Life Reinsurance

“The strength of expertise”

Assessment of Milliman’s Risk Score 3.0 as a Predictor of Life Insurance Mortality

Winter 2024

Introduction

Optimum Re has reviewed Milliman's Irix Risk Score, which is a proprietary algorithm developed by Milliman using an individual's prescription drug history, medical data and credit data to calculate a Risk Score as a predictor of future mortality. Such algorithm may be used by insurance companies to replace or complement traditional underwriting requirements. Optimum Re may assist insurance companies who wish to validate the predictability of Risk Score, with respect to mortality, on their own portfolio by performing a retrospective analysis and by guiding them in setting their mortality assumption. To have a summary of the highlights of this review, please refer to the **Conclusion** section.

In order to perform the assessment of the Risk Score, Milliman provided Optimum Re with a dataset having the following characteristics:

Table 1 – Data summary[1]	
Number of applicants	42M
Issue Years	2005 to 2020
Exposure Years	2005 to 2021
Units of Exposure	236M
# of claims	1.7M

Different versions of the Risk Score are provided in the dataset. For the purpose of this paper, we analyse the following:

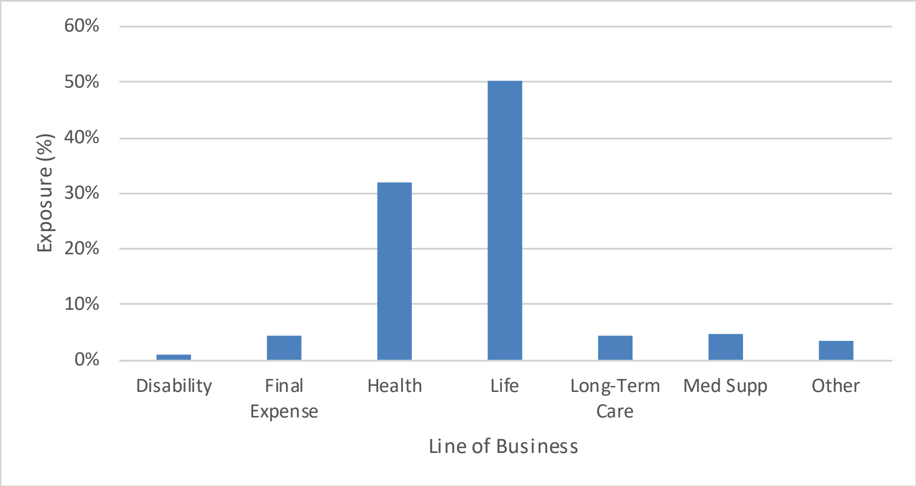
1. **2.2 Rx:** previous version of the Risk Score, using prescription drug history only
2. **3.0 Rx:** newest version of the Risk Score, using prescription drug history only
3. **3.0 Rx & Dx:** same as 3.0 Rx, with added medical data
4. **3.0 Rx & Cr:** same as 3.0 Rx, with added credit data
5. **3.0 Rx & Dx & Cr:** same as 3.0 Rx & Dx, with added credit data

Let's define an **Rx Hit**, **Dx Hit** and **Cr Hit** as the success in obtaining prescription drugs, medical and credit data, following a request to the database. For the purpose of this report, we do not consider an Rx Eligibility-Only hit, where the applicant was found in Milliman's database but no prescription drugs information was available, as an Rx Hit because it does not provide a Risk Score value.

A Risk Score will be available if at least one of its component has a Hit. Hence different versions of the Risk Score might be equal if they are based on the same underlying data. For example, an Rx Hit with no Dx nor Cr Hit will result in identical **Rx, Rx & Dx and Rx & Dx & Cr** Scores because the 3 scores are based on prescription drugs data only. An exception to this rule happens when the credit information is not ordered (for issue age below 18), where the **Rx & Cr** and **Rx & Dx & Cr** Scores are set at 0, regardless of the availability of the Rx or Dx data.

The dataset includes observations for different lines of business, with the majority labeled as "Life" (≈50%) and "Health" (≈30%) as seen in **Figure 1**. The Life data may include life insurance products with different levels of underwriting (Fully Underwritten, Simplified Issue, etc.) but excludes Final Expense products, which have their own label.

Figure 1 – Distribution of Exposure by Line of Business

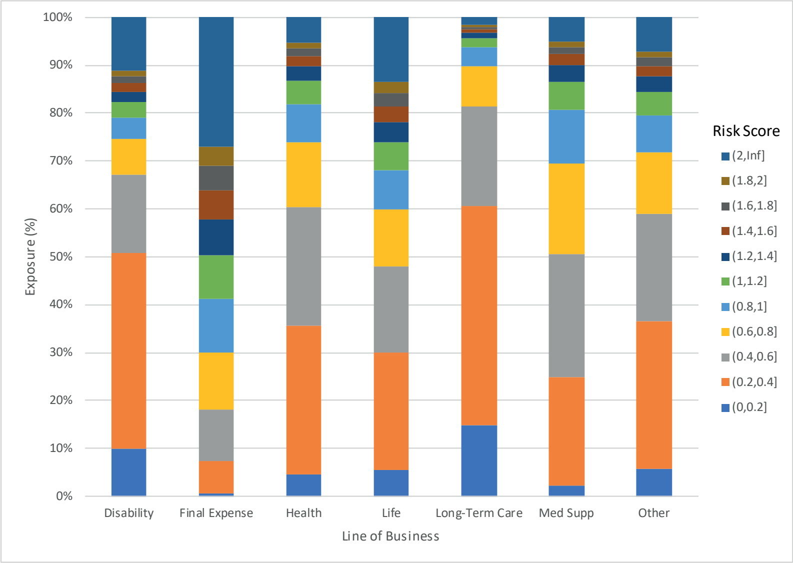


1 - For our assessment, we exclude applicants with Issue Age greater than 90 for credibility purposes.

Introduction (cont'd)

We can see in **Figure 2** that the distribution of Risk Score varies greatly from one Line of Business to another, with Final Expense being more skewed towards high scores than the rest. Given this, an analysis by line of business may be a wise choice to draw sound conclusions. Therefore, the focus of this report will be the Final Expense and Life lines of business with a deeper focus on the former, where Optimum Re's holds a high level of expertise. While the Final Expense portion of the data is more limited than the Life portion in terms of volume, it still includes around 10M units of exposure and 280K claims, which is plenty to draw credible conclusions.

Figure 2 – Distribution of Risk Score 3.0 Rx & Dx & Cr by Line of Business

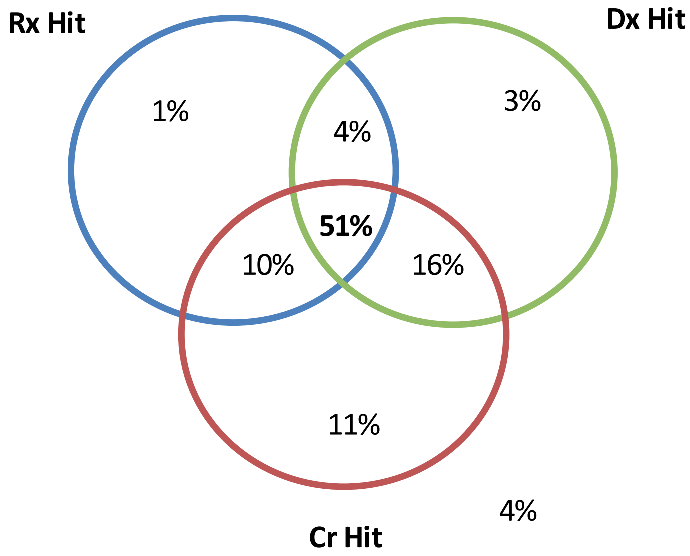


Results

1. Risk Score Hit Rate

Figure 3 presents the proportion of Final Expense Exposure in the dataset for which there is an **Rx, Dx** or **Cr Hit** in the form of a Venn diagram.

Figure 3 – Venn Diagram of Final Expense Exposure % of different Hit categories



A few notes:

66% of the exposure has an Rx Hit represented by the **blue circle**	74% of the exposure has a Dx Hit represented by the **green circle**	88% of the exposure has a Cr Hit represented by the **red circle**

In terms of scores:

1. **66%** of the exposure has a **2.2 Rx and 3.0 Rx** score:
   Rx Hit
2. **85%** of the exposure has a **3.0 Rx & Dx** score:
   Rx **or** Dx Hit
3. **93%** of the exposure has a **3.0 Rx & Cr** score:
   Rx **or** Cr Hit
4. **96%** of the exposure has a **3.0 Rx & Dx & Cr** score:
   Rx **or** Dx **or** Cr Hit

Hence if one were to wonder which version of Risk Score to use, for the purpose of maximizing hit rate, the **3.0 Rx & Dx & Cr** score comes out on top, followed by the **3.0 Rx & Cr** score and finally the **3.0 Rx & Dx** score.

Similar conclusions can be drawn for the Life line of business.

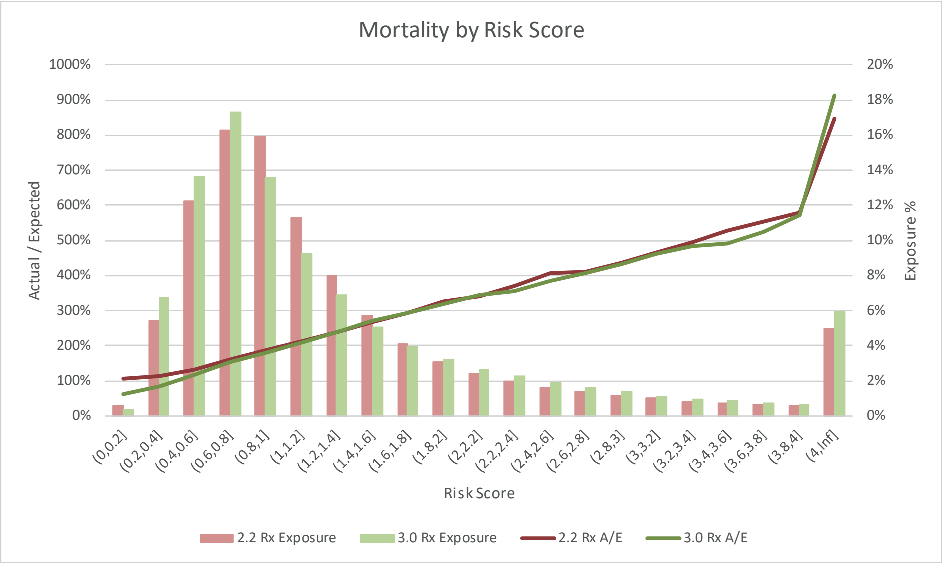
Results

2. Risk Score as predictor of mortality

In order to assess the relationship between Risk Score and mortality, we calculate actual to expected (A/E) ratios. The A/E ratio is the ratio of **actual observed deaths** to **expected deaths** based on the 2015 VBT Tables for Life, and the 2008 VBT LU for Final Expense[3]. By comparing the A/E ratios by different values of Risk Score, we can assess if Risk Scores are good predictors of mortality.

The link between mortality and Risk Score is clearly visible for the Final Expense business in **Figure 4**, where we can see an increase in A/E (represented by the curves) as the Risk Score increases.

Figure 4 – Actual / Expected by Risk Score – Final Expense



Furthermore, we see that **Risk Score 3.0 Rx** seems to be better than **Risk Score 2.2 Rx** at identifying applicants with the best and the worst mortality. Indeed, the new version of the Score classifies a larger proportion of applicants as relatively low risk (bars less than 0.8) with a lower A/E. We see a similar pattern at the other end of the risk spectrum (bars larger than 4).

For the Life line of business, the new version of the Score is better at identifying the best risks, and similar at identifying the worst risks.

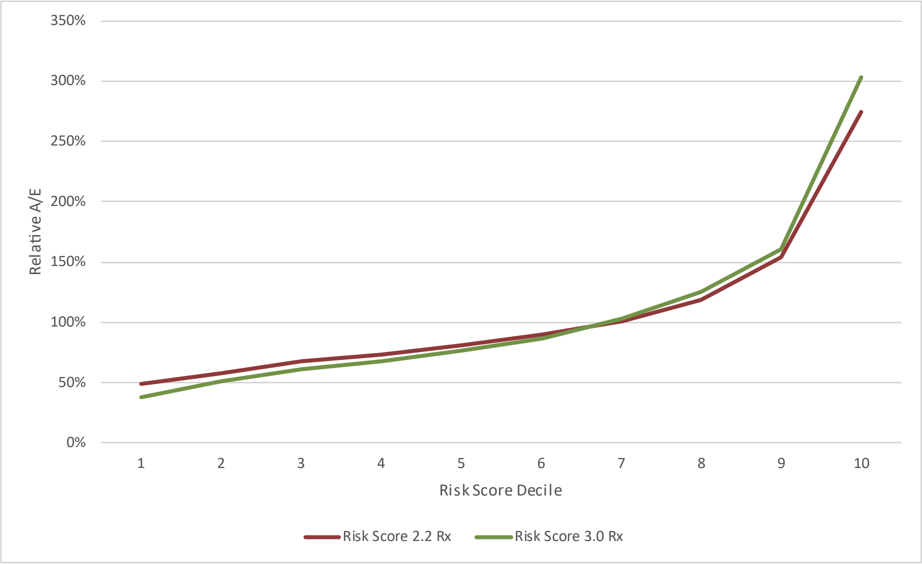
An alternative way to compare the power of each version of Risk Score as a tool for mortality segmentation is to group the scores into deciles from lowest to highest (10% of applicants in each decile) and then compare the A/E of each decile between scores. To account for the fact that each score is available for a different population, for a given score, we divide the A/E of each decile by the total A/E for that score and define it as the Relative A/E. This allows comparing each score on the same basis.

3 - Since the smoking status is not provided in Milliman's dataset, we use a blend of 85% Non-Smoker and 15% Smoker to obtain the 2015 VBT and 2008 VBT LU mortality rates. Additionally, the rates are adjusted for Mortality Improvement at a rate of 1% per year.

Results

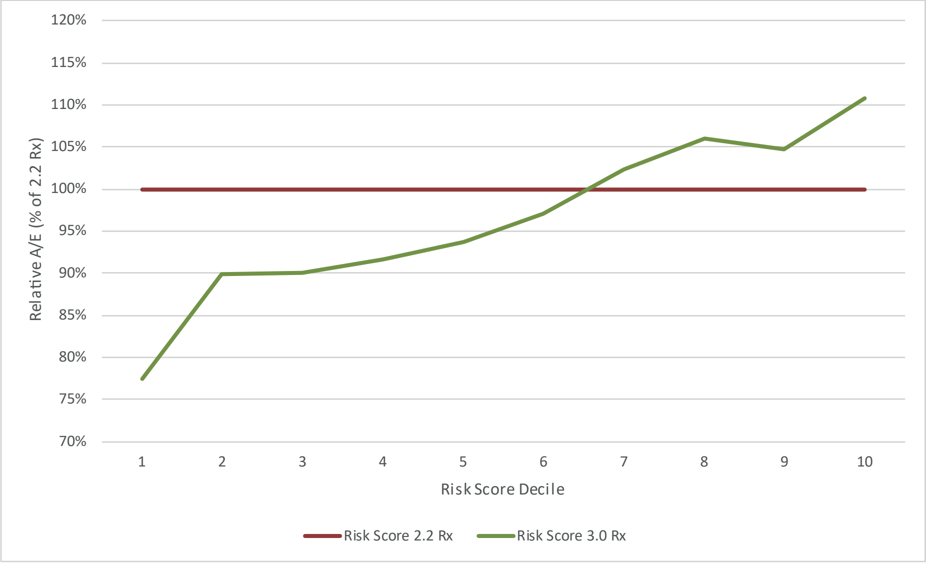
Figure 5 shows the Relative A/Es between **Risk Score 2.2 Rx** and **Risk Score 3.0 Rx**. When comparing the new version of the score to the previous one, the lower A/Es for deciles 1 to 5 and the higher A/Es for deciles 7 to 10 reinforce our previous conclusion that the new version of the Risk Score is better at segmenting mortality risk.

Figure 5 – Relative A/E by Risk Score – Final Expense



For a better visualization of the differences, we express the relative A/E as a percentage of the **Risk Score 2.2** Relative A/E in **Figure 6**.

Figure 6 – Relative A/E by Risk Score (% of 2.2 Rx) – Final Expense



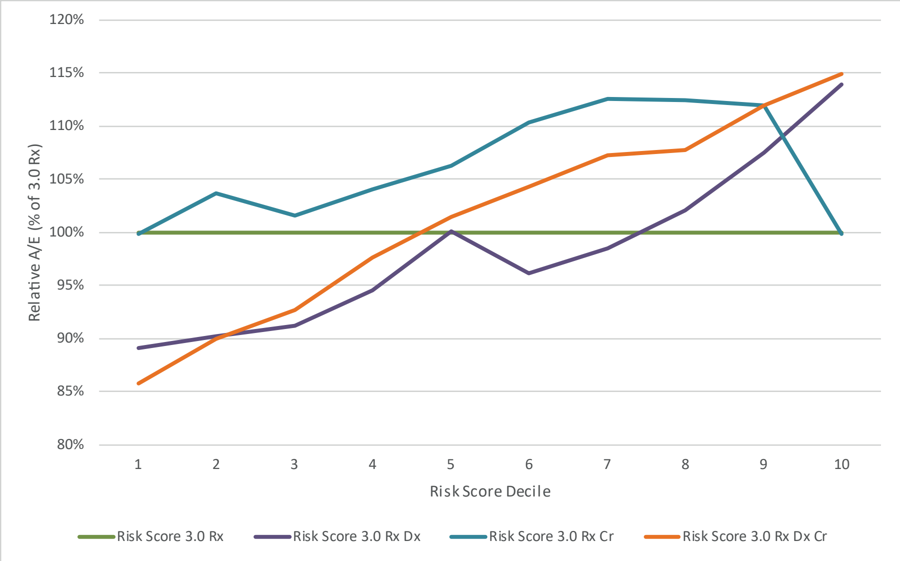
In this figure, we can clearly see that the Relative A/E for **Risk Score 3.0 Rx**, for decile 1 and decile 10 are respectively 78% and 111% of the Relative A/E for **Risk Score 2.2 Rx**.

Results

Using this approach and expressing the Relative A/E of each Risk Score as a percentage of the A/E of **Risk Score 3.0 Rx** we can compare the different versions of Risk Score 3.0 in **Figure 7**.

To consider a version of Risk Score 3.0 to be an improvement over the **Risk Score 3.0 Rx** baseline in terms of mortality segmentation, we would like to see it score applicants such that the lowest scores (low deciles) have lower A/Es and the highest scores (high deciles) have higher A/Es when compared to the baseline.

Figure 7 – A/E by deciles for different Risk Scores – Final Expense



Risk Score 3.0 Rx Cr performs generally better than the baseline at identifying the worst risks but generally worse at identifying the best risks and is therefore not a clear improvement over the baseline.

Risk Scores 3.0 Rx & Dx and **3.0 Rx & Dx & Cr** are clear improvements over the **Risk Score 3.0 Rx**. However, it is harder to assess the value the credit information brings since **Risk Score 3.0 Rx & Dx & Cr** seems to be generally better at identifying the worst risks but generally worse at identifying the best risks (except in decile 1).

Given that the use of credit information may come with higher costs, we find it pertinent to take a deeper dive into trying to assess if there are mortality segmentation gains coming from a transition from **Risk Scores 3.0 Rx & Dx** to **Risk Score 3.0 Rx & Dx & Cr** for the Final Expense line of business. To do this, we compare the two Risk Scores on the same population by selecting applicants for which **Risk Scores 3.0 Rx & Dx** and **3.0 Rx & Dx & Cr** are available and compare how each applicant would be grouped by each score into quintiles from lowest to highest (20% of applicants in each quintile).

Table 2 shows how the exposure from each quintile based on **Risk Score 3.0 Rx & Dx** (rows) would be grouped under **Risk Score 3.0 Rx & Dx & Cr** (columns).

Table 2 – Proportion of Exposure - 3.0 Rx & Dx vs 3.0 Rx & Dx & Cr

	3.0 Rx & Dx & Cr Quintiles				
3.0 Rx & Dx Quintiles	1	2	3	4	5
1	71%	**22%**	6%	1%	0%
2	**21%**	46%	**23%**	9%	1%
3	7%	**26%**	44%	**21%**	3%
4	1%	6%	**25%**	54%	**13%**
5	0%	0%	2%	**15%**	82%

We can see that a significant proportion of applicants would be grouped 1 quintile away under the two scores (in blue). For example, for quintile 2 under **Risk Score 3.0 Rx & Dx**, 21% and 23% would be classified as quintile 1 and 3 respectively under **Risk Score 3.0 Rx & Dx & Cr**.

Results

To assess the impact of these differences in classifications, we calculate the A/E for each cell as displayed in **Table 3**. Note that, for credibility purposes, we omit the A/E for cells in which there are too few claims.

Table 3 – A/E - 3.0 Rx & Dx vs 3.0 Rx & Dx & Cr						
	3.0 Rx & Dx & Cr Quintiles					
3.0 Rx & Dx Quintiles	1	2	3	4	5	**Total**
1	86%	126%	192%			**100%**
2	106%	135%	177%	290%		**144%**
3	136%	155%	188%	272%	487%	**192%**
4		197%	220%	279%	441%	**274%**
5			355%	354%	641%	**587%**
Total	**94%**	**141%**	**195%**	**288%**	**613%**	**239%**

For example, applicants that were classified in the 1st quintile by **Risk Score 3.0 Rx & Dx** (with copper borders) have an A/E of 100%. Some of these applicants were better classified by **Risk Score 3.0 Rx & Dx & Cr** in the 1st, 2nd and 3rd quintile, with respective A/Es of 86%, 126% and 192%.

However, the same observation can be made when looking at applicants that were classified in the 1st quintile by **Risk Score 3.0 Rx & Dx & Cr** (with thicker blue borders), where we can see improvements in classification under **Risk Score 3.0 Rx & Dx**.

More generally, while we see that some applicants are better classified by one score or the other, we can observe that, for a given quintile, the slope of the A/Es seem to be larger by rows than by columns, indicating larger segmentation gains under **Risk Score 3.0 Rx & Dx & Cr**. Moreover, we see that some of the applicants classified in the first 3 quintiles by **Risk Score 3.0 Rx & Dx** (in copper) are not following an appropriate segmentation pattern, with A/Es larger than those of the next quintile from the same column (ex: in the quintile 3 column, 192% for quintile 1 vs 177% for quintile 2).

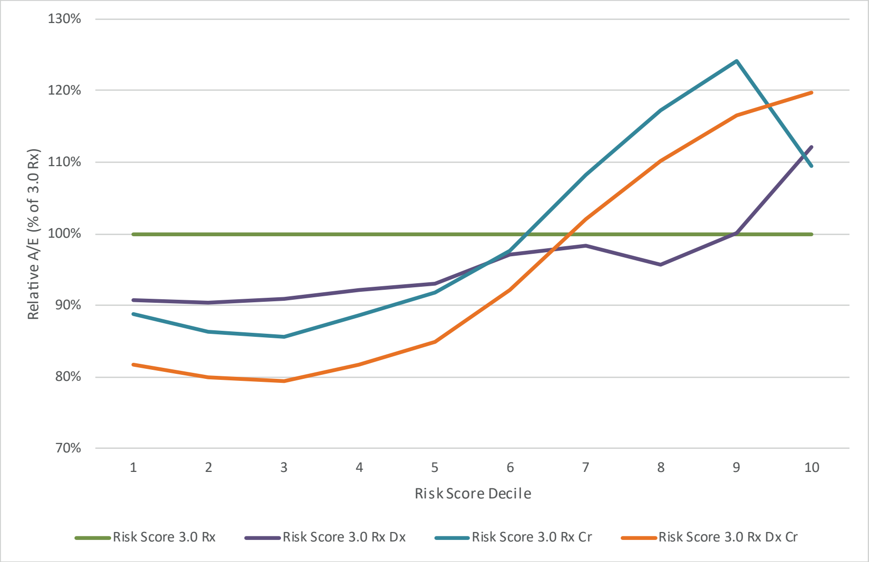
To summarise, for the Final Expense line of business:

- **Risk Scores 3.0 Rx & Cr** does not seem to improve much over **Risk Scores 3.0 Rx**
- **Risk Scores 3.0 Rx & Dx** and **3.0 Rx & Dx & Cr** are clear improvements over **Risk Scores 3.0 Rx**
    - Given the use of medical data by these two scores, one may conclude that medical data is a strong predictor of mortality in the Final Expense line of business.
    - **Risk Score 3.0 Rx & Dx & Cr** seems to have a small edge in terms of mortality segmentation power over its credit-less counterpart. However, the use of credit information may come with higher costs that must also be weighted in the decision process. Optimum Re is available to assist insurance companies in taking a decision that best fits their needs.

Results

Figure 8 shows the equivalent of **Figure 7** but for the Life business.

Figure 8 – A/E by deciles for different Risk Scores – Life



From **Figure 8**, we can draw the following conclusions for the Life line of business:

- The value of credit information is much clearer for the Life line of business when compared to the Final Expense line of business with **Risk Score 3.0 Rx & Cr** and **Risk Score 3.0 Rx & Dx & Cr** showing the largest improvements over the baseline.
  - The addition of medical data to the score seems to allow for a smoother segmentation or risks, especially at the two ends of the spectrum, making **Score 3.0 Rx & Dx & Cr** seem like the best overall choice.
- While **Risk Score 3.0 Rx & Dx** seems to generally improve over the baseline, it has a lower mortality segmentation power than the two other scores that include credit information.

3. Generalization of performance

We previously concluded that the Risk Score is a good predictor of mortality but one might wonder if this would also be the case when tested on data that was not included in the Risk Score's creation. Without train/validation/test labels in the Milliman data, we cannot truly assess the performance on new data but one approach we can use is to assess the predictive power of the Risk Score on a random subset of the data. Indeed, a score that is built to perform very well on a given dataset by overfitting could see a worse performance on a subset of this data.

Figure 9 shows the comparison of the Final Expense Relative A/E by **Risk Score 3.0 Rx & Dx & Cr** deciles on the whole data and on five random subsets of 10% of the data while **Figure 10** shows the same A/E as a percentage of the A/E on the whole data to highlight the differences.

Figure 9 – Risk Score 3.0 Rx & Dx & Cr A/E, Whole dataset vs five 10% random subsets – Final Expense

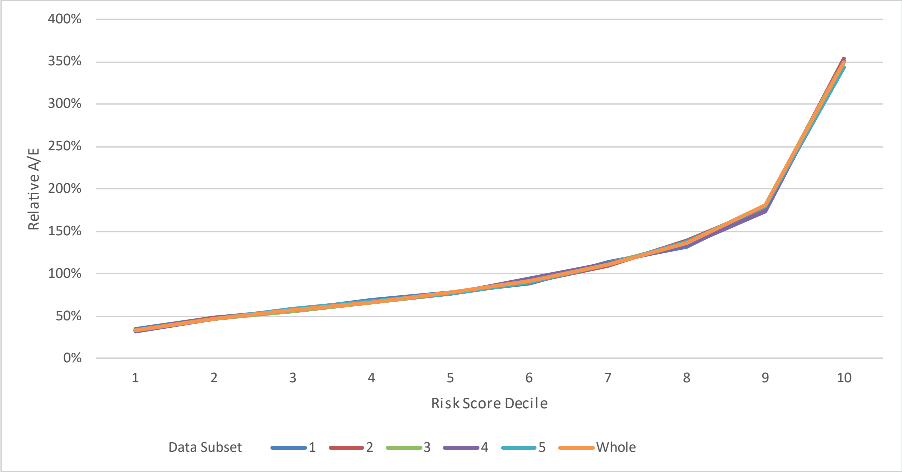
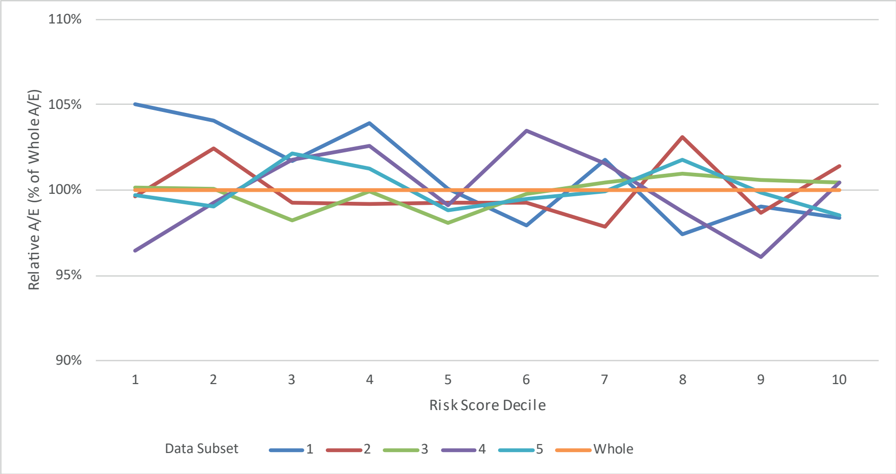


Figure 10 – Risk Score 3.0 Rx & Dx & Cr A/E, Whole dataset vs five 10% random subsets as % of Whole A/E – Final Expense



From **Figures 9** and **10**, we see that the Risk Score is a good predictor of mortality on the 5 subsets of the data, with minimal deviations from the results observed on the Whole dataset (lesser or equal than 5%). While this does not prove without a doubt that the Risk Score would perform well on new unseen data, it is a step in the right direction.

Similar conclusions can be drawn for the Life line of business.

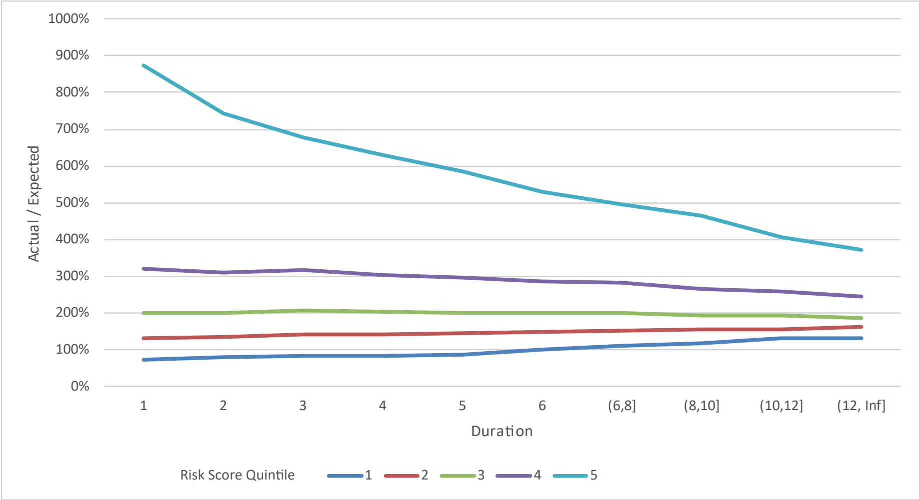
Results

4. Risk Score Duration Wear-Off

The distribution of Risk Score does not vary much by Duration. This makes sense given that the Risk Score is established at issue and does not vary across durations. Hence, the only change in Risk Score distribution by duration is caused by the fact that applicants with higher scores will die sooner than applicants with lower scores, so the proportion of high scores slightly decreases by Duration.

However, the mortality segmentation power of the Risk Score does vary by duration. **Figure 11** shows the A/E by Duration where each curve is a quintile of Final Expense applicants grouped by Risk Score from lowest to highest (20% of applicants in each quintile).

Figure 11 – A/E by Duration for Risk Score3.0 Rx & Dx & Cr quintiles – Final Expense



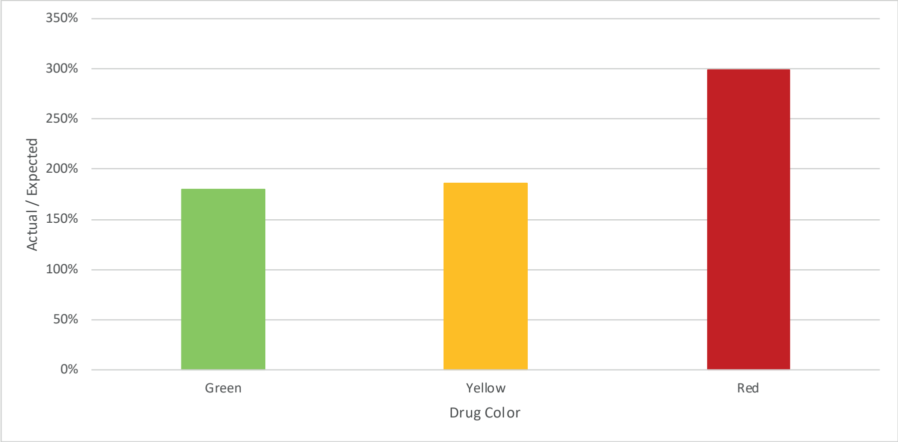
We can observe larger differences in A/E between Risk Score quintiles at lower Durations for the Final Expense business which seem to indicate that the predictability of the Risk Score wears-off over time. However, even at the latest durations, these is still a selection effect.

Similar conclusions can be drawn for the Life line of business, with a steeper wear-off effect.

5. Risk Score vs Rx Check

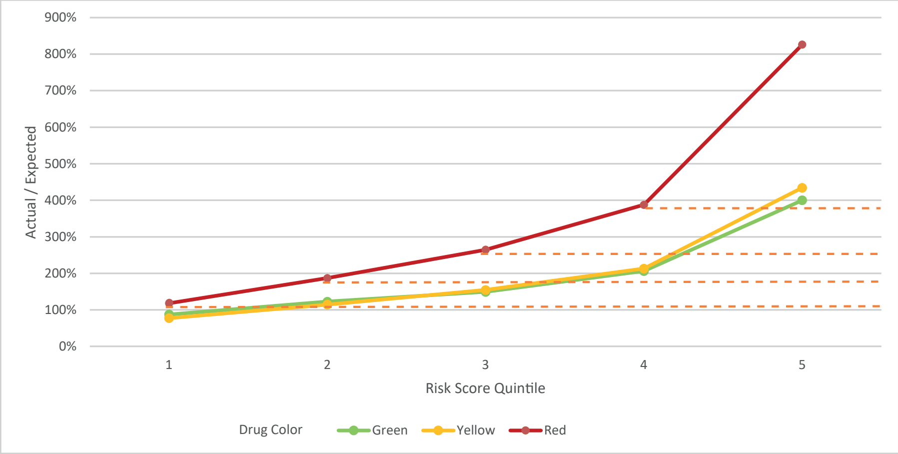
Insurers may wonder about the value of using a Risk Score compared to simply using Rx Check during the underwriting process. With Rx Check, the insurer receives a list of drugs taken by applicants with a color assignation corresponding to mortality risk. Green corresponds to a drug which is expected to yield a low mortality risk, Yellow a moderate risk and Red a high risk. In the dataset, we only have the color of the drug related to the most severe illness, which we use for the analysis. In **Figure 12**, we can see that the average A/E of Red drug applicants is much larger than the average A/E for Green or Yellow applicants.

Figure 12 – A/E by drug color – Final Expense



With the drugs colors information, underwriters may further investigate to refine the risk assessment of each applicant. We can demonstrate that some of that refinement can be readily available with the use of the Risk Score. For example, the Risk Score may take into account the interaction between different drugs as well as the timeline of the drugs intake. In order to illustrate the segmentation provided by the Risk Score within a drug color group, we further group the **Risk Score 3.0 Rx & Dx & Cr** into quintiles from lowest to highest (20% of applicants in each quintile) and then compare the A/Es in **Figure 13**.

Figure 13 – A/E by Risk Score quintile for different drug colors – Final Expense



As expected, we see an increase in A/E with an increase in Risk Score and the Red drug group has higher A/Es for a given Risk Score quintile. We also see large variations in A/E between the best and worst quintile of risks within a drug group (ex. best red at 118% vs worst red at 826%). An even more interesting observation is that the Risk Score allows identifying some applicants in the Red drug group with lower risk than applicants in the Green and Yellow groups. For example, the best 20% of the Red group has a lower A/E than the worst 60% of the Green and Yellow groups.

Thus, it is evident that the Risk Score significantly outperforms the Rx Check in the Final Expense line of business.
By offering enhanced segmentation of mortality risk, it aids in making well-informed decisions concerning risk assessment.

A similar conclusion can be reached for the Life business.

6. Application of the Risk Score – Thresholds and Pass Rates

In practice, users of the Risk Score will use Risk Score Thresholds to take underwriting decisions or classify applicants into various risk classes or product types. For example, applicants with a score lower than 2 may pass through the underwriting process without additional requirements for issue. A lower threshold leads to a lower pass rate and lower expected mortality and the opposite is true. Hence, users will need to select thresholds that are in line with mortality risk appetite and pass rate goals.

For example, let's assume a user wants to apply an underwriting decision or process to the best 80% of applicants. **Table 4** shows the Risk Score Threshold needed to achieve that goal, along with the resulting Relative A/E of the selected group for different versions of Risk Score 3.0 for the Final Expense line of business.

Table 4 – Threshold and Relative A/E by Risk Score, best 80% – Final Expense		
Risk Score	**Threshold (Risk Score <= x)**	**Relative A/E**
3.0 Rx	2.00	78%
3.0 Rx & Dx	2.26	75%
3.0 Rx & Cr	2.36	81%
3.0 Rx & Dx & Cr	2.44	76%

First, we can see that if a user of a Risk Score goes from using **3.0 Rx** to **3.0 Rx & Dx & Cr**, he could increase the Threshold to obtain the same 80% pass rate, while seeing an improvement in A/E.

Table 5 instead shows what would happen if a user were to make the same Risk Score transition but keep an existing Threshold of 2 in place.

Table 5 – Pass Rate and Relative A/E by Risk Score, Risk Score <= 2 – Final Expense		
Risk Score	**Pass Rate**	**Relative A/E**
3.0 Rx	80%	77%
3.0 Rx & Dx	76%	71%
3.0 Rx & Cr	73%	77%
3.0 Rx & Dx & Cr	73%	71%

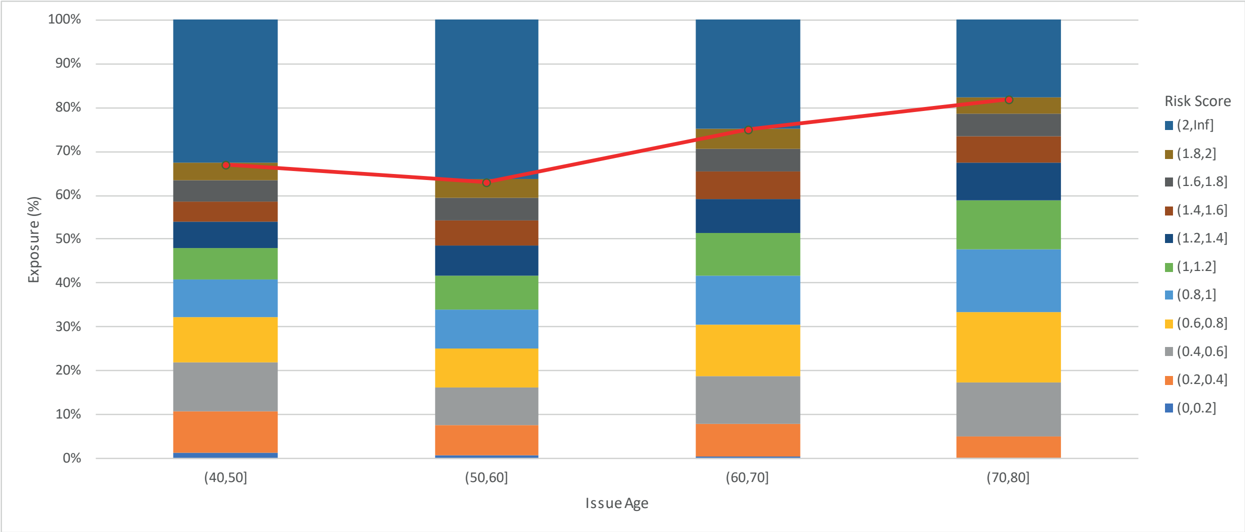
We can see that the gains in A/E would be higher (77% to 71%) but the pass rate would decrease from 80% to 73%.

In these examples, **Risk Score 3.0 Rx & Dx** slightly outperforms **Risk Score 3.0 Rx & Dx & Cr**. As seen previously for the Final Expense line of business, the addition of credit information generally improves mortality segmentation but performance may vary depending on the specific usage of the score.

Results

Now, a global 80% pass rate does not guarantee a 80% pass rate amongst all Issue Age groups. From **Figure 14**, we see that the distribution of **Risk Score 3.0 Rx & Dx & Cr** for the Final Expense business varies across Issue Age bands. Hence, a global 80% pass rate would lead to pass rates by Issue Ages varying from low 60% to low 80%, represented by the red curve.

Figure 14 – Distribution of Risk Score 3.0 Rx & Dx & Cr by Issue Age Band – Final Expense



Hence, depending on the pass rates goals, a user of the Risk Score may want to vary Thresholds by Issue Age but also by other characteristics such as Sex, Smoker Status, etc.

To summarize, Risk Score Thresholds must be carefully selected to ensure appropriate risk classification and pass rates and these thresholds will vary depending on the version of the Risk Score used. Optimum Re may guide insurance companies through an optimal use of the Risk Score.

Conclusion

Optimum Re confirms the mortality segmentation power of the Risk Score. More specifically, **Risk Score 3.0 Rx** is an improvement over the previous version, **Risk Score 2.2 Rx**. Additionally, **Risk Score 3.0 Rx & Dx & Cr** seems to perform the best for the Final Expense and Life Business with smaller gains over its credit-less counterpart in Final Expense. Finally, hit rates increase with the addition of new information (credit, medical) to the Risk Score, further increasing the value of **Risk Score 3.0 Rx & Dx & Cr**.

Furthermore,

- The mortality segmentation power of the Risk Score decreases with duration but some selection effects still remain at the latest durations provided in the dataset (17 years).
- The use of Risk Score provides significantly superior mortality segmentation power over the use of drug colors with Rx Check.
- Insurance companies should set Risk Score Thresholds with care in order to achieve desired pass rates and overall levels of mortality.

Even if the data on which this analysis is based was provided by Milliman itself, with no indication of which partition of the dataset was used to train the models, we found no indication that the performance of the Risk Score would vary significantly on new data, given it is not drastically different.

Optimum Re nonetheless recommends performing some retrospective analysis on your company's own portfolio prior to implementing Milliman's Risk Score. Optimum Re may assist your company in these tasks and provide recommendations on the mortality assumption to use in pricing your products.

Your Optimum Re Team for this report

Michel Hébert, FSA, FCIA
Vice President, Actuarial
mhebert@optimumre.com

Guillaume Ducharme, FSA, FCIA
Director, Research and Development
gducharme@optimumre.com

Who we are?

Since 1987, our company has been serving the US life and health reinsurance marketplace. Throughout the years, Optimum Life Reinsurance has maintained steady growth and profitability, earning a solid reputation as a reinsurer that provides efficient and professional client service.

Our mission is to support clients of all sizes by providing expertise and solutions in addition to reinsurance capacity. Year after year, we continue to successfully expand our presence in the market through organic growth, block acquisitions, and personalized solutions tailored to increase market share for our clients.

Optimum Life Reinsurance is the reinsurer of choice for more than 160 life insurance companies in the US and Caribbean markets.

About Optimum Financial Group

Optimum Financial Group is dedicated to the financial security of its clients since 1969. Global and privately-owned, it is active in the sectors of actuarial consulting, global asset management, general insurance, information technology, life insurance, life reinsurance, and real estate. The Group has 645 employees within diverse subsidiaries operating in 20 business places in Canada, the United States and in France. Its revenues are over 1 billion Canadian dollars, its assets under management in Canada, the United States and in France totalise nearly 8 billion Canadian dollars and its total assets rise up to nearly 6 billion Canadian dollars.

OPTIMUM®

Life Reinsurance

“The strength of expertise”

OPTIMUM RE INSURANCE COMPANY

1345 River Bend Drive, Suite 100
Dallas, TX 75247, U.S.A.
+1 214 528-2020
+1 214 528-2777

www.optimumre.com
www.optimumre.com/linkedin