Health History | Nervous System Disorder |
| Personal Health History | Reproductive Disorder |
| Personal Health History | Respiratory Disorder |
| Personal Health History | Urinary Tract Condition |
| Behavioral History | History of Smoking |
| Behavioral History | Motor Vehicle Convictions |
| Family History | Cancer |
| Family History | Cardiovascular Disease |
| Family History | Diabetes |

namely the Kaplan-Meier and Nelson-Aalen estimators, that compute these quantities directly from observed survival-related data.

The primary goal of predictive modeling in the survival context is to develop estimates of the survival, hazard, or cumulative hazard functions with respect to a set of observed covariates. In the underwriting-for-mortality setting, the covariates are medical and behavior attributes of life insurance applicants and the event is mortality (see below for a complete set of variables used to develop the life score). The techniques used to estimate these functions fundamentally require a different set of statistics as the time-to-event of mortality is unknown for most individuals. This is referred to as right-censored data because the date of birth is known but the date of death is unobserved for a large set of individuals. Missing survival information is a key characteristic of survival analysis, in which survival times may be censored at the beginning, end, or even middle of study periods.

There is a well-established set of methods commonly employed among academic and industrial practitioners of survival analysis. LifeScore Med360 is generated by a cutting-edge modeling methodology, rooted in statistics and machine learning research, that directly estimates the cumulative hazard function. The basis for this type of model has been used with success across various fields, including finance, medicine, sports, and ecology. This nonparametric, adaptive model captures interactions and non-linear dependencies that are more subtle and complex than can be described in published medical literature. Simultaneously, the models also recover conventional medical knowledge related to mortality risk factors and standard laboratory tests. Furthermore, this modeling approach vastly outperforms traditional statistical models that are capable of merely capturing global, population-level trends. This is especially true for high-dimensional, heterogeneous data composed of a large number of sub-populations such as the underwriting data set at MassMutual.

**Classes of model inputs**

The model described above relies on nearly 50 raw inputs and internally generates additional medically relevant features as combinations of variables (e.g., BMI). The main inputs are commonly collected through biophysical examinations, blood and urine specimens, and applicant health history questionnaires. Table 1 lists these classes of variables in more detail.

# The Life Score

The estimated cumulative hazard values produced by the mortality model are subsequently standardized to provide a consistent life score. The score has a range of 0–100 and reflects the relative risk among 5-year age band/gender/smoker cohorts. These cohorts capture the primary factors in actuarial mortality studies. Conditioned on cohort, LS360 is derived from the quantiles of the empirical distribution of the estimated 10-year cumulative hazard across the set of individuals used in model training (see the Data section below). Figure 1 demonstrates that the proportion of individuals in these cohorts are consistent across the range of LS360 scores.

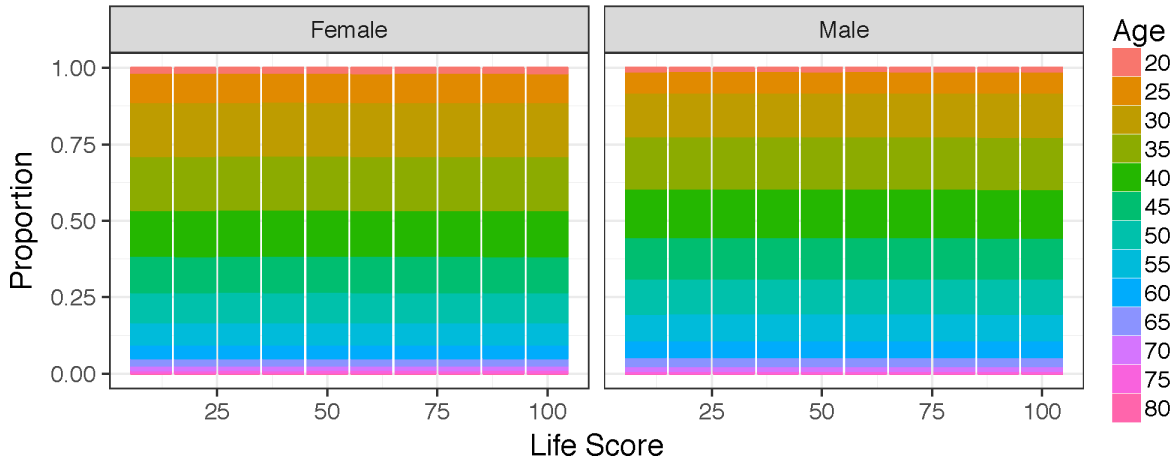


**FIGURE 1: The proportion of individuals in each decile of LS360 are consistent across 5-year age and sex bands.**

**Example**

If Carlos is a 55-year-old non-smoking male with an LS360 of 87, he can be compared directly against (and has lower mortality risk than) Barry, another 55-year-old male non-smoking male with a score of 53. However, if Amy is a 35-year-old non-smoking female with a score of 87, she does not present the same mortality risk as Carlos. Furthermore, if Amy's life score were 23, it would not necessarily mean she has a higher absolute mortality risk than Carlos.

TABLE 2: **Actual-to-expected ratios and 95% confidence intervals for each decile of LS360. The life score correlates with A/Es. Number of person-years for each stratum shown to provide credibility support.**

| Life Score Decile | Number of Person-Years | Actual-to-Expected ± 95% CI |
|---|---|---|
| 10 | 835,912 | 239.22 ± 4.351 |
| 20 | 847,924 | 124.31 ± 3.079 |
| 30 | 869,003 | 99.33 ± 2.697 |
| 40 | 887,603 | 84.56 ± 2.456 |
| 50 | 905,965 | 74.87 ± 2.270 |
| 60 | 922,019 | 67.46 ± 2.125 |
| 70 | 933,060 | 60.74 ± 2.007 |
| 80 | 944,886 | 55.40 ± 1.895 |
| 90 | 959,834 | 56.29 ± 1.896 |
| 100 | 973,739 | 49.42 ± 1.773 |

# Performance

As described in the previous section, the mortality model generates a score in the range of 0–100 relative to age, sex, and smoking status cohorts. The life score correlates inversely with mortality risk. Using the 2015 SOA VBT as a baseline for expected mortality, we show in Table 2 the actual-to-expected ratios and 95% confidence intervals across each decile of LS360. In aggregate, the score provides a linear decrease in actual-to-expected values for scores in the highest 8 deciles and ultimately pushes the cases with the worst mortality experience to the lowest 2 deciles.

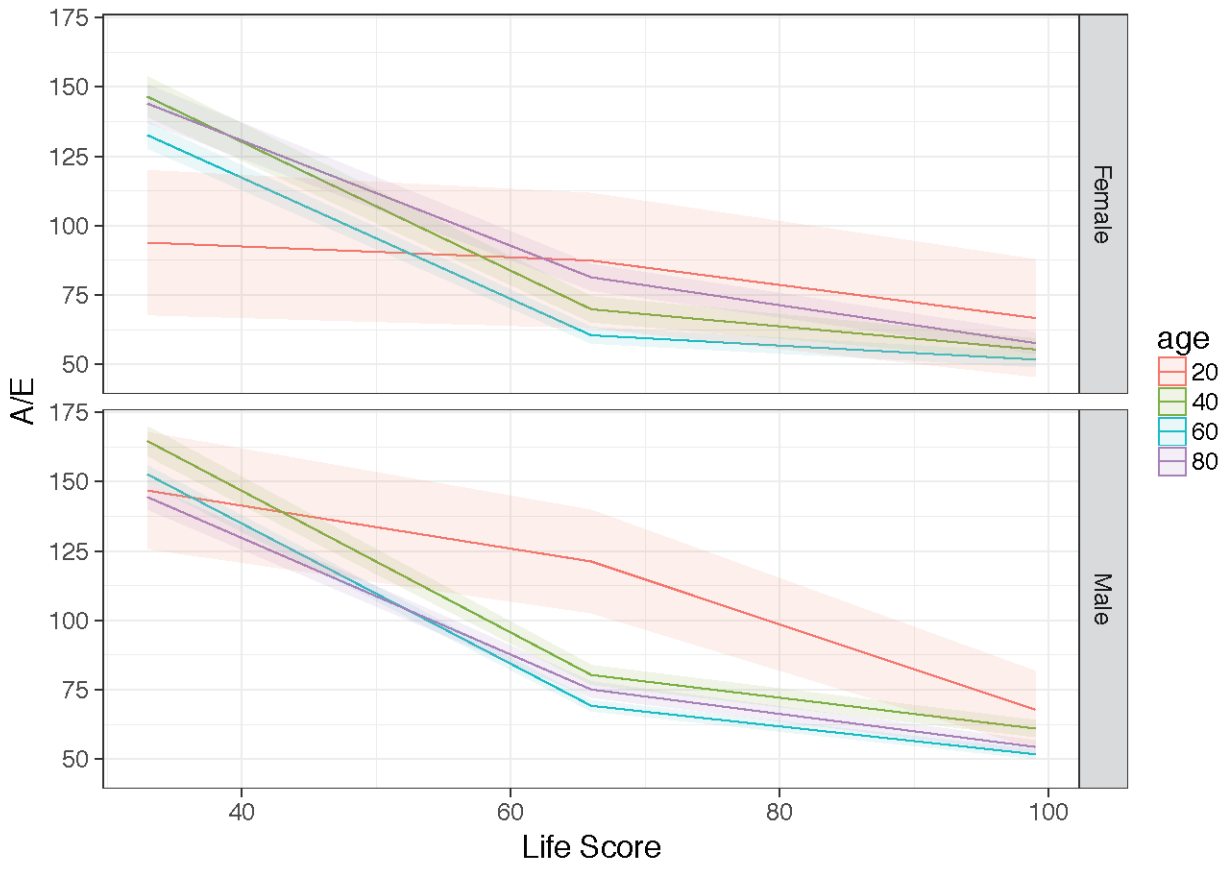


FIGURE 2: Actual-to-expected ratios as a function of life score stratified by applicant sex and 20-year age bands. Both graphics depict a monotonically decreasing relationship in A/Es as life score increases. Trends are shown with 95% confidence

Subdividing lives across different facets reduces the credibility (i.e., expands the confidence interval) of the actual-to-expected results. However, Figure 2 demonstrates that the life score is robust across age and sex. As the actual number of deaths is quite small in the 20-year age band, the confidence intervals are much larger than the other bands. The female actual-to-expected curves are systematically lower than the males, which is expected given that overall mortality risk in the data is lower for females than males (see the Data section below).

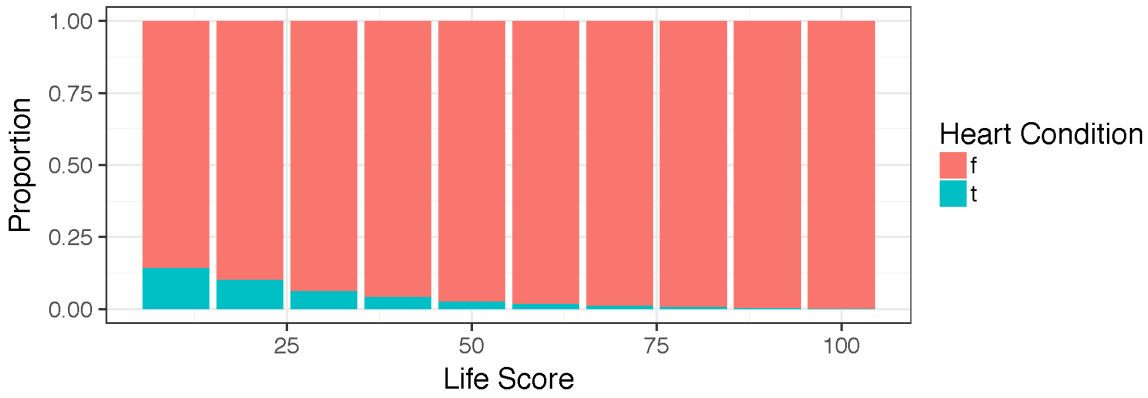


**FIGURE 3: Incidence of heart condition as life score decile increases. The proportion ranges from ~0.22% in the first decile to ~14.4% in the tenth decile, with a gradually increasing proportion in between.**

Additionally, we can demonstrate how medical impairments are stratified across the life score. In Figures 3 and 4, the relative proportion of heart condition incidence and BMI bands are displayed within each decile of the LS360. These statistics clearly depict how the score reflects the effect that BMI and heart condition have on mortality risk.

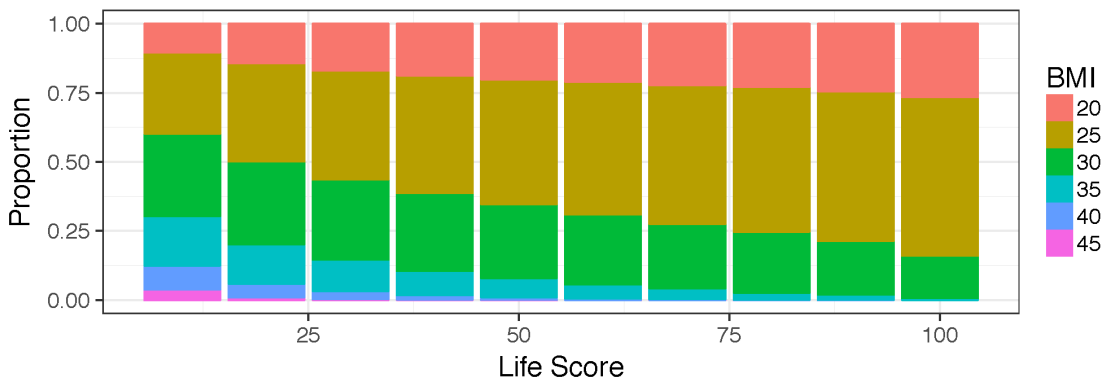


**FIGURE 4: Distribution of BMI as life score decile increases. The highest scores have a higher proportion of healthy-range BMI values. As LS360 decreases, the proportion of upper and lower BMI extremes increases.**

# Data

The mortality model was trained on 15 years of application history at MassMutual. The time period spans 1999–2014 and covers all cases for which a lab test was ordered on an individual. As a pure model of mortality risk, the training data includes every applicant for life and disability insurance. In total, this accounts for nearly one million applications. The final training data set, after removing applications with a high degree of missing values, totals 908,414 applications. The number of observed mortality events, total person-years, and aggregate actual-to-expected statistics are shown in Table 3. Note that the model training data represents a healthier set of individuals than the standard, fully underwritten life insurance population.

TABLE 3: **Total number of applicants, deaths, person-years of exposure and overall A/E in the model training data**

| Number of Applicants | Number of Events | Number of Person-Years | Actual-to-Expected ± 95% CI |
|---|---|---|---|
| 908,414 | 15,769 | 9,159,469 | 85.91 ± 0.684 |

Tables 4 and 5 display the actual-to-expected values stratified by applicant sex and 10-year age bands, respectively. We observe that females, in aggregate, represent a slightly lower relative mortality. There is a consistent relative mortality effect across age bands, with the exception of 50- and 60-year olds.

We can also subdivide the data by medically relevant attributes. While Figures 3 and 4 show the relationship between the life score and heart condition and BMI, the data also reflect a similar relationship to relative mortality, which is expected given the mortality model has learned these dependencies directly from the data. Table 6 shows a large relative mortality increase for both males and females that have a prior heart condition. The actual-to-expected ratios for males and females are correspondingly lower than the overall pool (shown in Table 4) when lives with a heart condition are removed. Table 7 and the associated Figure 4 display the actual-to-expected ratios for BMI stratified by applicant sex.

TABLE 4: **Total number of person-years of exposure and A/E ratios stratified by applicant sex in the model training data.**

| Sex | Number of Person-Years | Actual-to-Expected ± 95% CI |
|---|---|---|
| Female | 3,561,482 | 83.23 ± 1.198 |
| Male | 5,597,987 | 87.15 ± 0.833 |

TABLE 5: **Total number of person-years of exposure and A/E ratios stratified by applicant 10-year age bands in the model training data.**

| Age | Number of Person-Years | Actual-to-Expected ± 95% CI |
|---|---|---|
| 20 | 517,861 | 83.09 ± 6.610 |
| 30 | 2,442,984 | 88.58 ± 2.958 |
| 40 | 3,134,020 | 88.14 ± 1.832 |
| 50 | 1,755,659 | 83.76 ± 1.532 |
| 60 | 1,002,568 | 82.10 ± 1.302 |
| 70 | 223,199 | 89.34 ± 1.718 |
| 80 | 81,086 | 88.07 ± 1.723 |
| 90 | 2,093 | 97.18 ± 9.102 |

TABLE 7: **Total number of person-years of exposure and A/E ratios stratified by applicant sex and 5-point bands of BMI in the model training data.**

| Sex | BMI Band | Number of Person-Years | Actual-to-Expected ± 95% CI |
|---|---|---|---|
| Female | 20 | 494,482 | 76.82 ± 4.124 |
| Female | 25 | 551,938 | 74.64 ± 3.328 |
| Female | 30 | 241,201 | 73.06 ± 4.658 |
| Female | 35 | 98,682 | 105.84 ± 9.144 |
| Female | 40 | 37,047 | 117.57 ± 17.149 |
| Female | 45 | 12,562 | 166.56 ± 38.212 |
| Male | 20 | 190,683 | 86.70 ± 5.793 |
| Male | 25 | 1,070,765 | 80.08 ± 2.168 |
| Male | 30 | 709,593 | 84.38 ± 2.613 |
| Male | 35 | 200,486 | 104.03 ± 5.537 |
| Male | 40 | 49,241 | 146.96 ± 13.825 |
| Male | 45 | 11,613 | 197.92 ± 35.547 |

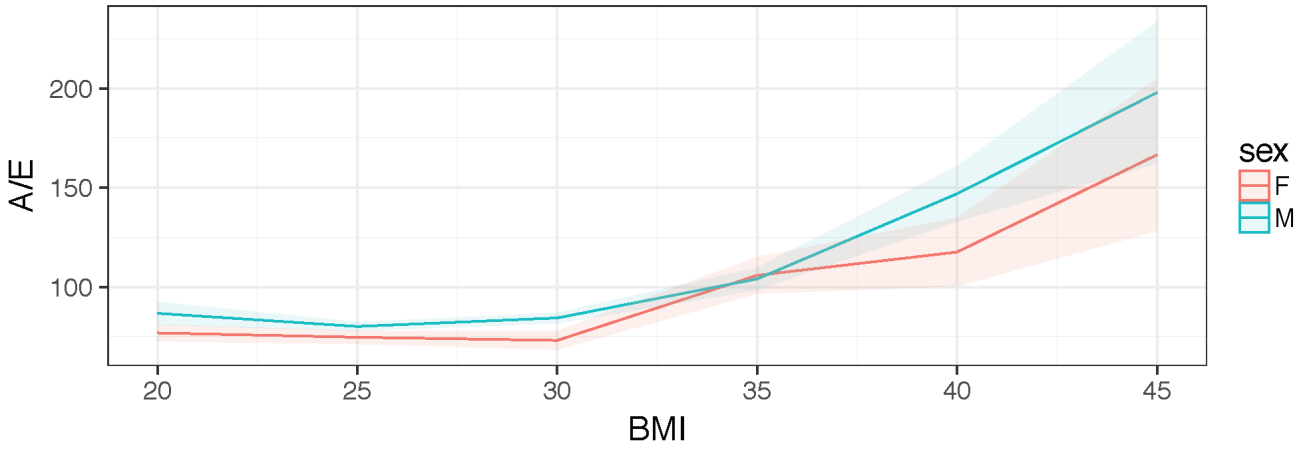


**FIGURE 5: Visualization of the data in Table 7, showing increasing A/E ratios by 5-point bands of BMI, stratified by applicant sex.**

# Conclusion

Leveraging a data set spanning 15 years of applications at MassMutual, this technical paper described an accurate, industry-leading model that estimates mortality risk. The output of this model generates an individualized life score that can directly compare applicants on a consistent basis relative to their demographic cohorts. LifeScore Med360 strongly correlates with relative mortality and captures a large number of statistical interactions among medical and behavioral attributes and their impact on mortality risk. Historical data paired with modern data science capabilities provides an unprecedented opportunity in the life insurance industry to disrupt the underwriting status quo.

THANK YOU