Are their life expectancy ratings accurate? How to tell

By Robert Shavelle, Ph.D.

The value of an existing life insurance policy depends critically on how long the insured will live. Various commercial rating services provide estimates of individuals' likely longevity, but the reliability of their estimates has rarely been correctly evaluated.

How should we compare a service's predictions for a large group of policyholders with the actual mortality experience observed during follow-up? That is, what is the right method to compare the actual number of deaths (A) with the expected number (E)?

Many approaches have been used, some quite wrong. The correct method, which we have explained in detail elsewhere,[1] seems not to be widely known or used in the life settlement industry. In what follows we provide a short non-technical discussion and illustration.

The Problem

In any population, persons will live longer or shorter than the life expectancy (the average), often much longer or shorter. An analysis of the insured's risk factors by a life expectancy provider – which can be referred to as a "rating" – results in not merely a single number, the life expectancy, but with an estimate of the entire survival distribution. That is, it gives the probabilities that the person will survive for one more year, two more years, and so on.

Suppose that a rating service has provided its customary survival estimates for 100 insureds. Suppose that these insured have been followed for 5 years, and we know that 10 have died. Is the number we expected to have died, or is it too high or too low? And could this have been solely due to chance?

To an extent that we find surprising, many investors have not obtained independent analysis of the accuracy of the life expectancy predictions they regularly rely upon. Some have incorrectly assumed that these estimates are instead a prediction of the insured's actual survival time (and used these in a term-certain analysis of cash flow), rather than just a convenient summary measure of the survival probabilities. Others merely average the life expectancies from the various providers, and assume that the result is near enough to the truth. Perhaps this is the reason for recent calls in the industry for a standard way to measure the accuracy of life expectancy reports.[2] As we noted at the outset, there already is a standard method. We now outline it.

The Solution

The correct method for comparing A to E does not work with individual insured persons, but instead with *person-years of exposure time*. For each year of each person's follow up, the actual mber of deaths is simply 1 if the person died in that year and 0 if not. The *expected number* of deaths for that year, which is approximately the same as the probability of death in that year, is in effect sup-

(Continued on page 20)

Table 1. Preliminary results

	Time Since Underwriting (Years)			
Statistic	All	0.0-0.9	1.0-2.9	3.0+
A	286	70	164	52
E	230	48	131	51
A/E	1.2	1.5	1.2	1.0

Robert Shavelle, Ph.D, is technical director of the Life Expectancy Project, San Francisco and a fellow of the American Academy for Cerebal Palsy and Developmental Medicine. Shavelle is a medical researcher on life expectancy, and author of 75 articles.

(Continued from page 19)

plied by the rating service, and it should reflect the individual's age, sex, and profile of medical factors. When the actual and expected numbers are aggregated over a group of interest (e.g., males, persons of age over 70, or persons with policies of a $1M or more), they are referred to as A and E respectively.

This is a standard method in epidemiology.[3] It is commonly used to compare the survival experience of a study group (e.g., those with a particular disease) to that of an appropriate reference group (e.g., the general population).[4] Other approaches to this task are at best incomplete and less powerful.[5] What makes the application here different is that we are not assessing whether the actual number of deaths, A, is consistent with the known expected number, E. Instead, *we are assessing whether the expected deaths predicted by the rating firm,[6] E, are consistent with the observed A.*

The ratio of A to E is often called a *standardized mortality ratio* (SMR).[3] If we observe more deaths than expected then the SMR is greater than 1 (or 100%). This indicates that the raters are underestimating mortality (or overestimating life expectancy). If conversely we have fewer deaths than expected, then the SMR will be less than 1 and the rates are overestimating mortality. One generally wishes to see SMRs "close" to 1.

If the ratings are "good" the SMR should be reasonably close to 1 for all groupings of interest. Such groupings would be the entire group and also, for example, only males, all persons during the first few years after policy purchase, those with heart disease, policies from a particular source, and policies with a large face value. A good rater, rating system, or rating firm is one that performs well in all tests.[7]

Example

Consider a large portfolio of persons who submitted their demographic and medical data in anticipation of selling their life insurance policies. A rating service

(Continued on page 26)

Table 2. Further Results

Risk Group	Average MM	A	E	A/E
Best Quintile	68	48	43	1.1
2nd Quintile	78	41	39	1.0
3rd Quintile	90	48	43	1.1
4th Quintile	107	54	46	1.2
Worst Quintile	165	95	58	1.6

FOOTNOTES

1 Shavelle RM, Strauss DJ (2009). Rating the raters: Evaluating the predictions from a life expectancy rating service. Journal of Insurance Medicine, Volume 41, Number 3. Available at http://www.LifeExpectancy.org/articles/rating.pdf.

2 Deal Flow Media. The Life Settlements Wire. "Life Expectancy Underwriters Support Standardized Accuracy Measurements". Posted September 25, 2008 2:45 pm. Available at http://LifeSettlements.DealFlowMedia.com/wires/092208.cfm. Accessed August 26, 2009.

3 Kahn HA, Sempos CT. (1989). Statistical methods in epidemiology. Oxford: Oxford University Press, page 96. Note that epidemiologists and others prefer to use the letter O to represent the "observed deaths", rather than A for "actual deaths".

4 See, for example, Finkelstein DM, Muzikansky A, Schoenfeld DA (2003). Comparing survival of a sample to that of a standard population. Journal of the National Cancer Institute, 95:1434-1439.

5 Consider, for example, the approach detailed by Finkelstein.[4] In this one plots the actual and expected survival curves. An equivalent method has been adopted in the life settlement industry by one of the major life expectancy providers. In both cases the methods provide no useful information either early on or after most persons have died. Further, their use is somewhat problematic in between because the fit may vary due to small sample sizes. Finally, there is no "overall" test that aggregates across time. We thus cannot answer the most important question: "How many deaths were we short or in surplus?"

6 The expected number of deaths is computed for each interval of follow-up time, for each person, based on the estimated mortality rates for that person during that period. It is beyond the scope of the present article to explain further. For details see the source article.[1]

7 Issues of sampling variation (or chance) error will arise if the sample sizes are relatively small. For these applications one should construct so-called confidence intervals for the SMRs. If the intervals include 1.0 (100%), then the putatively poor results could have been due to chance. If not, then they are unlikely to have been due to chance. See, for example, pages 98-100 of Kahn & Sempos.[3]

(Continued from page 20)

used this information to estimate their survival. Using the methods outlined here, the actual number of deaths was then compared with the expected number. Preliminary results by time since underwriting are (see Table 1):

Overall there were 20% more deaths than expected (SMR = A/E = 1.2). The rating service underestimated mortality in the first (1.5) and second/third (1.2) years post underwriting. Thereafter their ratings appeared to be good (1.0).

A second partitioning of the person-years was into five groups defined by the mortality risk. Thus the lowest risk or "best quintile" consists of the 20% of the person-years with the lowest predicted death rates, and so on (see Table 2.)

MM denotes the mortality multiplier, which essentially compares the predicted life expectancy to those implicit in the VBT2001 non-smoking mortality rates. As can be seen, the predictions were excellent for the first three risk quintiles – those with the lowest MM's. This indicates that (1) the chosen baseline mortality rate was a good fit to the empirical data, and (2) the relatively mild adjustments were done correctly. The predictions were unsatisfactory, however, in the worst quintile (SMR = 1.6). Additional analysis found that this excess was largely restricted to the first 3 years post underwriting and that the ratings were too optimistic at younger ages. It was thus possible to pinpoint the type of insureds where the provider's results were unsatisfactory. This, of course, is only an example of the myriad of analyses that could, and should, be done.

Summing Up

The method described here can be used to determine the accuracy of life expectancy estimates. It can thus be used to evaluate rating firms, individual raters, and the ratings themselves. We can also assess if a firm or its raters are improving with time, or if they are particularly good, or bad, for specific types of insureds (e.g., older males), diseases (e.g., Parkinsons), or policies (e.g., $5M+).

These scientific comparisons can be made by any investor or other person who has both (1) the life expectancy predictions, and (2) the observed mortality experience of each insured. We hope this article shows that it can be done easily and correctly. ■