



ExamOne's HealthPiQture Score — Stratification and Suitability

Evaluation of HealthPiQture Model Lift and a Novel Comparison to Traditional Insurance Exam + Lab Risk Identification

Executive Summary

- ExamOne's HealthPiQture is an ensemble model built on medical information (prescription history, lab results history, diagnosis history and medical claims).
- PartnerRe's independent evaluation¹ of the HealthPiQture model confirmed its association with mortality risk and its additional predictive lift over the individual constituent scores in isolation.
- PartnerRe also compared the mortality risk assessment from HealthPiQture (i.e., its score) with medical information typically captured during a life insurance application. The intent of this comparison was to assess, at a high level, the applicability of HealthPiQture for different use cases in an underwriting program. Our evaluation suggested that the HealthPiQture score and medical information used in traditional underwriting are consistent approximately 68% of the time, particularly for persons assigned to the lower risk categories. Examination of the 32% of cases where risk assessment differed between HealthPiQture and our internal estimations, indicated that each approach can identify different types of mortality risk in the sample population.
- PartnerRe's comparison of HealthPiQture to traditional exam and lab tests indicated that careful analysis should be performed when HealthPiQture is being considered as a replacement for existing exam and lab requirements, since the identified risks may differ. However, by the same token, the addition of HealthPiQture may provide meaningful protective value even on fully underwritten life cases with exam and fluids tests.

¹ This evaluation is intended to objectively present the results of our detailed review of the model and does not serve to promote this model.

Objectives

ExamOne developed a mortality risk model – HealthPiQture – that relies on three component scores – Rx, Labs, and Dx. These three scores are based on:

1. prescription history from ScriptCheck – ExamOne's proprietary prescription history offering (Rx),
2. clinical laboratory test history from LabPiQture reports (Labs), and
3. diagnosis codes from LabPiQture and/or medical claims data available through ExamOne databases (Dx).

The HealthPiQture score is based on these three constituent scores, and this score is meant to estimate the overall mortality risk. A score of zero is the median mortality rate of the population used by ExamOne to develop the model, which is equivalent to an actual/expected deaths (A/E) value of 116% (E = 2015 Unismoke Valuation Basic Table (VBT)). HealthPiQture scores that are lower or higher than zero reflect, accordingly, risk that is lower or higher than 116% of the 2015 Unismoke Valuation Basic Table. HealthPiQture could, hypothetically, be an alternative to medical information commonly utilized during life insurance underwriting process. Using data provided by ExamOne, PartnerRe conducted an independent assessment of the HealthPiQture model. Our two goals were to:

1. verify whether HealthPiQture scores are associated with mortality risk, and
2. compare risk segmentation between the ExamOne product and medical information typically used in life insurance underwriting for life insurance applicants.

Data

ExamOne provided PartnerRe with a sample of data utilized for their own mortality study². The shared mortality dataset contained 4,147,023 individuals, who were scored between 2005 and 2019. For each person, we were provided with their gender, age, death indicator, four ExamOne scores (HealthPiQture, Rx, Labs, Dx), scoring date, and follow-up time. The data contained 61,466 deaths and 17,302,205 person-years of exposure. The follow-up time varied between individuals; the median was 3 and the maximum was 15 years.

In addition to the mortality data, ExamOne provided anonymized data on 95,436 life insurance applications. This included information collected during the insurance application process, such as BMI, blood pressure, smoking indication, viral markers, variety of clinical lab tests including results of comprehensive metabolic panel, lipids panel, sugar metabolism, urinalysis, and others. For each record, ExamOne provided the four scores (HealthPiQture, Rx, Labs, Dx), and the records' historical data on prescriptions, clinical laboratory

tests, diagnoses, and medical claims. These data lacked mortality information. After we cleaned and filtered out rows with scarce medical information acquired during the insurance application process, the data contained 84,658 records.

Results

HealthPiQture score composition and mortality risk stratification – Analysis of the mortality data

We assessed the composition of HealthPiQture scores and HealthPiQture's ability to stratify mortality risk based on the mortality data provided to us by ExamOne. In **Table 1**, we show that not all three constituent scores are present simultaneously. All three constituent scores are present for approximately 5% of all individuals, while the majority of HealthPiQture scores are generated from a combination of two scores – Dx and Labs (43%) or Rx score only (48%). Dx and Labs scores are often generated from the same information (e.g., laboratory tests to confirm a diagnosis), and as such, it is unsurprising that these were the most common two-score combination.

Score Combination	Rows		Exposure		Deaths
	Number	Proportion (%)	Person-Years	Proportion (%)	
Dx, Labs	1,782,675	43.0	11,824,679	68.3	47,051
Rx	1,985,593	47.9	3,936,261	22.8	8,907
Labs	175,069	4.2	1,139,049	6.6	4,532
Rx, Dx, Labs	183,412	4.4	362,702	2.1	880
Rx, Labs	20,274	0.5	39,514	0.2	96
HealthPiQture/Total	4,147,023	100.0	17,302,205	100.0	61,466

Table 1: The frequency, exposure in person-years, and number of deaths for different combinations of constituent HealthPiQture scores that were used to generate the HealthPiQture score in the mortality data.

² These data were entirely deidentified following HIPAA standards.

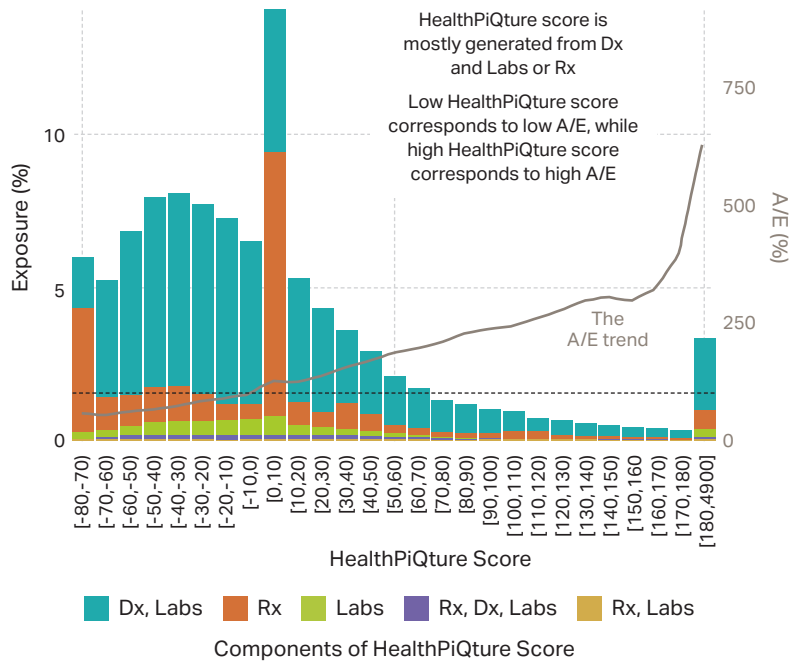


Figure 1: Exposure for binned HealthPiQture scores and the corresponding A/E, where E is measured using the 2015 Unismoke Valuation Basic Table. The dashed line indicates A/E = 100%. Source: PartnerRe.

We further examined the composition of the HealthPiQture score and how well it stratifies mortality risk (**Figure 1**).

The HealthPiQture score ranges from -80 to 4,900. Any score that is computed as lower than -80 is reported as -80. Likewise, any score that is computed as greater than 4900 is reported as 4900. We grouped HealthPiQture scores into 10-unit intervals but truncated records with a score at and above 180 because of a long tail. The grey line shows the A/E deaths (E = 2015 Unismoke Valuation Basic Table). Bars in Figure 1 show the proportional contribution of constituent scores to overall exposure and each color represents a different combination of the HealthPiQture score constituents.

Risk segmentation in score quartiles

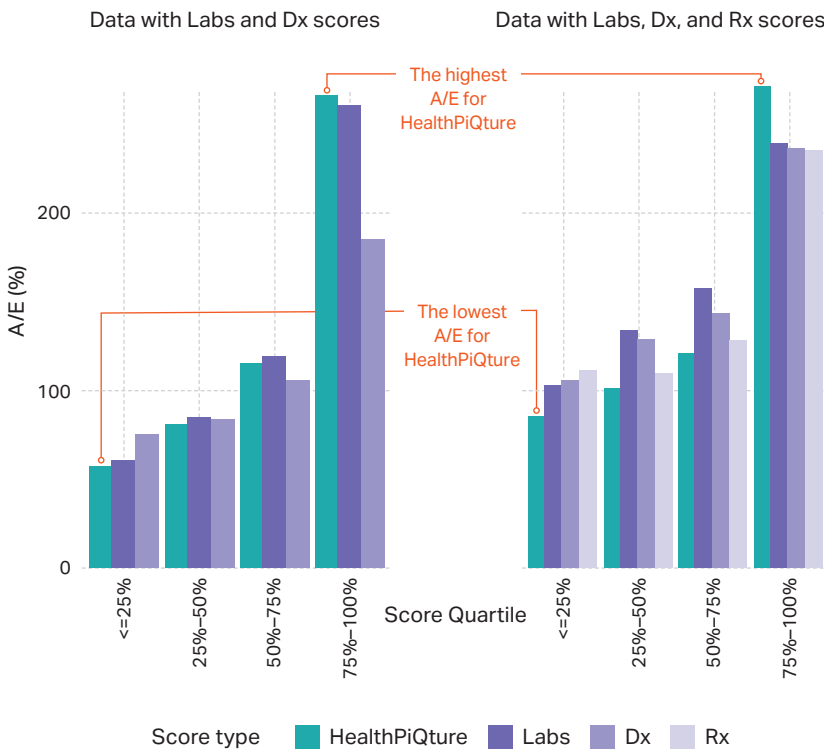


Figure 2: Mortality risk segmentation in scores quartiles. Source: PartnerRe.

First, we show that the HealthPiQture score does segment mortality risk well – we see a steady increase in A/E with increasing score. We do observe both an exposure spike and mortality bump at the 0-10 bin, which could be attributed to individuals with medical information (largely individuals with prescription history only) that is insufficient to shift the mortality risk below or above ExamOne's population median (i.e., score 0). We also show that apart from records with a score from -80 to -70 and from 0 to 10, we have more exposure for records that had an overall score generated from Dx and Labs. Rx only HealthPiQture scores are most dominant (i.e., have the highest exposure) when scores are between -80 and -70 or 0 and 10. Although not shown in this chart, Rx scores between 0 and 10 are more common for men than for women.

We also compared risk segmentation among the four scores, which is shown in **Figure 2**.

We selected two subsets of the mortality data: rows with only Labs and Dx scores (1,782,675 records) and rows with all three constituent scores: Rx, Labs, and Dx (183,412 records). By grouping scores into quartiles, we show that individuals assigned to lower quartiles have lower risk than individuals assigned to higher quartiles. This risk segmentation is the greatest for HealthPiQture – this score had the largest mortality difference between the lowest and highest quartile. Consequently, if we use the HealthPiQture, versus Labs, Dx, or Rx scores in isolation, we will observe the lowest mortality among persons with the lowest 25% of scores. Likewise, for the persons with the highest 25% of HealthPiQture scores, we will observe a higher mortality than those in the top 25% of Rx, Labs, or Dx scores. Thus, a carrier, when estimating the mortality risk, should also consider the underlying HealthPiQture composition. If the carrier would like to assign applicants to a certain risk group, they must adjust the threshold accordingly. For instance, based on mortality data, for the population with all three scores available, the A/E (E = 2015 Unismoke Valuation Basic Table) for persons with scores below -10 is approximately 90% if the threshold is set based on the ensemble of all three scores, and 107% if based on only the Rx score.

HealthPiQture score vs. information used in traditional medical life underwriting – Analysis of data on insurance applications

We had two goals for our analysis on the insurance applications data provided by ExamOne. Our first goal was to determine if the composition of HealthPiQture scores were the same for the insurance application and mortality datasets. Second, we aimed to interrogate the ways in which risk assigned from HealthPiQture would compare to that based on medical information used in traditional life underwriting.

Score Combination	Rows	
	Number	Proportion (%)
Rx, Dx, Labs	58,401	69.0
Dx, Labs	7,365	8.7
Rx, Labs	7,079	8.4
Rx	5,650	6.7
Rx, Dx	4,682	5.5
Labs	879	1.0
Dx	602	0.7
HealthPiQture/Total	84,658	100.0

Table 2: The frequency of different combinations of constituent scores in the insurance application data that were used to generate the overall score. The majority, 69%, have all three constituent scores.

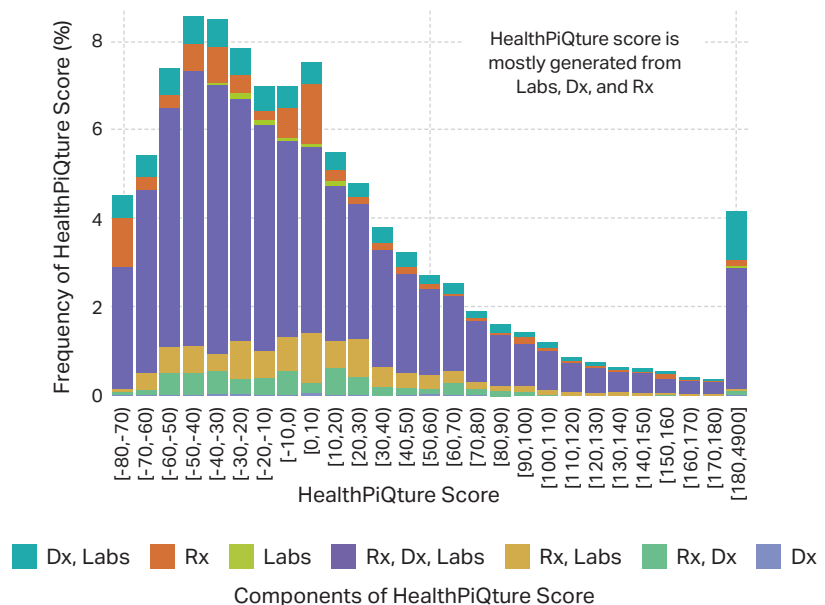


Figure 3: Frequency of the 10-unit HealthPiQture score bins for the life insurance applicants' data. Source: PartnerRe

We assessed whether records in the insurance applications data had the same distribution of constituent HealthPiQture scores as the mortality data. We examined the proportion of records that had an HealthPiQture score based on various combinations of the constituent scores. In **Table 2**, which presents the mixture of the constituent scores used to generate the overall scores, we show that most records had all three constituent scores.

We examined whether the trend we observed in **Table 2**, where most insurance records had HealthPiQture scores resulting from all three constituent scores, was present across different score ranges. In **Figure 3**, which presents the HealthPiQture score distribution and its composition for 10-unit bins, we show that all ranges of HealthPiQture scores were largely the result of the three constituent scores. Thus, we can conclude that the composition of

the overall score for the life insurance records differs from that in the mortality data. While the presence of all the three constituent scores in the mortality data is rarely observed, it is the most frequent combination in the provided sample of insurance application data. Because risk segmentation is strongest when HealthPiQture is based on all three constituent scores (Figure 2), we would expect the predicted mortality risk from HealthPiQture to be more reliable for the typical insurance applicant than for the larger population.

Our final step was to examine how congruent risk assigned by a HealthPiQture score would be to risk assigned with medical information used in traditional life underwriting. Using a PartnerRe internal model, which reflected the information used in traditional life underwriting (i.e.,

age, gender, smoker status, blood pressure, BMI, laboratory tests), we assigned a risk score to each record. Since the HealthPiQture score can be informed by medical information that is both typical (e.g., glycohemoglobin) and atypical (e.g., complete blood count) in traditional life underwriting, it could be seen as an alternative to the laboratory tests that are typically requested during an insurance application. Our comparison between HealthPiQture and our internal model provides insights on the overlap between the risks evaluated from these two approaches.

We determined which score, from both models (i.e., HealthPiQture and our internal model), would be equal to the same mortality rate to establish cut points for low and high risk. We measured mortality as predicted deaths to expected deaths (P/E, E = 2015 Unismoke Valuation

Basic Table) and set the cut point as 150% P/E. Records with a score below the cut point were assigned to the low-risk group, and those with a score above the cut point were assigned to the high-risk group. This resulted in four possible scenarios, which are presented in **Figure 4**:

Agreement:

1. HealthPiQture score and risk score indicate low risk – lower left quadrant.
2. HealthPiQture score and risk score indicate high risk – upper right quadrant.

Disagreement:

3. HealthPiQture score indicates high risk, risk score indicates low risk – lower right quadrant.
4. HealthPiQture score indicates low risk, risk score indicates high risk – upper left quadrant.

Our internal model and HealthPiQture were consistent in assigning risk for 68% of the records, as shown in Figure 4. We caution that this consistency in risk assignment is disproportionately higher for persons with lower risk – our internal model and HealthPiQture both identified 60% of records as low risk. The 8% of records that both models agreed were higher risk only accounted for approximately one third of the cases that either model assigned to the high-risk category.

We also show in Figure 4 that HealthPiQture and our internal model usually identify different types of risk. Nearly the same proportion of records were assigned a different risk group by both models (i.e., 16% for our internal survival model; 17% for HealthPiQture). Because HealthPiQture does flag records that our internal model assigns as low risk, we can conclude that it has the potential to capture mortality risk not identified by medical tests typically used during insurance underwriting.

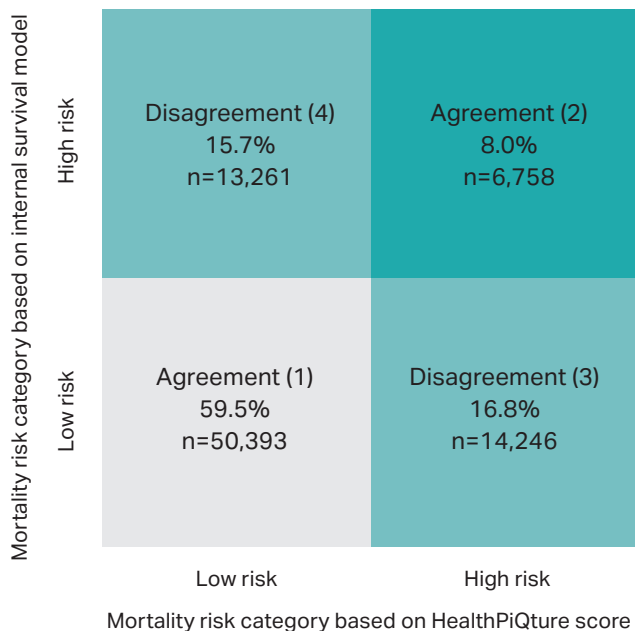
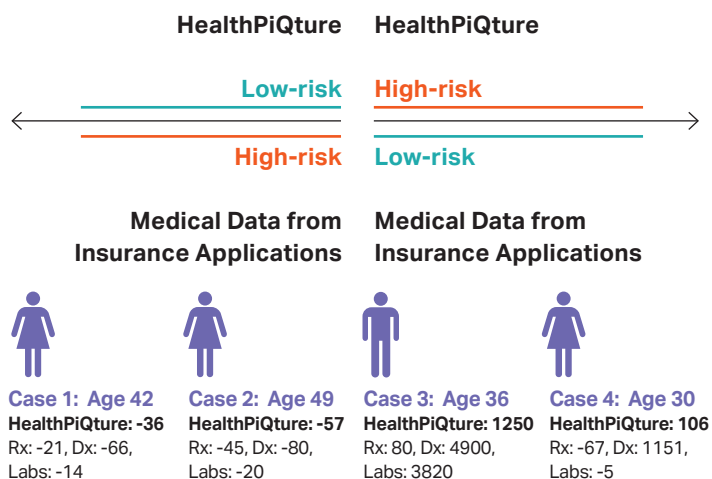


Figure 4: Distribution of the life insurance applications between low and high mortality risk. The low- and high-risk categories were assigned based on a PartnerRe internal model and the HealthPiQture score. Source: PartnerRe.

Case examples

The two models diverged on risk for 32% of the insurance application records, and we explore a few explanations for this divergence using four case examples in **Table 3**. For each case, we compared the medical information available to generate the HealthPiQture score with the medical information that would have been available using the typical insurance lab panel. To identify abnormal results, we used guidelines from U.S. Department of Health and Human Services; thus, which tests results are highlighted in the table as abnormal are not reflective of any underwriting guidelines. We caution that these case examples are meant for **illustrative purposes only** and are not meant to represent all scenarios in which the risk assigned by HealthPiQture would differ from insurance medical exams.



	Insurance Lab Test	Normal Range	Exam/ Lab Result	Exam/ Lab Result	Exam/ Lab Result	Exam/ Lab Result
Cardiovascular Risk	Body Mass Index (kg/m ²)	18.5-24.9	30	18	30	25
	Systolic Blood Pressure	90-120	Normal	Normal	Normal	123
	Diastolic Blood Pressure	60-80	Normal	Normal	82	84
Diabetes	Serum Glucose (mg/dL)	<140	244	Normal	Normal	Normal
	Glycohemoglobin (%)	<5.7	7.9 (diabetes)	5.8 (prediabetes)	Normal	Normal
Hyperlipidemia/ Cardiovascular Risk	Total Cholesterol (mg/dL)	<200	Normal	Normal	Normal	Normal
	HDL Cholesterol (mg/dL)	>40 (Males) >50 (Females)	29	Normal	34	Normal
	LDL Cholesterol (mg/dL)	<100	Normal	Normal	Normal	Normal
	Total to HDL Cholesterol	<4	4.59	Normal	Normal	Normal
	LDL to HDL Cholesterol	<3.5:1	Normal	Normal	Normal	Normal
	Triglycerides (mg/dL)	<150	214	Normal	Normal	Normal
Kidney Function	Microalbumin to Creatinine (mg/g)	<0.010 (Males) <0.012 (Females)	Normal	-	Normal	Normal
	Urine Albumin (ug/mL)	<30	Normal	-	Normal	Normal
	Blood Urea Nitrogen (mg/dL)	7-20	Normal	Normal	Normal	Normal
	Serum Creatinine (mg/dL)	0.8-1.2	0.5	0.6	Normal	Normal
	Urine Creatinine (mg/dL)	40-300 (Males) 27-250 (Females)	Normal	Normal	Normal	275
Liver Function	GGT (U/L)	<30	110	180	Normal	Normal
	ALP (U/L)	20-140	Normal	237	Normal	Normal
	ALT (ALT)(U/L)	29-33 (Males) 19-25 (Females)	73	207	17	11
	AST (AST)(U/L)	<35	42	62	Normal	Normal
Liver/Kidney Function	Total Bilirubin (umol/L)	1.7-20.5	Normal	Normal	Normal	Normal
	Globulin (g/dL)	2.3-3.5	Normal	3.9	Normal	Normal
	Total Protein (g/dL)	6-8.3	Normal	Normal	Normal	Normal
Smoker	Cotinine (ng/mL)	<10	Normal	Normal	Normal	Normal
	Smoker	Non-Smoker	Normal	Normal	Normal	Normal
Other Diseases	Hepatitis C Antibody	Negative	Normal	-	-	Normal
	Hepatitis B core antibody	Negative	-	-	-	Normal
	HIV Result	Negative	Normal	-	Normal	Normal

Table 3: For purely illustrative purposes, four case examples of divergent risk from HealthPiQture and the PartnerRe internal model utilized for this analysis, a model that uses information typically used in traditional medical life underwriting. Using guidelines from the U.S. Department of Health and

Human Services, we highlighted lab tests that were outside of the normal range. These normal/non-normal lab tests results do not therefore reflect any underwriting rules or guidelines. Source: PartnerRe.

High risk based on insurance medical information and low risk based on HealthPiQture

Case 1

When ExamOne's clinical lab history is limited, the HealthPiQture score for an applicant could suggest below-average risk, contrary to a risk assessment from medical tests typically used in life underwriting.

Using medical information from the insurance application process, we would assume that case 1 (a 42-year-old female) is high risk. We see evidence of poor liver function and a lipid disorder. Because of their elevated serum glucose and glycohemoglobin, we posit that they are diabetic. However, ExamOne's most recent clinical labs history (2 years prior to the scoring) did not show any abnormalities in their complete blood count panel (red and white blood cells, hemoglobin, hematocrit, platelets, etc.) or hormone blood tests, though there was an elevated level of one of the lactate dehydrogenase (LDH) isoenzymes, which interpretation is highly context dependent. The applicant diagnosis history indicates ovarian cysts, which might explain why the LDH tests were performed. While Exam One's database for this individual highlighted that there were tests that are not typically included in the insurance lab panel (e.g., complete blood count, hormone tests), the numerous tests used to ascertain if an applicant is at risk for cardiovascular issues (e.g., BMI, HDL), diabetes (e.g., glycohemoglobin), or liver/kidney failure (e.g., GGT, BUN) were absent. Thus, it is unsurprising that the clinical lab history was insufficient to flag the applicant's high blood sugar, lipid disorder, or poor liver function (Labs score of -14). The prescription history covers records only from 5-6 years prior to the scoring and includes prescriptions for medicines to treat high blood sugar. However, that did not suggest high mortality

risk (Rx score -21). The diagnosis history did not indicate that there were any concerning illnesses (Dx score -66). Thus, we posit that the resulting below-average HealthPiQture score may be because the applicant's historical medical information were insufficient to flag their cardiovascular, diabetic, liver function, and kidney function risk.

Case 2

We might also observe below-average risk from HealthPiQture, even when lab results that are systematically applied in the traditional medical underwriting process suggest above-average risk, when we lack recent information on an individual.

This 49-year-old female, based on medical information collected during underwriting, is assumed to be high risk. They are underweight; and in nearly all the liver function tests, their scores are outside of the normal range. Based on ExamOne's data, this individual's most recent clinical laboratory history was 2.5 years before they were assigned a score. At that time, they underwent an HPV test. In addition, their clinical labs history contained results from urine analyses performed 4 years prior to scoring. The results did not suggest any ailments (i.e., Labs score of -20). In and around the period these clinical tests were done, they were diagnosed with pelvic pain and hematuria, which did not warrant high mortality risk (i.e., Dx score of -80). Similarly, the Rx history included prescriptions for antibiotics, capsules to treat certain stomach problems, and single entry for medicine to control high blood sugar – none of these resulted in high mortality risk (i.e., Rx score of -45). Thus, we posit that the applicants' HealthPiQture score that suggests below-average risk (i.e., -57), though their insurance lab tests suggest above-average risk, may be explained by a recent change in their health status that is not captured in their historical medical information.

Considerations

When an applicant's clinical labs history is limited or dated, carriers may wish to consider using additional rules when incorporating HealthPiQture into their underwriting program. As shown with Case 1 and 2, there may be instances when the HealthPiQture score suggests a lower risk than an insurer would have assigned using medical exams, and this could be because the relevant tests were either absent or dated in the clinical lab history. Though the clinical labs may have been limited or dated, the prescription histories did suggest some risk (i.e., medicines used to treat diabetes). To manage these cases, an insurer could add a rule that flags applicants with concerning prescriptions in their history for further review, especially when the clinical labs history is limited or dated. Therefore, depending on the carrier's mortality assumptions, additional rules might be necessary when using the HealthPiQture score.

Low risk based on insurance medical information and high risk based on HealthPiQture

Case 3

The HealthPiQture score might suggest above average risk when an applicant's clinical lab history is inclusive of tests that are not included in insurance panel but are crucial in diagnosing some illnesses. This 36-year-old male, based on the medical information collected during underwriting, is low risk, even with a slightly abnormal result for one of the liver functions tests. Their Labs score is 3,820 and the historical clinical labs suggest poor cholesterol levels, low ALT, abnormal hormones levels (i.e., testosterone, luteinizing hormone, follicle stimulating hormone) and a divergent red cell distribution width. Additionally, there are past diagnoses that were not picked up by the tests conducted during underwriting, for example – disorder of the adrenal

gland, long term drug therapy, and other (Dx score of 4,900). Thus, their HealthPiQture score of 1,250 is inclusive of the numerous concerning diagnosis and abnormal clinical labs, which were not captured by the insurance lab panel.

Case 4

As with Case 3, when an applicant's clinical laboratory history includes test that are not typically used in traditional medical underwriting, the HealthPiQture score could suggest above average risk.

This 30-year-old female had a few concerning insurance lab results, but they were only slightly outside of the normal range. In this case, we could assume that this applicant is at average or low risk, depending on the underwriting rules. Similarly, their Labs (i.e., -5) and Rx score (i.e., -67) suggest low risk. However, within the past 1.5 years, they were diagnosed with a few illnesses (disorders of lipoprotein metabolism and other lipidemias, elevated blood glucose level, iron deficiency anemia, HPV, and disorder of thyroid) that increase their mortality risk, which could explain why their Dx score (i.e., 1,151) was poor. Because of these diagnoses, this individual was assigned as high risk based on their HealthPiQture score risk (i.e., 106).

Conclusion

Based on the provided data, we conclude that the HealthPiQture score can stratify mortality risk and is better at risk segmentation than any of its constituent scores (Rx, Dx, Labs) alone.

While HealthPiQture, in isolation, cannot identify all individuals that would be of concern, it can identify additional mortality risk when:

1. the traditional exams and tests used in the life insurance application process are not sufficient to identify or diagnose an illness, and
2. there is limited or no medical information available on an applicant, for instance with simplified issue policies.

Our findings, therefore, distinguish two clear use cases of HealthPiQture: as an additional layer of risk protection on an existing underwriting program, or as part of an electronic, data-driven, simplified issue product offering, as well as accelerated underwriting programs.

When considering the use case of HealthPiQture as an exam and lab replacement, our analyses indicates that the risks identified may not always overlap with those found by traditional paramedical exams and blood/urine tests. Thus, additional analysis may be necessary to fully understand the impact on expected mortality under this scenario.

Limitations

The insurance sample in this study were provided by ExamOne, and the trends we uncovered in these data will not be consistent for each insurer. In addition, the rules used in HealthPiQture may differ from insurers' underwriting rules and criteria. Consequently, we recommend that insurers interested in implementing the HealthPiQture score conduct analyses to understand how best to set thresholds, and incorporate new rules for these thresholds, in the view of their own underwriting criteria.

Contact us for more details

Though not presented in this evaluation, we at PartnerRe have conducted additional analyses on the HealthPiQture model. Consequently, we are positioned to provide your organization with greater insights on considerations when deploying this tool within your organization. Please reach out to us as we will happily discuss our insights with you.

Contributors

Tom Fletcher, PhD.
VP, Data Analytics, North America Life

Malwina Staniuk, PhD.
Data Scientist, Europe Life

Jody Daniel, PhD.
Data Scientist, North America Life

Ehren Nagel, FSA, MAAA
Consulting Actuary, North America Life