RGA's Actuarial Validation of Milliman Irix® – Risk Score 3.0 with Credit

Mortality-Risk-Based Predictive Models Using Prescription Drug, Medical Billing, and Credit Histories

Mark Ma
Vice President and Managing Actuary

Guizhou Hu
Vice President, Head of Risk Analytics

Taylor Pickett
Vice President and Actuary

www.rgare.com

Executive Summary

In December 2022, RGA published the whitepaper, RGA's Actuarial Validation of Milliman Irix® – Risk Score 3.0 . Since then, RGA assessed Milliman Irix - Risk Score 3.0 with Credit, which includes four different models and scores, Rx3.0 based on prescription drug data only; RxDx3.0 based on prescription drug and medical billing data; RxCr3.0 on prescription drug and credit data; RxDxCr3.0 on prescription drug, medical billing data, and credit data when at least one of them is available.

All Milliman Risk Score 3.0 products are effective in segmenting mortality. The most powerful one is the RxDxCr3.0. Adding credit into Rx and RxDx scores could significantly improve the risk segmentation. RxDxCr has the lift of 17 times while RxDx has the lift of 15 times.

Credit and medical-based scores rank mortality risks differently. In other words, if targeting at the same mortality level, using RxDx and RxCr would end up with different populations. RxDx and RxCr has similar lifts. However, credit has higher hit rates.

RGA assisted clients in using different versions of Risk Scores for a variety of purposes. Our experience shows that it is critical to understand the comprehensive objectives of an underwriting program to choose the best solutions. RGA continues to see value in customized analysis to understand the impact of using scores for each company's products, market, and use cases.

Milliman Irix® Risk Scores – Data and Models

RGA's US Mortality Market provides an unbiased assessment of risk segmentation tools. Our underwriting and pricing teams routinely engage clients and vendors in better understanding the industry trends and tools our clients use. In December 2022, RGA published the whitepaper, RGA's Actuarial Validation of Milliman Irix® – Risk Score 3.0. Since then, RGA assessed Milliman Irix – Risk Score 3.0 with Credit, which currently encompasses four mortality-risk-based predictive models:

- Risk Score 3.0-Rx ("Rx3.0"), utilizes prescription drug data ("Rx") only.
- Risk Score 3.0-Rx & Dx ("RxDx3.0"), utilizes prescription drug and medical billing data ("Dx") when either or both are available.
- Risk Score 3.0-Rx & Cr ("RxCr3.0"), utilizes prescription drug and credit data ("Cr") when either or both are available.
- Risk Score 3.0-Rx & Dx & Cr ("RxDxCr3.0"), utilizes prescription drug, medical billing data, and credit data when at least one of them is available.

Milliman also provided RGA the scores of the previous generation, the Risk Score 2.2 version of the model ("Rx2.2") and Risk Credit Score 1.0 version of the model ("RxCr1.0").

The dataset Milliman provided includes scores for 42.3 million individual insurance applicants, spanning application dates between 2005 and the end of 2020. Death data was also provided, covering calendar years 2006 through the end of Q1 2021, allowing RGA to perform a mortality study with 236M exposure years and 1.7 million deaths. When compared to the validation data for the Risk Score 3.0 RGA received from Milliman in 2022, the new data, for Milliman Irix® Risk Score 3.0 with credit, has the same population, business, and demographic information. The life insurance line of business in the dataset represents applicants of a variety of underwriting methods, ranging from full underwriting to non-medical and simplified issue.

Evaluation Method

Based on the dataset from Milliman, RGA conducted an independent validation by developing a mortality study of the data through calendar year 2020. It is the same mortality study used in the previous whitepaper, i.e., the U.S. population mortality table as the expected basis. The analysis in this paper is based on the entire dataset Milliman used for training, testing, and validation.

Hit Types and Mortality

Not every individual has all Rx, Dx, or Cr histories. Below is the summary of hit types and their corresponding percentages of exposures.

There are three types of Rx hits:

- those who were not found in the Rx database (No Rx hit), 23.3%.
- those who exist in the enrollment data but do not have any Rx history (Eligibility only), 13.7%.
- and those with both enrollment and Rx history (Rx hit), 63.0%.

As enrollment data was not provided for Dx, there are only two types of Dx hits:

- those with Dx history, 66.4%.
- and those without, 33.6%.

There are three types of credit hits:

- those whose credit history are not ordered (credit not ordered). They are people younger than 18 at issue, 4.6%.
- those whose credit was ordered but do not have any credit history (credit no hit), 9.7%.
- and those with credit history (credit hit), 85.6%. A small percentage of lives before 2008, even though they have credit history, was not scored. We left the discrepancy in some analysis below as they are immaterial for our observations. Credit hit rates tend to be high for middle ages but low for young and old ages.

An individual could have one or two types of history but not the other(s). If one type of history is available, the scores with that data type would be produced. For example, if a person has Rx history but no Dx or credit history, the person will have all four scores, i.e., Rx, RxDx, RxCr, and RxDxCr. If a person has a credit hit but no Rx or Dx hit, the person will have RxCr, RxDxCr, but no Rx nor RxDx scores. Below, we compare the exposure distributions by the types of evidence.

Rx versus RxCr

Exhibit 1 below illustrates the relationship between Rx3.0 and RxCr3.0.

Rx Hit Type				
Credit Hit Type	**No Rx Hit**	**Eligibility-Only Hit**	**Rx Hit**	**Total**
Credit Hit	16.5%	11.9%	57.2%	85.6%
Credit No Hit	4.8%	1.2%	3.8%	9.7%
Credit Not Ordered	2.0%	0.7%	2.0%	4.6%
Total	23.3%	13.7%	63.0%	100.0%

The green highlighted cells, about 91.4% of exposures, have either Rx or credit history. They have RxCr3.0 scores. For those with Rx hit but no credit histories (either no credit hit, or credit not ordered), about 5.8% of exposure, their RxCr3.0 scores are the same as Rx3.0 scores. RxCr3.0 scores range from 0.106 to 326.743.

Generally speaking, the higher the hit rate of a type of evidence, the higher the mortality of those without a hit. For example, Rx only model has the hit rate of 63.0% (the last column of Exhibit 1). Those without Rx scores have the mortality of 107.8% of the aggregate mortality. RxCr has a significantly higher hit rate, 91.4%. Those without RxCr scores have the mortality of 126.6% of the aggregate mortality. Therefore, when comparing different models or setting mortality assumptions, it is important to understand the business purposes, target markets, as well as how no hit and not ordered are handled.

RxDx versus RxDxCr

Exhibit 2 below illustrates the relationship between RxDx3.0 and RxDxCr3.0. A vast majority of exposures have at least one of Rx, Dx, or Cr history. For those credit not ordered, no RxDxCr score is populated. That leads to the hit rate of 91.5% (green highlighted cells in exhibit 3). RxDxCr scores of those with either Rx or Dx hit but no credit hit are the same as RxDx scores. RxDxCr score ranges from 0.068 to 434.108.

Exhibit 2: Exposure by Rx, Dx, and Credit Hit Type

Rx Hit Type			
Credit Hit Type	**Rx or Dx Hit**	**Neither Rx nor Dx Hit**	**Total**
Credit Hit	71.3%	14.4%	85.6%
Credit No Hit	5.9%	3.9%	9.7%
Credit Not Ordered	3.3%	1.3%	4.6%
Total	80.4%	19.6%	100.0%

Like the comparison of Rx to RxCr, including credit into RxDx could either increase or decrease a score, for those with all three data elements.

Model Performances

Aggregate Performances

Exhibit 3: Relative Mortality by Decile Buckets

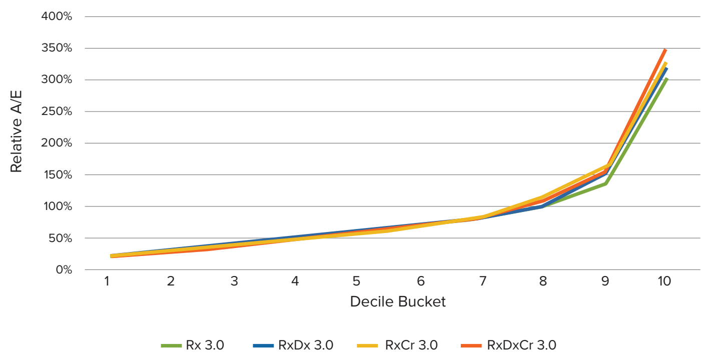


Exhibit 3 shows the relative mortality by decile buckets for each score. For example, the green line represents the mortality lift curve of the Rx3.0 model. Each data point represents 10% of the exposure for the model it represents. As explained in Table 1, 63% of the total exposure is scored by Rx3.0. Therefore, each dot on the green line represents 6.3% of the total exposure in the dataset RGA received. Each line is adjusted to its own aggregate mortality level to derive relative mortality.

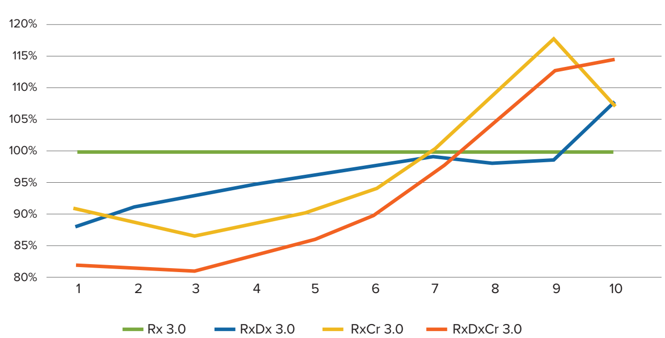
All the lines are monotonically increasing, indicating strong segmentation of mortality risks. RxDxCr3.0 has the steepest slope, suggesting the strongest segmentation power.

All lift curves have a hockey stick pattern, where the relative mortality jumps significantly to the right side of the chart.

The differences among the four lift curves are significant, even though they are visually overlapping one another. It is mostly because the scales are dominated by the buckets with high relative mortality. Exhibit 4 presented the same data as Exhibit 3 but in a different format. Each model takes ratios to its correspondent Rx3.0 relative actual to expected ("A/E").

Exhibit 4 presented the same data as Exhibit 3 but in a different format. From decile 1 to decile 6, RxDx3.0, RxCr3.0, and RxDxCr3.0 can do a much better job in capturing exposure with lower mortality risks. At the higher end, they do a better job in capturing exposure with higher mortality risks.

Exhibit 4: Relative Mortality Using Rx3.0 as Baseline



The models with credit data have higher segmentation power than the counterparts without credit. For example, in the first decile, the mortality of RxDx3.0 is 88% of that of Rx3.0, while that of RxDxCr is 82%. This suggests that RxDxCr is more effective in finding people with lower mortality risks.

Another way to summarize the segmentation power is to examine the level of lift, defined as the ratio of the relative mortality of the top 10% exposure to that of the bottom 10% exposure. Below is a summary of lifts for each score.

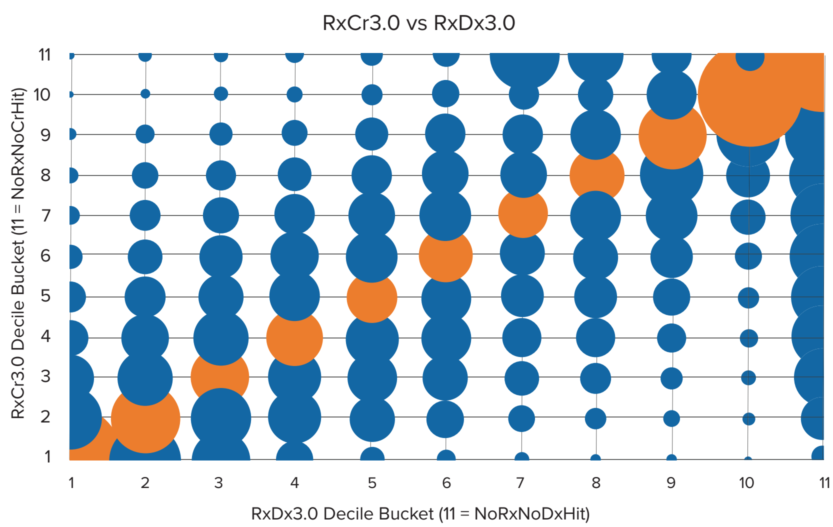
RxDxCr 3.0	16.9
RxCr 3.0	14.3
RxCr 1.0	11.1

RxDx 3.0	14.8
Rx 3.0	12.1
Rx 2.2	7.2

RxDx3.0 has a lift of 14.8 times. RxCr3.0 has a lift of 14.3 times. RxDxCr3.0 is the most powerful among all, with a lift of 16.9 times.

Credit and medical-based scores rank mortality risks differently. Exhibit 5 below compares the distribution of RxCr and RxDx. Each of the first 10 columns represents a decile of exposure by RxDx scores. The last column is for those without RxDx scores. Similarly, each of the bottom 10 rows is a decile of RxCr scores. The top row is for those without RxCr scores. Since both RxDx3.0 and RxCr3.0 scores have the Rx component, they are correlated. People with low RxDx3.0 intend to have low RxCr scores and vice versa. For those with neither Rx nor credit hit, the exposure is mostly in RxDx3.0 decile bucket 7 to 9. This is consistent with the observation that people without those hits, especially those without credit hits, tend to have higher mortality risks. For those with no Rx and no Dx hit, the exposure by RxCr3.0 decile bucket are more evenly distributed. The people with neither Rx nor Dx scores still have elevated mortality but not as high as those without credit hits.

Exhibit 5 – Distribution of exposure by RxCr and RxDx scores



Incremental Segmentation of Cr and Dx, on top of Rx

In this section, we try to answer the question that assuming Rx as table stake, how much additional segmentation power do we see in RxDx versus RxCr models? With that in mind, we focus on life business with no red drugs in the history, between ages 20 and 69.

For those with all three hits (e.g., have Rx, Dx, and Credit hit), we group them into 5 categories, ranked by Rx3.0 scores. For example, the 20% exposure with the lowest Rx scores are categorized into group 1. Within each category, people are re-grouped into deciles of RxDx3.0, RxCr3.0, and RxDxCr3.0, to calculate the lifts of each model. Exhibit 6 compares the lifts of the three models, RxDx, RxCr, and RxDxCr within each group.

Exhibit 6 – Lifts by Rx3.0 Groups for those with Rx, Dx, Cr hits

Rx3.0 Group	RxDxCr3.0	RxDx3.0	RxCr 3.0	Rx3.0
1	4.2	3.6	3.4	2.1
2	5.4	4.3	3.7	1.4
3	8.5	5.8	5.5	1.3
4	11.8	7.5	7.2	1.4
5	18.1	11.1	11.1	3.5

All three models perform better when Rx scores raise. In other words, among people with Rx history indicating a poor risk, considering Dx and Cr could further segment the mortality risks. Group 5 is more diverse in terms of Rx3.0 score, as illustrated by the lift of 3.5 times. The heterogeneity of Rx scores partially explains the observed increases of lift from the other three models in group 5. Interestingly, RxDx and RxCr are comparable overall.

Early in this article, we called out that the different hit rates would correlate to mortality for those without hits. For the subpopulation in this section, credit has a significantly higher hit rate, 93.4%, than medical billing data does, 78.3%. Therefore, those with no credit hit or credit not ordered would have higher mortality than those without a medical billing data hit. Since credit hit rates vary by ages, the mortality differential for those without Cr or Dx hits vary by ages too.

Exemplary Case Studies

RGA routinely assists carriers in designing and improving underwriting programs, empowered by a variety of vendor solutions.

Many carriers are seeking the greatest possible mortality lift to offset the lack of fluids. We are engaged with multiple carriers in adding Dx into their underwriting programs that currently utilize credit and Rx data. Some supplemented the risk-based scores with medical rules to better reflect their underwriting philosophy and meet their unique business needs. Carriers may find retrospective studies helpful in determining the mortality implications of these program updates and assessing the combined impact of medical rules and a risk-based score. As part of that process, RGA is well-positioned to assist carriers through the calibration process with a dedicated, multi-discipline team that has designed dozens of underwriting programs in close collaboration with our clients.

Some carriers desire to remove certain evidence without significant additional mortality slippage. RGA sees potential in adding medical data to mitigate the impact of removing credit data. Especially for those who are currently using previous generation Rx and credit information, the results of this study suggest that the decrease of segmentation due to dropping credit as a requirement can be largely offset by using the upgraded model with both Rx and Dx. Note that the change of evidence could still have pricing implications. The target market business distributions, such as age mix and exclusivity relative to other tools used, all play a vital role in evaluating the impact of changing evidence. The types of scores involved are not independent. It is important to evaluate people without certain hits. RGA's experience with a large variety of risk segmentation tools and deep commitment to clients enable customized solutions to meet each carrier's unique objectives.

Summary and Limitations

All Milliman Risk Score 3.0, including Rx3.0, RxDx3.0, RxCr, and RxDxCr, are effective in segmenting mortality. Adding credit into Rx and RxDx scores could further segment both the low-risk end of the spectrum and the high-risk end. The most powerful one is the RxDxCr3.0. Milliman Risk Score 3.0 can significantly segment mortality risks across the business attributes RGA examined, even though the level of segmentation varies by populations and durations.

Traditionally, the protective value of credits and Rx tend to be assessed separately. The Milliman Irix RiskScore 3.0 with Credit data contains Rx, Dx, and credit together. It allows us to analyze the exclusivity among those three evidences within the framework of the models. However, the increased complexity of scores could be a drawback. Generally speaking, the more evidence a predictive model includes, the better it can segment risks, but the harder it is to explain the predictions. Regulators and other stakeholders increasingly demand transparency, explainability, and fairness in using AI. Some carriers might desire more than a score to make an underwriting decision. This analysis does not consider the expense associated with ordering more evidence, and the cost differential of different scores. It is important to understand the comprehensive objectives of an underwriting program, to choose the best solutions.

Moreover, a carrier's application pool and insured population might not have the same underlying characteristics as the population in this study. Model performances vary by business attributes. The exclusivity of the evidence and models must be considered in the context of the underwriting process. The value of the scores will be impacted by other evidence used in underwriting. Therefore, RGA continues to see value in customized analysis to understand the impact of using scores for each company's products, market, and use cases. RGA is experienced in helping clients with deeper and more contextual analysis.

Reference

1. http://www.rxhistories.com/wp-content/uploads/2023/01/RGA_Actuarial_Validation_of_MillimanIntelliScript_Irix_RiskScore3.0.pdf
2. http://www.rxhistories.com/wp-content/uploads/2023/01/RGA_Actuarial_Validation_of_MillimanIntelliScript_Irix_RiskScore3.0.pdf
3. http://www.rxhistories.com/wp-content/uploads/2023/01/RGA_Actuarial_Validation_of_MillimanIntelliScript_Irix_RiskScore3.0.pdf