JOURNAL OF INSURANCE MEDICINE

J Insur Med 2009;41:178–190

MORTALITY

Rating the Raters: Evaluating the Predictions from a Life Expectancy Rating Service

Robert Shavelle, PhD, MBA; David Strauss, PhD, FASA

There is a growing market for the resale of life insurance policies to third party investors. A key factor in the valuation of a policy is how long the insured is likely to live. Various commercial rating services offer to provide estimates of individuals' likely longevity, but the reliability of their estimates has rarely been correctly evaluated. The question is how to compare the estimates provided for a large group of policy-holders with the observed ''truth data'' — the actual mortality experience observed during follow-up.

Various approaches to this have been used in practice, some of them quite wrong. The correct method does not seem to be practiced or widely known in the life settlement industry. It is based on a comparison of observed and expected deaths computed on the basis of person-years of exposure rather than of individual persons. The method is explained in detail here and illustrated with the results from a large portfolio of policies.

Address: Life Expectancy Project, 1439 – 17th Avenue, San Francisco, CA 94122; ph: 415-731-0240; Shavelle@LifeExpectancy.org.

Correspondent: Robert Shavelle

Key words: Life expectancy, rating, litigation, senior, settlement, industry, rating service, figures, observed, expected, actual, deaths, standardized mortality ratio, SMR, overestimating, underestimating, mortality, survival, life table, A to E, O to E.

Received: October 6, 2009

Accepted: November 1, 2009

1. INTRODUCTION

Readers of this journal will be aware that there is a growing market for the resale of life insurance policies to third party investors. Such investors pay a lump sum to the policyholder, and assume annual premium payments, in exchange for becoming the beneficiary. These transactions are referred to as life settlements (or senior settlements).

A critical factor in valuing such an investment is how long the insured is likely to live. If he lives much longer than expected the investor not only pays more premiums than expected, but also must wait longer to receive the death benefit. To an extent that we find surprising, many investors have not obtained independent analysis of the accuracy of the predictions.

There is, of course, a risk that these rating services lack the expertise to provide reliable ratings. Perhaps equally seriously, there is often a conflict of interest because the customers of the rating services are primarily the *sellers* of the policies, not the buyers. The shorter the policyholder's life expectancy is considered to be, the more valuable the policy becomes. Further, the brokers' commissions are generally based on the selling price. Thus both brokers and vendors prefer short life expectancies, a preference that is often made clear to the rating service. Perhaps as a result of this, there have been numerous lawsuits brought by investors against rating services, alleging that – as a result of either negligence or fraud – the service has systematically underestimated

the life expectancies, to the detriment of the purchasers of the policies.

Suppose there is a portfolio of life insurance policies, for each of which a rating service has provided an estimate of life expectancy (or some other measure of prognosis for survival, such as the 80th percentile of the survival distribution). Suppose, too, that there is some follow-up on the policyholders (''truth data''). For example, it may be that the assessments were made four years ago, and it is known whether each policyholder is now alive, together with date of death for those who died. The question then arises: How do we use this information to assess the performance of the rating service? How do we rate the raters?

In extreme cases it will be obvious that the ratings are not consistent with the truth data. Suppose, for example, that there are 20 individuals rated, and all are estimated to have a life expectancy of exactly two years. If we find that 17 of the 20 individuals are still alive 5 years later, it is clear that survival times have been grossly underestimated. In general, however, it is not immediately obvious how the ratings and the truth data are properly to be compared. This may be why there have been recent calls in the industry for a standard way to measure the accuracy of life expectancy reports.[1]

Our purpose here is to show precisely how this should be done. To our knowledge, the correct method has not been explained in previous literature or textbooks. We will also show that some alternative approaches, which may seem plausible at first sight, are in fact wrong. We illustrate the correct methodology with several examples, including a large-scale study of persons who offered to sell their life insurance policies. We note the methodology can also be applied to situations other than the life settlement industry. It can, for example, be used to assess the ratings provided by life insurance underwriters or other providers of life expectancy estimates.

2. WHAT SHOULD THE RATING SERVICE PROVIDE?

Perhaps the first issue to consider is what precisely the rating service should provide for a given individual. In practice it is common to report a single number, such as the median survival time (ie, the time at which there is a 50% chance of being alive) or the life expectancy (the *average* number of additional years lived in a large group of similar persons). Even if these estimates are correct, however, they are not sufficient to permit a rational analysis of the value of the investment. What is needed is an estimate of the individual's mortality risk in the present year and for every subsequent year. Equivalently, *one requires the individual's complete life table*.

We give a simple example to illustrate this point. Suppose that we are the beneficiary of a life insurance policy with a $1 death benefit, and the life expectancy is known to be exactly 20 years. Assume a discount rate of 3%; that is, a dollar to be paid to us next year has a present value of 1/(1.03) = roughly 97 cents. For simplicity, let us ignore the premiums that must be paid while the insured is still alive. What is the present value of the death benefit?

It turns out that we do not have sufficient information. Consider, for example, Scenario A, in which the insured is known to have exactly 20 years of life left. In this case the present value is $1 times 0.97 raised to the 20th power, which is 54 cents.

By contrast, consider Scenario B in which there is a 50% chance that the insured will die tomorrow and another 50% chance that he will live 40 more years. The average survival time, which by definition is the life expectancy, is still exactly 20 years. But this time the expected present value is: $0.50*0.97^{0} + 0.50*0.97^{40} = 0.50*1.0 + 0.50*0.30 = 65$ cents.

That is, half the time we receive the benefit immediately, with present value $1, and the other half of the time we must wait 40 years, with present value 30 cents. The average is thus 65 cents, considerably more than the

Table 1. Life Table for 65-Year-Old Non-Smoking Males

Age	l(x)	d(x)	q(x)	m(x)	L(x)	T(x)	e(x)
65	100000	247	0.0025	0.0025	99877	2000634	20.0
66	99753	408	0.0041	0.0041	99549	1900757	19.1
67	99345	579	0.0058	0.0058	99055	1801208	18.1
68	98766	759	0.0077	0.0077	98387	1702153	17.2
69	98007	940	0.0096	0.0096	97537	1603766	16.4
70	97067	1117	0.0115	0.0116	96509	1506229	15.5
80	74810	4182	0.0559	0.0575	72719	621855	8.3
90	26280	4463	0.1698	0.1861	24049	109640	4.2
100	1762	570	0.3235	0.3909	1477	4062	2.3

54 cents under Scenario A. The example illustrates that a single summary statistic, such as the life expectancy, is not sufficient: a complete life table is needed.

An abbreviated life table for 65-year-old, non-smoking males who recently qualified for standard insurance[10] is in Table 1.

Two quantities are of particular note here. The first is the survivorship column, l(x), which begins with 100,000 persons alive at the initial age. With a simple re-scaling, these values may also be viewed as a survival curve beginning with 100%.

The second is the mortality rate at each age x, denoted by m(x). These are computed as the number of deaths per year, and may also be referred to as ''occurrence-exposure rates'' or the ''instantaneous force of mortality''. Note that they differ from the annual mortality probabilities, q(x), or the number of deaths per person, which must lie between 0 and 1 inclusive. Readers interested in further technical details of the life table are referred to standard sources.[2,3]

For the remainder of this article we assume that a complete life table has been provided for each insured. In practice the life table may have to be constructed on the basis of certain assumptions. For example, if the rating service provided only the life expectancy, it may be reasonable to multiply some standard mortality rates at every age by a suitable constant to produce the life table that yields this target life expectancy. There are many other possible methods.[4] This exercise involves extra work and extra assumptions, but we emphasize that a complete analysis of the survival estimates provided by a rating service is not possible without it.

3. THE FRAMEWORK FOR RATING THE RATERS

We have information on the survival of n persons. Let t_i be the time of death or censoring (ie, the end of the follow-up period) for the i^{th} person. We assume that the rater has provided a *survival function* $S_i(t)$ for the i^{th} person, where $S_i(t)$ is the probability that he will be alive at time $t > 0$.

If a life table has been provided, the function $S_i(t)$ is obtained from the ''l(t)'' column of the life table; it is simply l(t)/l(0), where the radix l(0) is 100,000 in most life tables. For example, if the life table indicates that there are 40,000 persons still alive at age t, then S(t) = 0.40. As noted previously, in practice most rating services do not provide a complete life table for each subject, and the table must be inferred from information that they do provide.

Let d_i be 1 if the person died in the follow-up period, and 0 if he did not. The observed number of deaths is $O = \Sigma d_i$. [Note: In the life settlement industry it is common to refer to the *actual* number of deaths, denoted by A, but we use the more standard notation ''O'' here.] The more difficult question is how to compute E, the *expected* number of deaths.

In the simplest possible case, that of a group of persons who are the same age and sex, and followed for exactly one year each, the expected number of deaths is merely the number of people (times one year each) multiplied by the expected annual rate of death.

In general, one computes E as the sum – over all possible combinations of age and sex – of the product of the mortality rate and the exposure time. This is because a mortality rate is, by definition, the number of deaths per exposure time and, thus, the rate multiplied by the exposure time gives the expected number of deaths.[5,6]

The methodology to be described here is analogous to the traditional epidemiological task of comparing the observed survival experience of a study/target group to that of a standard reference group/population (eg, the general population).[7] Readers of this journal are familiar with the method in the context of mortality abstracts.[8] The difference in our application here is that the comparison is not *to* a standard reference group, but rather is to determine the accuracy of the ratings made *for* the reference group. That is, we are not assessing whether the observed deaths, O, are consistent with the known expected number, E. Instead, we are assessing whether the putative expected deaths, E, are consistent with those observed, O.

The comparison of interest will be that of the total observed number of deaths during the follow-up time period (O), with the expected number (E). The ratio of these, O/E, is often called a *standardized mortality ratio* (SMR).[5,p.97]

An SMR greater than 1 (or 100%) indicates more deaths than ''expected'', and thus that the raters are underestimating the actual mortality. Conversely an SMR less than 1 indicates fewer deaths than expected and an overestimation of mortality. SMRs can be computed for specific groups (eg, only males, or only heart patients) or specific follow-up times (eg, the first 2 years after rating). The key point in the evaluation of a rating system is that *a good system is one that produces an SMR of approximately 1.0 for the whole group and also for every relevant subgroup*.

Of course there are issues of sampling variation here. For example, in one small group it may be that E = 1.0 but O = zero. This alone hardly constitutes evidence that the rating system is inferior. It is clear that technical issues of standard error and statistical significance are to be considered.

In Sections 4 and 5 below, we note some superficially plausible but unsatisfactory methods for comparing the observed and expected numbers of deaths. Next, in Section 6, we demonstrate the correct method and give two simple examples of its use. In Section 7, we apply the correct method to a large data set of insureds. We conclude in Section 8 with some comments on how these ideas should be used in practice.

4. SOME UNSATISFACTORY APPROACHES

Although we have not yet discussed the computation of the expected number of deaths, E, we have indicated the correct approach to rating the raters: it is based on a systematic comparison of O versus E for the whole group and for relevant subgroups. We think it may be helpful, however, to note some incorrect approaches that may have some superficial appeal, in part because of their simplicity. We have frequently come across these wrong approaches in our consulting work.

a. ***Do the individuals generally die at a time close to their life expectancies?***

As a simple example, suppose that the rating service concludes that the life expectancy is 5 years for every insured. Is it reasonable to expect that the actual survival times t_i should be close to 5 years?

The answer is, of course, no. Just as in the general population, a wide range of survival times is to be expected. When

this very pattern is observed in practice, it is in no sense evidence that the ratings are incorrect. We mention this here because we have seen financial models predicated upon the insureds each living exactly to his life expectancy, with no allowance made for departures from this ''term certain'' approach, or sensitivity analyses to determine the possible ill effects. The example in section 2, where half the individuals died tomorrow and the other half live 40 more years, shows how this leads to an incorrect valuation.

b. ***Is the average of the actual survival times fairly close to the average of the life expectancies estimated by the raters?***

This test has more merit than (a) above, but there are two problems. Firstly, there is a major technical problem: many, and perhaps most, of the individuals will not have died at all during the study period, so it is not possible to compute the average time until death. Secondly, even if all the subjects have been followed until their deaths, and the average survival time matches that estimated by the raters, it is possible that the *pattern* of mortality over time is wrong. As an extreme example, consider again the situation of Section 2, where the test criterion has been met but the expected value of the portfolio is still incorrect.

c. ***Incorrect computation of E, the expected number of deaths***

We have come across the following mistaken argument more than once. Suppose again that each subject is followed for a time period t_i, which is the time of death or of censoring. Before we had any information on his survival, we would have estimated his chance of being alive at time t_i as $S_i(t_i)$. So the expected number of deaths that this person contributes is:

$$1 * \text{Probability [person dies before time } t_i] = 1 - S_i(t_i).$$

Hence the total expected number of deaths is $E = n - \sum S_i(t_i)$.

It is easy to see that this approach is wrong. Suppose, for example, that the follow-up period is long enough that all the subjects die before it elapses. In this case the observed number of deaths, O, will be equal to n, the total number of persons in the sample. But the above E will always be less than n, because each $S_i(t_i)$ – which is the probability that the i^{th} person is alive at time t_i – will be always be greater than 0. Thus the method will in this case always lead to the conclusion that mortality has been underestimated, even if the reverse happens to be true.

An even worse mistake would be to apply the above method using the time of death, t_i, as the end of the period of observation. According to this approach, if the person dies at his median survival time he would contribute 1 death to O but only ½ a death to E. One would thus conclude that mortality is being underestimated by 50% (SMR = O/E = 2).

5. A USEFUL THOUGH INCOMPLETE METHOD

The data we are given – survival times, together with an indicator variable d_i of whether the individual lived or died during the study period – are in a suitable format for the computation of a *Kaplan-Meier survival curve*.[9] This statistical method, which properly takes account of censoring, gives a composite survival curve, S(t), that takes the value 1.0 at time 0 and diminishes over time as a step-function. The function is known to be an unbiased estimator (in fact, the ''maximum likelihood estimator'') of the expected proportion of subjects who will be still alive at time t.

To be useful, this empirical curve must be compared (and plotted alongside) a suitable ''expected'' curve. By this we mean the expected proportion of persons alive at each

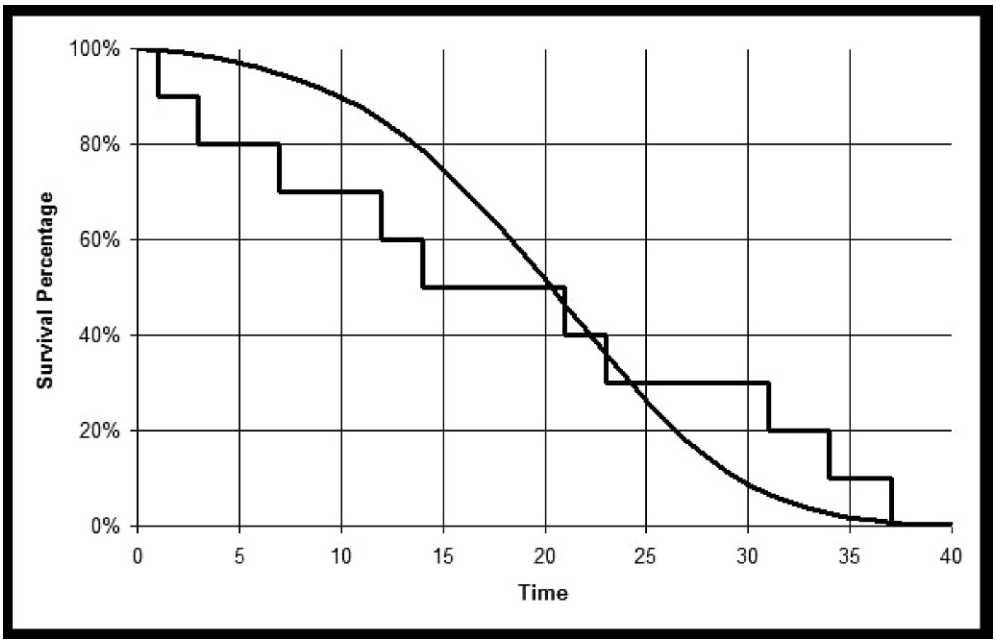

Figure. *Comparison of observed and expected survival curves.*

time on the assumption that the ratings are correct. The interpretation of this curve requires care because the n subjects are not a sample from a common distribution. Instead, each has his own survival function $S_i(t)$, specified by the rater. In general, the expected proportion alive at time t is correctly computed as:

$$S(t) = (1/n) \sum S_i(t).$$

Thus, for each time t, S(t) is simply the *average* of the n probabilities of survival.

The Figure above is an example. The smooth expected survival curve is that for 65-year-old non-smoking males who recently qualified for insurance at standard rates.[10] The observed Kaplan-Meier curve, which is in fact a step-function, is for a hypothetical group of 10 persons who died after survival times of 2, 4, 8, 13, 15, 22, 24, 32, 35, and 38 years. As can be seen, the observed survival percentage closely agrees with the expected values at times 0, 20, and 40 years. One might therefore think that the survival experience is ''about as expected'' and thus that the SMR is close to 1.0. But it is not. It can be shown that for this data E = 15 (see section 6) and thus the SMR = O/E = 10/15 = 0.67. Further, the plot reveals quite clearly that the observed survival was much worse than expected during the first 10 years, say, and much better in the later years. That is, the pattern of mortality is *not* as expected. It turns out that the 3 persons who lived the longest contributed 12.4 of the 15.0 expected deaths.

This method, which is recommended by Finkelstein et al.,[7] is certainly sound and can be useful in practice. There is even a simple statistical test for goodness-of-fit.[7] The limitation of the method is that it focuses solely on the *time since the rating was performed* as the variable of interest. Thus it would be sensitive to deficient ratings that, as here, seriously *under*estimate mortality in the early years and *over*estimate it in later years. But suppose that instead there was no time effect, and the ratings were, for example, too high for males and too low for females. The graph would not reflect this at all.

An equivalent method has been adopted in the life settlement industry by one of the major life expectancy providers.[11] They compare the cumulative actual (A) percentage of deaths with the percentage expected (E). This is, of course, merely the comple-

ment of the above method (ie, 1 minus the survival values). They then compute their ''A to E'' by taking the ratio of these two quantities. In statistical parlance, this is the ratio of the empirical distribution function, F(t), to the expected (or null) distribution function, $F_o(t)$, or $F(t)/F_o(t)$.[12] Many one-sample goodness-of-fit statistical hypothesis tests have been developed for this very application,[13,14] but the provider does not appear to have used them.

Under either of the above two methods, the actual and expected curves will agree perfectly at the outset (100% survival or 0% mortality, respectively) and after sufficient time has elapsed for all insured to die (0% survival or 100% mortality, respectively). In the interim, the ratio (or the curves) may vary, sometimes quite significantly if the sample size is relatively small.

Thus the above two methods provide no useful information at the two extremes of follow up time (ie, at time t = 0 and also when t is very large). Further, their use is somewhat problematic at intermediate values of t because the assessment of fit may vary over time. 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 CORRECT COMPUTATION OF EXPECTED VALUES (E)

As noted in Section 3 above, the correct method to rating the raters is based on comparison of the observed number of deaths, O, to the expected number E. These should be sufficiently close, both over the entire study group and also within subsets of interest. We now turn to the computation of the E's. As will be familiar to readers of the *Journal*, the correct units of analysis are *person-years of exposure time* (rather than, for example, the persons themselves).

We introduce this with a simple example. Suppose that an individual's mortality rate in any given year is 40%, and that he has been followed for exactly 4 years.[15] If he is still alive, then O, the observed number of deaths, is zero for him, of course. If he died at the end of the four years, then O = 1.

The *expected* number of deaths in each of these four years of exposure is 0.4, for a total of 1.6 deaths. It may seem odd that the expected number of deaths for one individual can exceed 1, but this is in fact perfectly reasonable. Suppose, for example, there are 100 persons like our individual, all of whom died after exactly 4 years. Then O = 100 and E = 160. This leads to the correct conclusion that the rating method has *underestimated* survival, or *overestimated* mortality, even though all the subjects have died. They have, after all, lived rather longer than expected: the average survival time can be shown to be only 2.5 years.[16]

For the general case, let the population size be n people. The total expected number of deaths is $E = \Sigma E_i$, where the summation is over i=1, 2, …, n, and E_i is the number of deaths expected over the follow-up period of t_i years for person i.

Note: One typically thinks of E as applying only to groups of persons. However, for each single person there is an expected number of deaths. Further, there is an expected number of deaths for each sub-interval of the person's exposure time.

For simplicity assume that t_i is an integer.[17] Then E_i is the sum of the first t_i mortality rates in the life table for person i.[18] That is, $E_i = \Sigma m_{ij}$, where m_{ij} is the estimated mortality rate for person i at follow-up time j, and the sum is over j=1, 2, …, t_i.

The mortality rates m_{ij} are shown in the life table for person i, or can be derived from survival or mortality probabilities.[19] Note that the E_i are not probabilities or rates of any kind. *The only interpretation of E_i is the expected number of deaths based on the total exposure time for the i^{th} individual.*

Example #1. This is given to illustrate the computation of the expected number of deaths for one person using the mathematical relationships described above.

- According to the U.S. life table, 2.3% of persons will survive from birth to age 100. That is, S(100) = 0.023.
- If a given individual is followed from birth and is still alive at age 100 (or dies at that age), the expected number of deaths during this period is the sum of all the mortality rates in the life table up to age 100. That is, $E = \Sigma m_j = -\ln[\exp(\Sigma m_j)] = -\ln[S(100)] = -\ln(0.023) = 3.8$, where ''ln'' denotes the natural logarithm and ''exp'' is its inverse, obtained by raising the transcendental number ''e'' to the given power. See Footnote #18 for the technical details.
- Thus even if the person dies at age 100, contributing 1 death to the observed total O, he contributes much more than that – ie, 3.8 deaths – to the expected total. This reflects that his survival is much better than average.
- This is perfectly reasonable; if we had a large group of persons, all of whom lived to age 100, we would correctly conclude that their collective survival is much better than predicted from the U.S. life table. In fact, if all died at age 100 we would have SMR = O/E = 1/3.8 = 0.26, much less than 1.0.

Example #2. This is given to illustrate the correct and incorrect methods on a small data set.

- Suppose a group of 4 identical elderly males were rated as having life expectancies of exactly 2 years each.
- For simplicity assume that the estimated mortality rate is constant. If so, it is 0.5 per year.[20]
- Suppose that the 4 die after survival times of 1, 1, 1, and 3 years.
- What is the SMR?
- *Correct answer*: O = 4, E = 1*0.5 + 1*0.5 + 1*0.5 + 3*0.5 = 3. So the SMR = O/E = 4/3 = 1.33, or 133%. There are 33% more deaths than were expected. Thus, the ratings *slightly* underestimate mortality and thus overestimate the survival rates and life expectancy.
- *Using the incorrect method in section 4(c) above*: Again, O = 4. The chance of dying in the first year is approximately 40%, and the chance of dying in 3 years is approximately 80%.[21] Thus, the incorrect estimate of E is 0.4 + 0.4 + 0.4 + 0.8 =2.0. The (incorrect) SMR is therefore O/E = 4/2.0 = 2. This incorrect SMR would lead us to the incorrect conclusion that the method of estimating life expectancy *greatly* overestimates survival.

As noted, in any application the observed number of deaths will be known. We have described how to compute correctly the expected number of deaths for each person. We emphasize that the observed and expected number are each attached, essentially, to a person-year, not to a person. The person-year contains all the current information about the person (age, sex, time since underwriting, medical risk factors, etc.).

These person-years can then be partitioned and compared (SMR = O/E) in many ways, according to the follow-ups times and person characteristics (which include age, sex, and all their lifestyle factors and medical conditions).

There will typically be hundreds or even thousands of individuals being assessed. A first step is to compare the *overall* observed and expected numbers of deaths as above. This may show that a given rating service systematically overestimates (SMR $<$ 1) or underestimates (SMR $>$ 1) the actual mortality.

This, however, is only a first step. Suppose that a given rating system gives an SMR of 1.0. This looks satisfactory. But it may be that there is overestimation of mortality for males and underestimation for females. Or the ratings are good during the first 3 years after ''underwriting'' and poor thereafter. Or it works well for healthy people and not for those with serious medical conditions.

Thus the data should be partitioned in many ways – eg, by age, sex, severity of health conditions, time since rating, source of the policy, type of underwriter or underwrit-

Table 2. Insured Information and Life Expectancy Estimates

	Covariates					
#	Age	Sex	Education	Medical Factors	MM	e
1	86	M	College	…	87	8.0
2	74	F	Grad School	…	129	13.3
3	63	F	College	…	164	18.0
4	74	M	College	…	103	14.4
5	80	M	High School	…	81	12.0

ing, face value, life expectancy (high/low), rating service, etc. *We then look for patterns of systematic deviation of the SMRs from the ideal value of 1.0*. This is partly an art, and partly involves statistical criteria such as significance tests and confidence intervals for the SMRs.[22] It is not our intention to give all the details here, but we believe that the examples that follow are sufficiently illustrative.

7. APPLICATION TO A PORTFOLIO OF POLICIES

We analyzed a portfolio of 4000+ persons who submitted their demographic, medical, and other information in anticipation of selling their life insurance policies. There were 286 deaths over the 5 years of available mortality follow-up.

Risk factor information was used to provide estimates of life expectancy. Table 2 shows partial information on the first 5 insureds. For brevity the ''Medical Factors'' are not shown here. MM is the mortality multiplier (MM=100 being ''standard insurance'') and ''e'' is the resulting life expectancy.

As noted, however, a life expectancy estimate alone is not sufficient to evaluate the ratings. Table 3 shows the estimated mortality rates for each insured over the 5 years of follow-up (although the life table provided mortality rates up to age 110, not all of these were necessary because the follow-up time was only 5 years).

To evaluate the above ''predictions'' we also required ''truth data'': the actual mortality experience of the group. Table 4 shows the survival time (ST) in years for each insured. The resulting exposure time in years, by time since underwriting, is given in the subsequent columns.

As can be seen, individual #1 survived 3 full years and then was censored (that is, he was still alive as of the date of the analysis). He contributed a full year of exposure time for years 1, 2, and 3, and nothing thereafter.

Table 3. Estimated Mortality Rates for Each Insured by Time Since Underwriting

		Time Since Underwriting (Years)				
#	e	0.0–0.9	1.0–1.9	2.0–2.9	3.0–3.9	4.0–4.9
1	8.0	0.03	0.04	0.06	0.08	0.10
2	13.3	0.01	0.02	0.02	0.02	0.03
3	18.0	0.10	0.12	0.12	0.12	0.14
4	14.4	0.01	0.01	0.02	0.02	0.03
5	12.0	0.01	0.02	0.02	0.03	0.03

Table 4. Exposure Time for Each Insured by Time Since Underwriting

#	Survival Time	0.0–0.9	1.0–1.9	2.0–2.9	3.0–3.9	4.0–4.9
1	3.0	1	1	1	0	0
2	3.5	1	1	1	0.5	0
3	0.4	0.4	0	0	0	0
4	3.9	1	1	1	0.9	0
5	5.0	1	1	1	1	1
Total		4.4	4.0	4.0	2.4	1.0

Table 5. Expected Deaths (E) for Each Insured (= Value in Table 3 * Value in Table 4)

#	Survival Time	0.0–0.9	1.0–1.9	2.0–2.9	3.0–3.9	4.0–4.9	Total
1	3.0	0.03	0.04	0.06			0.13
2	3.5	0.01	0.02	0.02	0.01		0.06
3	0.4	0.04					0.04
4	3.9	0.01	0.01	0.02	0.02		0.06
5	5.0	0.01	0.02	0.02	0.03	0.03	0.11
Total		0.10	0.09	0.12	0.06	0.03	0.40

Table 6. Observed Deaths (O) for Each Insured by Time Since Underwriting

#	Died	0.0–0.9	1.0–1.9	2.0–2.9	3.0–3.9	4.0–4.9	Total
1	0	0	0	0	0	0	0
2	1	0	0	0	1	0	1
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	1	0	0	0	0	1	1
Total		0	0	0	1	1	2

Person #3 was censored after only 0.4 years of follow-up, and thus contributed only 0.4 years in the first year.

The expected number of deaths (E) for each insured is simply the appropriate mortality rate in Table 3 multiplied by the exposure time in Table 4. These are shown in Table 5. The missing entries are in fact all zeros; these have been left blank to emphasize that there was no exposure during the given time period for the given person. Notice that there is an expected number of deaths for each person-year.

Lastly, we must use information on the timing of deaths. Table 6 presents this information. As with the expected number above, there is also an observed number of deaths for each person-year.

From the limited data given above, we see that there were 2 deaths, compared with the total expected number of 0.40. Hence the overall SMR = O/E = 2/0.40 = 5.0, 500%, or 5 times as many deaths as predicted.

We now present results for the entire group of 4379 insureds.

Table 7 provides preliminary results. As can be seen, overall there were 20% more deaths than expected (SMR = 1.2), but the excess was largely in the first year post underwriting (SMR = 1.5) and the second and third years post underwriting (SMR = 1.2). Thereafter, the SMR was approximately 100%.

Many analysts would either choose to end their analysis at this point, or be forced to do so by the limitations of their chosen methods. As noted above, however, there is no need to restrict the analysis to the single issue of time since underwriting (TSU). The data can usefully be partitioned in many different ways. Table 8 shows the results of 3 other ways of stratifying the data. The first is by risk group. The best quintile is the 20% of the group with the lowest mortality multipliers (equivalently, they had the highest age-adjusted life expectancies), the 2nd quintile is the next group, and so on. The second is by age (at time of rating) group, and the third is by sex.

Table 7. Preliminary Results

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

To aid in the interpretation of these results, we include the number of people (n), the average mortality multiplier in the group, and the observed and expected numbers of deaths overall. We do not show O and E separately for the SMRs by TSU.

As can be seen, the predictions were excellent for the first three risk quintiles – those with the lowest mortality multipliers (average MM of 68, 78 or 90) – the SMRs being either 1.0 or 1.1. This indicates that (a) the chosen baseline mortality rate was a good fit to the empirical data, and (b) the relatively mild adjustments were done correctly.

The predictions were unsatisfactory in the worst quintile, the SMR being 1.6. This, however, was largely restricted to the first 3 years post underwriting. Further, we found that ratings were best for the older ages (70+), and tended to be far too optimistic at the younger ages (O/E = 3.3 at age 60–64, for example, though this was based on only 10 actual deaths).

We also examined whether the mortality estimates were accurate for a given medical condition (eg, coronary artery disease) and

Table 8. Further Results

Risk Group	n	Average MM	O	E	Overall O/E	O/E by TSU		
						0	1–2	3+
Best Quintile	879	68	48	43	1.1	1.1	1.0	1.3
2nd Quintile	875	78	41	39	1.0	1.1	1.0	1.0
3rd Quintile	875	90	48	43	1.1	1.0	1.2	1.0
4th Quintile	875	107	54	46	1.2	1.4	1.1	1.1
Worst Quintile	875	165	95	58	1.6	2.6	1.7	0.8
Age								
60–64	154	127	7	2	3.3	3.7	3.1	3.6
65–69	512	116	21	11	2.0	1.7	2.2	1.7
70–74	1031	104	36	31	1.2	0.9	1.4	1.0
75–79	1268	97	80	53	1.5	1.7	1.5	1.3
80–84	1003	95	76	73	1.0	1.6	1.0	0.8
85–89	339	97	53	48	1.1	1.5	1.0	0.7
90+	71	86	11	14	0.8	0.9	0.7	1.1
Sex								
Male	3836	107	184	145	1.3	1.6	1.3	1.0
Female	543	93	102	85	1.2	1.3	1.2	1.1

the type or severity of the condition (eg, 2-vessel). We found no clear pattern, and thus do not show those results here. We did not have information on the specific source, type, or face value of the policy, and thus could not stratify the analyses by these factors. Lastly, and as expected, the incomplete method of section 5 yielded rather different results.[23]

8. DISCUSSION

The rigorous testing described here is not merely for evaluation *of* the raters. It can and should be used by the rating firms themselves to improve the quality of their ratings. That is, the company can assess whether its ratings are correct, either overall, by disease, or by other stratification. Further, assuming that the raters adhered to company policies and the rating manual, the company can then assess whether the ratings themselves are adequate.

Additionally, the service can individually evaluate its own raters/underwriters. Are they systematically too optimistic or pessimistic? Do they rate some types of cases better than others? Are they becoming more accurate with experience?

The firm can also assess whether its accuracy varies by the type of investor, source of the policy, type of policy, face value, or other factors. If specific subsets tend to die early or live too long, further questions will arise.

Several rating services have advertised the accuracy of their predictions. However, it appears that they have done their testing on the same data that they used to determine the ratings. In statistics this would be considered an inadequate test, as fitting and testing are two distinct activities. It is a basic principle that the training or calibration data set should be separate from the validation set. Otherwise one is merely describing the fit of the model to the extant data – which can always be made better by post-hoc adjustment – rather than independently testing the methods or models.

Ideally, an independent actuarial firm would have a rating contest, contestants being invited to rate a portfolio of, say, 1000 persons whose actual mortality experience was known. Actuaries would then

evaluate the accuracy of the ratings and publicize their findings. One would be surprised, however, if many rating services would agree to such a test. But investors can already perform this test themselves. Investors typically have multiple ratings (''life expectancy predictions'') on each insured. And they have mortality information on each insured. They are thus in a position to evaluate the various rating services using the same set of insureds. Questions that can be answered include: What are the overall SMRs for the various service providers, and are they acceptably close to 1.0? Does one service do better on one type of insured (eg, older males) or policy (eg, large face value)? Is a particular company getting better with time? Are the newer companies more or less accurate than the older ones? Certainly this is a more principled and accurate approach to evaluation than merely using the average of the respective life expectancy opinions, as some investors tend to do.

FOOTNOTES

1. 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.
2. Anderson TW (2002). Life expectancy in court: A textbook for doctors and lawyers. Vancouver BC: Teviot Press.
3. The Life Table. Available at http://www.LifeExpectancy.org/LifeTable.shtml. Accessed July 8, 2009.
4. If a complete life table has not been provided, one can be constructed in several ways, such as (a) rating up [finding the age in the general population that has the same life expectancy as predicted, and using the subsequent mortality pattern for that age]; (b) adding a constant excess death rate to some baseline mortality rate to give the same life expectancy; (c) multiplying the baseline rates by some constant relative risk [called the mortality multiplier (MM)] to give the correct life expectancy; or (d) some other pattern of mortality, such as linearly declining log relative risk or proportional life expectancy.
5. Kahn HA, Sempos CT. (1989). Statistical Methods in Epidemiology. Oxford: Oxford University Press, page 207.
6. Breslow NE, Day NE (1980). Statistical methods in cancer research: Volume 1 – The analysis of case-control studies. Lyon, France: IARC Scientific Publications No. 32, page 45.
7. 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.
8. Using *Journal of Insurance Medicine* notation, the expected number of deaths (d') is computed as the product of the exposure patient-years (E) and the expected mortality rate (q'). Here, however, we use E to represent the expected number of deaths, and have used standard biostatistics terminology by denoting the mortality rate by m and the mortality probability by q. Because similar terms and symbols are used (e.g., O/E, SMR), a life actuary may believe that none of this theory is new. This may be so, but we were unable to identify any published literature or textbooks on the topic. More importantly, we have not seen a comprehensive and correct application of the method.
9. Collett D (1994). Modelling survival data in medical research. London: Chapman and Hall. page 19.
10. Society of Actuaries Mortality Task Force (2001). 2001 Valuation basic mortality table (VBT). Schaumburg, Illinois: Society of Actuaries.
11. In actuality they count the observed number of deaths, and compare to the expected number, both computed as a percentile. In practice, this means that they must compute the separate percentile times for each prediction for each person, then see if the person was observed long enough so that a death could be observed (e.g., if they are still alive today, 3 years later, and 3 years is their 50^{th} percentile, then they obviously cannot contribute anything to the 60^{th} percentile on the graph). This seems far too complicated. While plotting against percentile may be interesting in theory, it eliminates the obvious time scale – years – where people are both familiar with and interested in cash flow.
12. The ratio of the survival in a target population to that of a reference population is commonly referred to as the relative survival rate, and can be presented in interval or cumulative fashion. The concept dates to Greenwood (1926). See, for example, page 93 of reference #13 below, or the treatise by Fred Ederer (1961), Cancer Institute Monographs.
13. Lee ET (1992). Statistical methods for survival data analysis. New York: John Wiley and Sons, pages 182–191.

14. Rohatgi VK (1976). An introduction to probability theory and mathematical statistics. New York: John Wiley and Sons, page 539.
15. We use the term ''rate'' in the strict technical sense. In the life table, the mortality rate at age x is denoted by ''m(x)'', or simply m. It is the number of events (deaths) per unit time, and thus takes values 0 and larger. This is distinct from the mortality probability at age x, ''q(x)'', or simply q, which is constrained to lie between 0 and 1 inclusive. Note that in some insurance literature the authors often derive the mortality probability q, but mistakenly refer to it as a rate.
16. The average survival time, ie the life expectancy, is $e=1/m=1/0.4=2.5$ years. Note that if all persons lived exactly to their life expectancy, then it can be shown that (for cases with increasing mortality risk with age) the expected number would be less than n.
17. We have considered the simple case where all survival times are integers. For non-integer values (eg, 2.4 years), the expected number of deaths in the final year is merely the fractional value (eg, 0.4*m).
18. The mathematically-minded reader will recognize that E_i is the *cumulative hazard function* (that is, the area under the hazard rate function up to a given time), normally denoted $H(t_i)$. As the individual's survival time T is a random variable, H(T), the cumulative hazard function at time T, is also a random variable; it is the expected number of deaths at the time that our individual actually died. That $H(t) = -\ln[S(t)] = \Sigma m_j$ is a well-known result in survival analysis. See reference #13, page 13. Note also that $S(t) = \exp[\Sigma m_j]$. One might expect that, on average, there will be exactly one death per person, and this is correct. That is, the mathematical expectation of H(T) is equal to one. Proof: $E[H(T)] = \int H(t)f(t)dt = -\int \ln[S(t)]f(t)dt = -\int \ln[1-F(t)]f(t)dt$. Now use substitution with $u = 1-F(t)$ to obtain $E[H(T)] = -\int \ln(u)du = [u-u*\ln(u)]|$, where the integral and subsequent evaluation is from 0 to 1. Taking the limit as $u \rightarrow 0$, and using L'Hopital's Rule, yields the desired result.
19. Using the notation in note #15 above, $m = -\ln(1-q)$. If $q=0$, then $m=0$. As $q \rightarrow 1$, $m \rightarrow$ infinity.
20. This is based on the exponential (''memoryless'') distribution, wherein $e=1/m$ or $m=1/e$. See, for example: Ross SM (1989). Introduction to probability models, 4th Edition. San Diego, CA: Academic Press, page 33. This fact was used in Footnote 16.
21. Using note #19 above, $q = 1-\exp(-m) = 1-\exp(-0.5) = 0.39$. Over 3 years, the survival probability is $(1-0.39)^3 = 0.23$, so the mortality probability is 0.77.
22. For confidence intervals, see, for example, (a) pages 98–100 of reference #5 above, (b) Singer RB (1992). The Application of Life Table Methodology to Risk Appraisal. In: RDC Brackenridge, RS Croxson, & R Mackenzie (Eds.), Medical Selection of Life Risks, Fifth Edition, pages 45–70. New York: Palgrave, or (c) Breslow NE, Day NE (1987) Statistical methods in cancer research, volume II. Lyon, France: IARC (reprinted by Oxford University Press, New York), pages 69–71.
23. The incomplete method of section 5 gives an overall ''A to E'' of 110%, which is rather smaller than the true value, 120% (see Table 7). Note that we write ''A to E'' in quotes to distinguish this measure from the correct one. The deficiency is even more pronounced in the worst quintile, with actual 5-year survival of 81% (19% mortality) compared with 84% expected (16% mortality), which gives an ''A to E'' of $19/16 = 1.2$. But the correct A to E is $136/83 = 1.6$, which is evidently quite different. The problem with the incomplete method is best illustrated by extending the above findings for 10 years. If hypothetically the 10-year actual mortality curve ''flattened out'' or otherwise met with the expected 10-year mortality curve, one would have an ''overall'' ''A to E'' at that point of 100%, or perfect agreement. But this would wholly misrepresent the evidence.