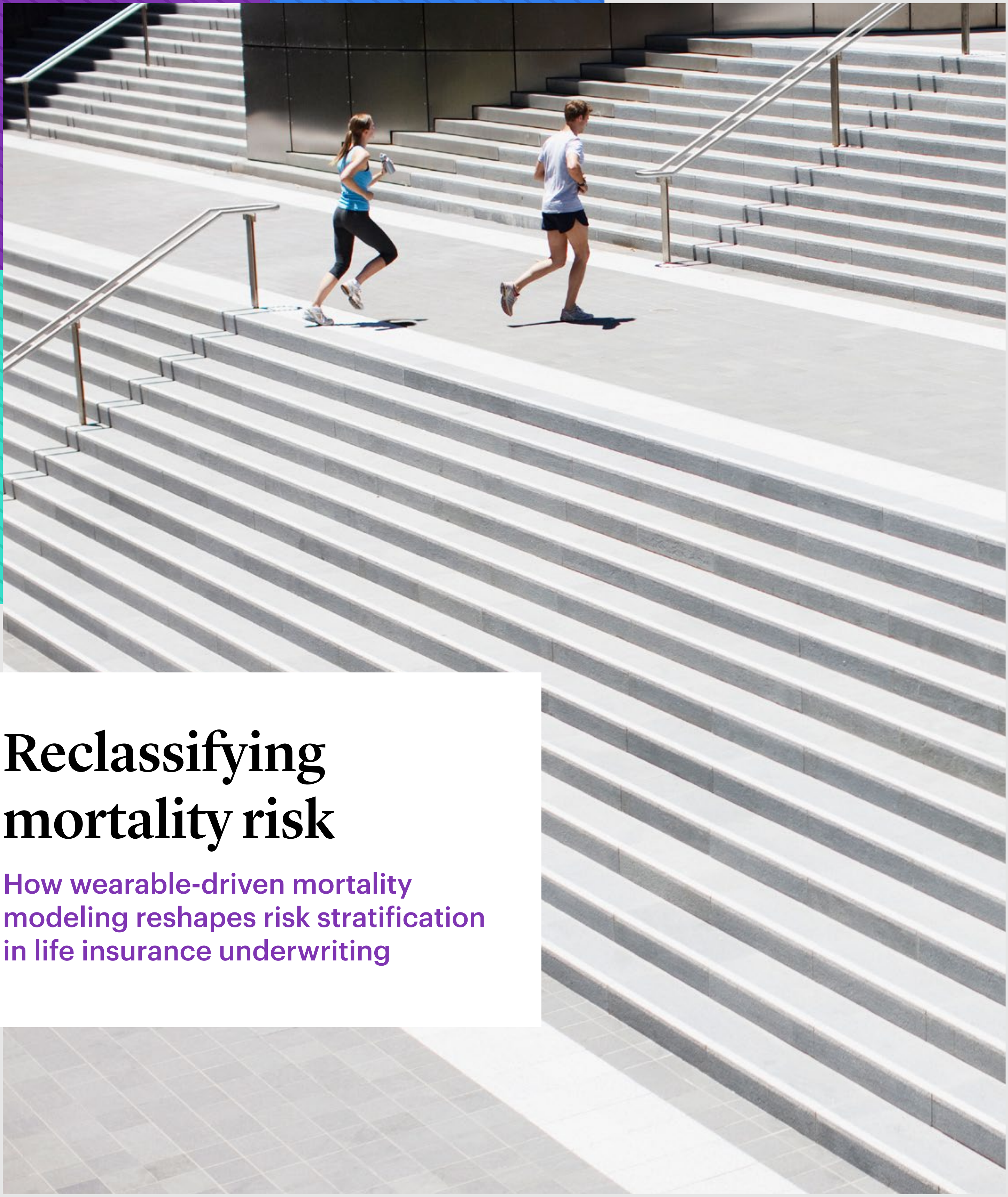


Reclassifying mortality risk

How wearable-driven mortality modeling reshapes risk stratification in life insurance underwriting

CONTENTS

Section 1:
Executive summary

“Sitting is the new smoking” is a catchphrase often used to encourage people to get some level of physical activity.

Medical personnel, underwriters, actuaries and mortality researchers understand activity level is an important measure to assess one’s health and expected longevity. Unfortunately, activity level information is often overlooked or has been measured only through self-reporting or correlation to other measures such as body mass index (BMI) in the life insurance risk selection process.

Since the proliferation of multiple risk classes, companies have used traditional measures such as cholesterol level, blood pressure, BMI, tobacco usage, and personal and family history, to name a few, for stratifying and determining risk class criterion and placement for applicants. While each is an important health metric, these traditional approaches and metrics often misclass applicants because the measures only provide part of an individual’s health profile and often miss important individualized measures such as resting heart rate, heart recovery rate, sleep and activity versus inactivity levels.

Over the past 10-plus years, the industry has been moving toward changing the underwriting process. For life insurance, this has meant rethinking the data sources used, improving the customer’s experience and shortening the time from application to policy issue. This has added to the challenges for risk classification and difficulty in truly differentiating the risk profiles of the preferred risks as well as the healthier impaired.

The need to rethink the risk stratification process in the life insurance industry has become increasingly evident over the past decade. With the proliferation of new data sources and advancements in technology, there is a significant opportunity to enhance the accuracy and efficiency of underwriting processes. The Klarity model aims to address this need by leveraging nontraditional data to produce individual-level mortality scores that can predict and classify risks more effectively than traditional methods.

Over the past year, WTW's Insurance Consulting and Technology (ICT) team has analyzed a new risk scoring tool developed by Klarity, which incorporates data obtained from a wearable device such as a fitness watch, a smartphone or other device that captures activity levels, sleep patterns, heart rate and pulse data. The model, originally trained on U.K. national health data, was built on data spanning over 12 years, covering 6.1 million life years, focused on ages 40 to 70. Within this data set, there are over 37,000 deaths.
To analyze the model and its applicability for life insurance, WTW partnered with Klarity to test the efficacy of the model’s mortality score predictions on data from the U.S., leveraging the National Health and Nutrition Examination Survey (NHANES) data set.

Key observations of WTW's analysis show:

- The Klarity model has the ability to better **classify risks** and reduce the overlap inherent in today's risk classification systems. In our analysis, when activity level and the Klarity model risk score are considered:
  - 34% of the second-best nonsmoker risks and 16% of the residual standard class risks are identified as having a better risk score and exhibit similar mortality risk profiles to the best nonsmoker risks
  - 6% of risks currently classified in the best and 13% classified in the second-best preferred nonsmoker classes are identified as having a lower risk score and exhibit actual-to-expected (A/E) ratios, more akin to a residual (not preferred) risk
- The level of activity, including Step Count and Activity Duration, **provide high correlation to mortality outcomes**, even more than some traditional markers such as cholesterol, BMI, and family history of heart disease and diabetes
- Though trained mostly on U.K. data, the Klarity model and model risk score are effective when applied to a U.S. population, and the model will **continue to get stronger** with increased U.S. insured data over time
- **Stratification improves** even among those who would be classified under the same risk classification using traditional underwriting methods
- There are many benefits to utilizing the Klarity risk scoring model, such as:
  - Improved stratification of risk and ability to improve the accuracy and assignment of risks to risk categories
  - Enhanced triaging or streamlining the triage process or risk class quoting to improve consumer and applicant expectations
  - Improved accuracy and selection of the healthier impaired risks
  - Reduced friction in the data collection process for measuring and analyzing mortality and longevity risk
  - In-force engagement
  - Potential for new products and pricing strategies focused on healthy longevity

The Klarity model demonstrates a promising approach to improving risk stratification in the life insurance industry. One of our key findings is that while the Klarity model validates the risk ranking of traditional underwriting classes used to assess mortality risk by insurers, the Klarity model can improve and further differentiate risks significantly even within a single risk class. We found the risk assignments by the Klarity model to be highly correlated with actual mortality results relative to mortality demographic baselines.

By utilizing a broader range of data sources and advanced predictive techniques, the model can provide more accurate and individualized risk assessments, ultimately leading to better pricing, improved customer engagement and enhanced in-force management.

The following documents WTW's analysis and findings; the white paper is structured as follows:

Background
This section provides an overview of the current state of the life insurance industry, highlighting the challenges and opportunities presented by new data sources and methodologies

Model
This section details the Klarity model, including its data sources, methodology and the results of its application to the U.S. NHANES data set

Use cases and next steps
This section explores potential use cases for the Klarity model and outlines the next steps for its implementation and further development

About Klarity
This section provides background information on Klarity, the company behind the model

About WTW
This section offers insights into WTW, the global consulting organization that collaborated on the review and testing of the model

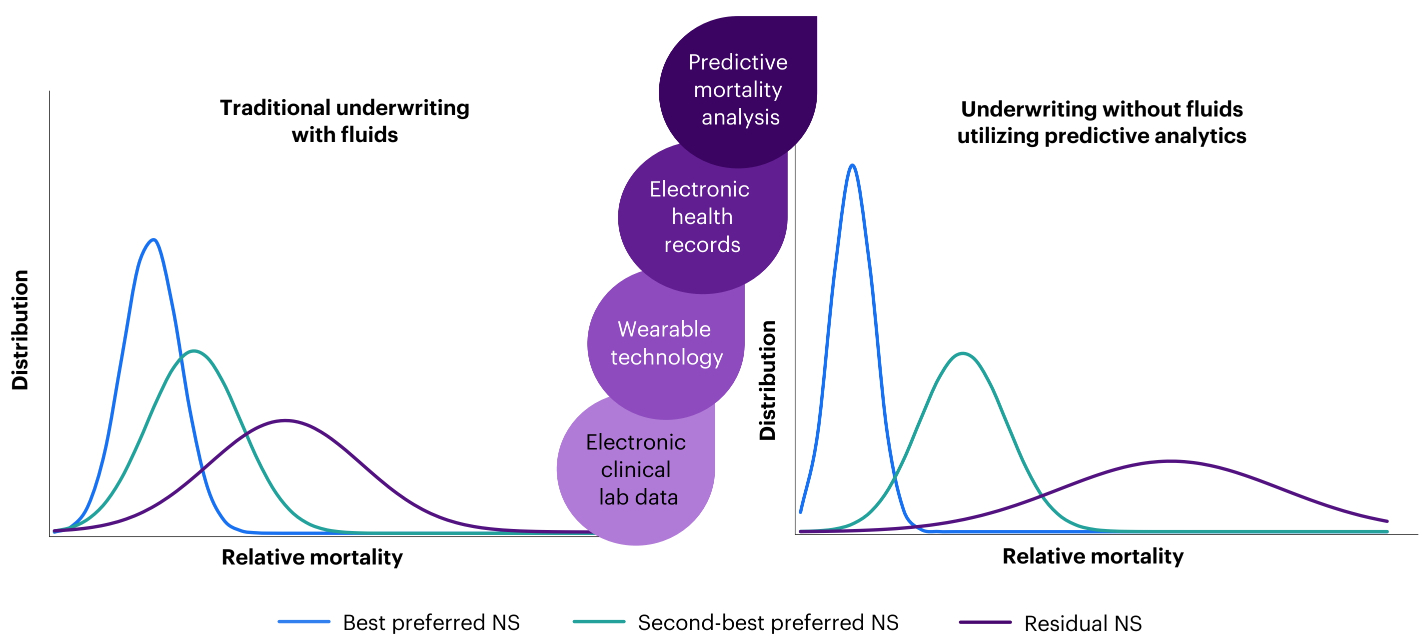
Section 2:
Background

Over the past few decades, the amount of data to help us better understand an individual's health and predict mortality has increased immensely, most notably over the past 10 to 15 years. The continued growth for insurtech companies continues to present new data and methodologies to improve the insurance industry.

Current state of the industry

The life insurance industry has incorporated and utilized new data sources but at a slower rate than other industries and other types of insurance. In addition to use of external data sources such as MIB and motor vehicle records, some of the largest movements in adding data for risk selection include credit risk attribute scores, pharmacy data records, medical billing data, clinical laboratory results data and electronic health record data — most of which are heavily leveraged for some variation of accelerated or automated underwriting. Companies have had vastly different experiences with accelerated and automated underwriting programs; some have observed only slight differences ("mortality slippage") between the mortality for policies underwritten using traditional criteria and data elements compared against policies incorporating more data and model driven criteria, while others have observed significant mortality slippage. Additionally, the process for accurate risk classification/stratification has been quite challenging for many companies, especially between the preferred risk classes that historically were differentiated by factors obtained through more invasive bodily fluid collection and measurements such as height, weight and blood pressure.

Figure. 1: **Enhanced risk stratification with new data** — This illustrative example shows how new digital data sources provide clearer differentiation of risk classes, enabling faster, less invasive and more accurate risk classification



Source: WTW analysis

More recently, a positive trend has emerged: the explosion of wearable technology has resulted in individuals increasing their awareness of and focus on being well informed of their individual health and wellness. Access to wearable data is now widely available via smart devices, including watches, bands, scales, rings, cell phones, computer applications and clothing. From 2017 to 2024, fitness tracking usage increased from 35 million users to 62 million in the U.S. and is projected to increase to over 92 million by 2029.[1] Along with the proliferation of devices, the friction in accessing this data has reduced significantly, and while a company's ability to directly leverage this data has some hurdles, we believe the increased focus and tracking of health-related data has created an opportunity for the industry both in terms of risk selection and ongoing consumer engagement. In addition to the availability of new data sources, the ease of availability and the quantity and quality of the data have continued to improve. The data has the potential to be leveraged to help companies improve their risk selection process, improve the alignment of risks with pricing, narrow the dispersion around the mean of a particular risk class and improve in-force management and engagement. The additional and better data can be used to partner with insureds instead of leveraged to gain a competitive advantage by providing a fair price for the risk and helping provide coverage for each individual's unique protection needs both at issue and throughout the individual's lifetime as needs change.

Using new data sources allows carriers to offer policyholders more individualized pricing and tailored coverage that evolves with their needs.

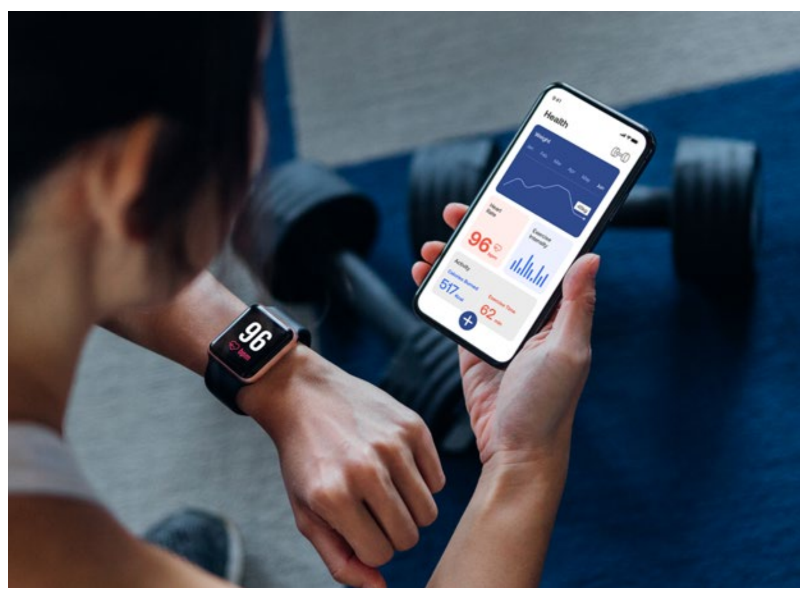
Though many companies have expressed interest in changing the paradigm of consumer engagement around aligned interest in better health management, this has often become a lower priority due to the other priorities such as new regulatory developments, other product profitability, complex product structures, and financial advisor or producer conflicts. If companies are willing to invest the effort at this time, the trend of individual's focusing on their health may prove to be a strong growth opportunity for both new business and customer engagement and loyalty, along with possibly new, innovative product designs.

Opportunities for improvement

One area WTW believes could be significantly improved by using additional data is in the risk class stratification process during underwriting. The current methodologies are capable of allocating a cohort of applicants into risk classes that directionally align with expectations. However, many measurements for the accuracy of the methods focus on the average of the overall cohort and fail to realize the materially different risks within the cohort, as the methodology for underwriting often has difficulty in only allocating the best individual risks into most preferred risk classes. While nearly all companies' current approaches are acceptable and reasonably align the risk class cohorts' mortality with pricing expectations, there is an opportunity to better allocate the risks into risk classes by leveraging activity and inactivity levels, sleep and health-related data from wearable devices.

Using new data sources allows carriers to better align risk stratification and expected mortality.

WTW has partnered with Klarity to review and test a risk scoring model that was initially built for the U.K. insurance industry. We believe this model can also be utilized in the North America life insurance market to help improve the risk stratification, reduce the dispersion around the mean for expected mortality of various risk classes, and create engagement programs for better in-force management and understanding of the health profile of an in-force cohort, which can lead to many other potential longer-term benefits.



1. Statista: www.statista.com/forecasts/1425172/fitness-tracker-users-by-segment-us

Section 3:
The Klarity Model

Overview

The Klarity model produces an individual level mortality score that can be used to predict and classify risks using limited or nontraditional data. The model, initially trained on U.K. population data and validated on U.S. data, determines risk scores, which can effectively stratify risk in a way that strongly correlates to actual mortality and demonstrates an ability to improve the stratification of risks over traditional underwriting methods using similar data inputs.

Data and methodology

Klarity's model is trained on data predominantly from the U.K. Biobank,[2] which is a prospective longitudinal study of de-identified individuals drawn from the U.K. population representative of the U.K. general population.[3] This data set contains 502,369 lives and 37,897 deaths from a population spanning enrollment mainly of ages 40 to 70 and observed an average period of 13.5 years to track health outcomes, including mortality. Data was too sparse outside of this age range to ensure a stable model. Individuals were recruited for this study between 2006 and 2010 and followed through the end of 2022. There are currently over 6.1 million life-years of data, after accounting for deaths, with collection still ongoing. This longitudinal data set also allows for analysis of cumulative exposures over life years, offering insights into the relationship between sustained risk factors and mortality outcomes.

WTW partnered with Klarity to test the efficacy of the model's mortality score predictions on data from the U.S. In an effort to leverage a large set of publicly available U.S. data, WTW tested the model's predictions against the NHANES[4] data set, which is a longitudinal survey run by the National Center for Health Statistics, part of the Centers for Disease Control and Prevention (CDC).[5] NHANES has run continuously since 1999 and includes approximately 5,000 participants each year, combining surveys with physical examinations. The survey uses a random, scientific process to select the individuals who are invited to participate. NHANES also contains mortality data on the testing population to allow for analysis of the actual mortality experience. Our testing data set combines each NHANES survey from 1999 to 2015. Mortality data is available for a subset of participants (those 18 and over at the time of the survey), where available. The identification of deaths for all participants were verified as of December 31, 2019, by the CDC,[6] which is the most recent available update on identifying deaths that have occurred. The death data provide the number of months completed from survey participation until death. The following table summarizes the participants within each two-year period for the 1999–2018 data:

Survey year	Participants	Deaths
1999–2000	9,965	1,675
2001–2002	11,039	1,624
2003–2004	10,122	1,420
2005–2006	10,348	1,027
2007–2008	10,149	1,126
2009–2010	10,537	861
2011–2012	9,756	628
2013–2014	10,175	467
2015–2016	9,971	276
2017–2018	9,254	145
Total	**101,316**	**9,249**

Source: Klarity analysis

The NHANES data was further filtered from the 101,316 lives to align with the U.K. data set and also to focus the study on a typical life insured population. As can be seen in the following table, the NHANES data were filtered for starting ages 40 to 70. Overall, the filtering resulted in a data set that contained 25,785 lives.

NHANES data set walkthrough	Change	Cumulative Total
Full data set (all NHANES cohorts)	**101,316**	**101,316**
Remove individuals under 18	*(42,112)*	59,204
Remove individuals missing mortality status	*(140)*	*59,064*
Remove individuals missing cholesterol data	*(2,811)*	*56,253*
Remove individuals under 40 and over 70	*(30,468)*	*25,785*
Data set used for testing	**25,785**	**25,785**

Source: Klarity analysis

2. Research conducted under Application Number 88308
3. Information about U.K. Biobank can be found at www.ukbiobank.ac.uk/enable-your-research/about-our-data
4. Note this is a distinct study from the NHANES Epidemiologic Follow-up Study (NHEFS)
5. Information about NHANES can be found at www.cdc.gov/nchs/nhanes/about/index.html
6. CDC source: www.cdc.gov/nchs/data-linkage/mortality-public.htm

The following tables provide a summary of the demographic details of the data set that was used to test the Klarity model on the U.S. data set.

Age at observation period	Participants	Deaths
40–44	4,506	203
45–49	4,234	264
50–54	4,277	366
55–59	3,519	442
60–64	4,827	855
65–70	4,422	1,185
Total	**25,785**	**3,315**

Source: Klarity analysis

The lives were spread fairly evenly across the decennial age bands.

Attained age	Deaths
40–44	28
45–49	95
50–54	218
55–59	312
60–64	440
65–69	610
70–74	738
75–79	497
80+	377
Total	**3,315**

Source: Klarity analysis

The WTW study included slightly more females than males. WTW analyzed results on an aggregate basis.

Gender	Participants	Deaths
Male	12,591	1,953
Female	13,194	1,362
Total	**25,785**	**3,315**

Source: Klarity analysis

Self-reported smokers were just over 20% of the reduced data set used in the WTW study. The definition was set to be aligned with the CDC definition of "smoking ≥100 cigarettes during a lifetime and now smoking cigarettes either every day or some days."

Smoking status	Participants	Deaths
Smoker	5,780	1,181
Nonsmoker	20,005	2,134
Total	**25,785**	**3,315**

Source: Klarity analysis

Given that the Klarity model has been trained on U.K. data, translating the data to address the U.S. market creates a potential disconnect. For our testing, we leveraged the Klarity model to classify relative risks — one of several potential applications of the model — rather than to determine the absolute level of mortality risk, and thus the differences in overall mortality between the two countries should not have a sizable effect. As Klarity gains usage in the U.S. market, additional training of the model with U.S. data and insured lives data will further expand its potential value in the U.S. For our testing of the Klarity model, the 2019 population mortality table published by the Social Security Administration (from the 2022 Trustees Report[7]) was chosen as an expected basis to compare against actual mortality in NHANES. The 2019 population mortality table was chosen due to the lack of insured risks and underwriting within the NHANES data set, and it most closely reflects the latest pre-COVID population mortality. It is worth noting that while lifestyle choices and demographics may differ between the two countries, we expect the physiological associations with mortality risk captured by the Klarity model to remain relatively stable.

The inputs from the NHANES data leveraged by the Klarity model include variables for age, sex, smoker status, BMI, cholesterol, statin usage, blood pressure, alcohol use, personal history of disease, family history of disease and wearable data such as daily activity, step count, sleep duration and inactive time. In some instances, the NHANES data were not complete. For this analysis, we compared model results using both imputed data and by allowing the model to deal with the missing data on its own. Upon encountering a missing feature value, the Klarity model does not discard the data point. Rather, it has learned through the training how to best handle missing values by assigning them to the path that results in the lowest prediction error. This ensures the model can still generate accurate predictions from incomplete data. The model minimizes a regularized objective function that penalizes prediction error while also controlling model complexity, to reduce overfitting. Thus, the predictive metrics from the model were stronger when letting the model handle missing data. The following summarizes the key fields where imputation was required due to the underlying NHANES data having missing values:

- **Step count:** Available in only the 2005-2006 NHANES study cohort, leading to 92% missing values and gaps when analyzing multiple cohorts
- **Activity duration**: 74% missing data for individuals ages 40 to 70 due to skipped questions or incomplete tests
- **Drug coverage:** Some statins missing in specific cohorts (e.g., BEZAFIBRATE only in one cohort; FLUVASTATIN absent in 2011, 2013, 2015; ROSUVASTATIN absent in 1999, 2001)

7. Source: www.ssa.gov/oact/STATS/table4c6_2019_TR2022.html

The Klarity model is a classification model, which predicts whether the person will be alive or dead at the end of the time horizon.

As a starting point, our testing focused on a 10-year horizon, which was chosen to balance the long potential duration of a life insurance contract with the importance of selecting risks for favorable short-term mortality risk. Any deaths in the training data set occurring after the time horizon are not used by the model in predicting deaths. Similarly, when testing against the NHANES data, deaths occurring after the time horizon from the initial observation of model characteristics or predictor variables were removed. During testing, Klarity trained its model on a five-year horizon to allow for analysis of using different horizons. For these tests, we observed that the model performed comparably although shorter horizons exhibited more volatility, which is expected given the shorter time period. We believe that predicting early duration death claims has been challenging for many companies in the U.S., and thus an additional tool for better predicting likelihood of early death claims would still be useful. The remainder of this report focuses on the analysis using the 10-year horizon.

Mortality scoring

The model outputs a score for each risk ranging between zero and one. This score is compared against scores from other risks with the same age, sex and smoking status characteristics from the training data and transformed using a Box-Cox transformation (a statistical technique used to transform non-normal data to resemble a normal distribution) to normalize the results. The Box-Cox transformation method was chosen because it produced a better result than logarithmic transformation. This allows the model to assign a relative risk for a person given their age, sex and smoking status. The relative risk is given in standard deviations from the expectation, which was based on the results from the U.K. model, for that risk based on age, sex and smoking status. Relative risks were used over absolute risks based on the focus of determining whether the model could select the best mortality risks within a given cohort.

In addition to the mortality scoring, the model outputs a SHAP (Shapley Additive Explanations) value for each predictor for that risk.

The SHAP value (or "score") is commonly used to evaluate machine learning models and indicates the direction and size that the given predictor is contributing to the overall risk prediction for an individual, in terms of a distance from the mean prediction. A positive SHAP score indicates that the variable is increasing the total mortality prediction for that individual. This provides the model with the ability to explain the key contributing factors to its assignments of each risk.

How does the model compare with current preferred guideline risk assignment?

We obtained four publicly available preferred criteria guidelines from popular life insurance companies. We assessed these guidelines for reasonableness relative to our understanding of guidelines generally in place by carriers today. Additionally, we used a set of guidelines provided by the Society of Actuaries (SOA) that were contemporary to the NHANES data set to compare the ability of the Klarity model to stratify the risks versus more traditional underwriting guidelines used by insurers.

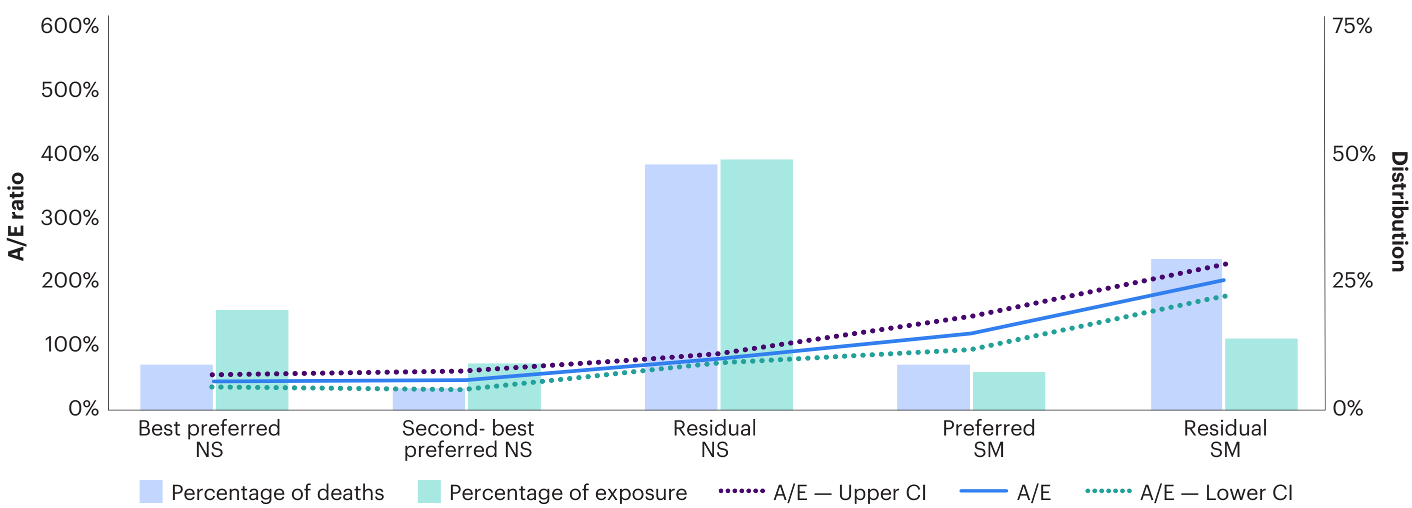
Some of the more common preferred criteria utilized in the U.S. include BMI, blood pressure, cholesterol, smoking/tobacco history, medical history, family history and driving history. Since not all the data features were available for the testing NHANES data set, some professional judgment was utilized in classifying the individuals into specific risk classes. For example, we utilized the NHANES question focused on if the person was ever diagnosed with cancer whenever company guidelines referred to a diagnosis within N years.

Figures 2 and 3 below show the results of our analysis for two of the five sets of preferred underwriting criteria. These results were fairly representative of the remaining three which we modeled. The analyses confirm that the application of the guidelines produced a reasonable stratification of risks when applying the publicly available traditional preferred underwriting guidelines to the NHANES data set. The line graph shows the A/E ratio with 95% confidence interval (CI) using the 2019 U.S. population mortality table as the expected basis. The bar graphs show the percentage of exposures (life-years) and deaths in each underwriting class.

In *Figure 2*, company 1 shows a strong level of overall stratification; however, there is little distinction between best preferred nonsmoker (NS) and second-best preferred NS, as well as residual NS and preferred smoker (SM).

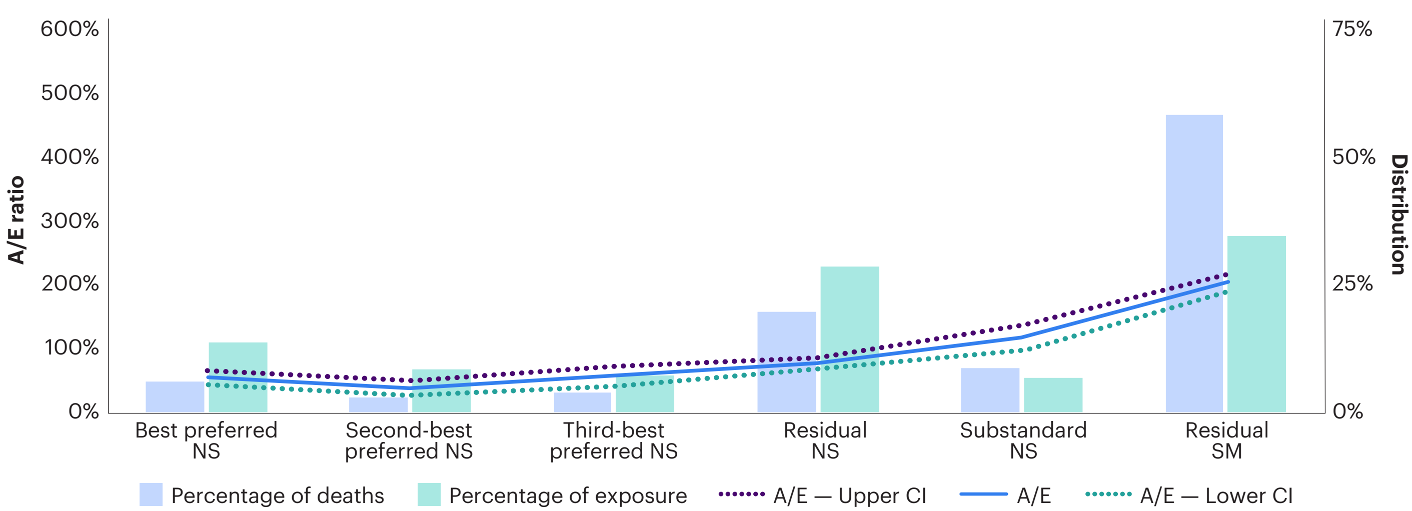
In *Figure 3*, company 2 shows a more complex underwriting structure. The trend aligns with expectations in that the better classes have lower mortality risk, with the exception of Best Preferred NS. This is evidence traditional underwriting may struggle with differentiating preferred risks. We believe that the Klarity model, leveraging activity and inactivity data is a tool to improve accuracy in risk class assignment and differentiation.

Figure. 2: **Life insurer's underwriting guidelines — Company 1**



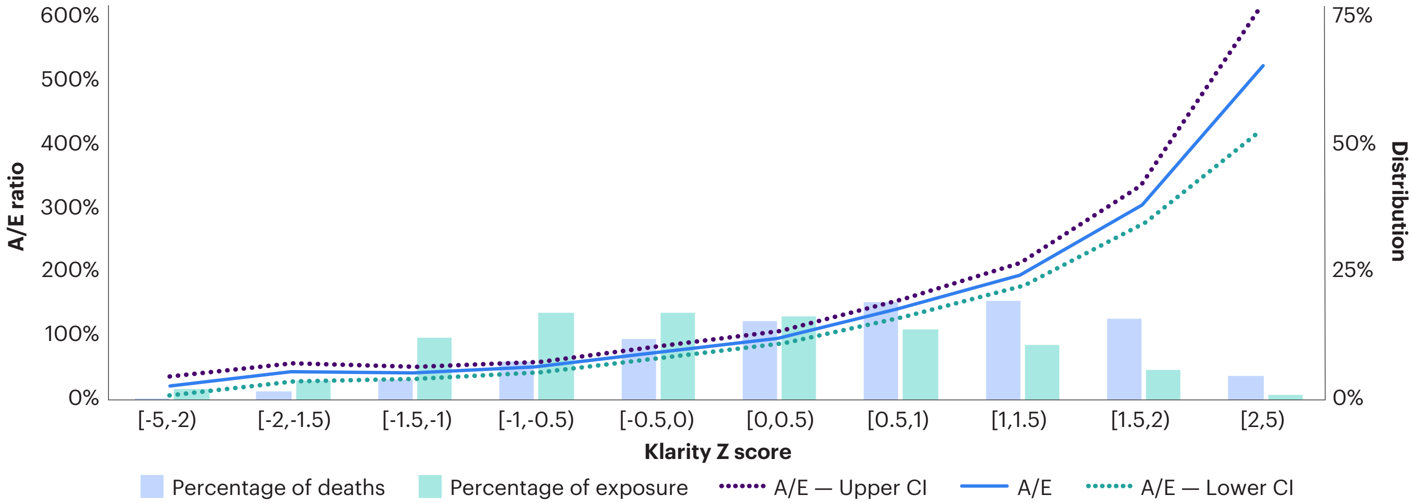
Source: WTW analysis

Figure. 3: **Life insurer's underwriting guidelines — Company 2**



Source: WTW analysis

Figure. 4: **Klarity 10-year horizon model**



Source: WTW analysis

Figure 4 replaces the traditional class assignments with mortality prediction scores from the Klarity model. As can be seen above, the mortality increases as the score increases, including identifying some individuals which are significant mortality risks.

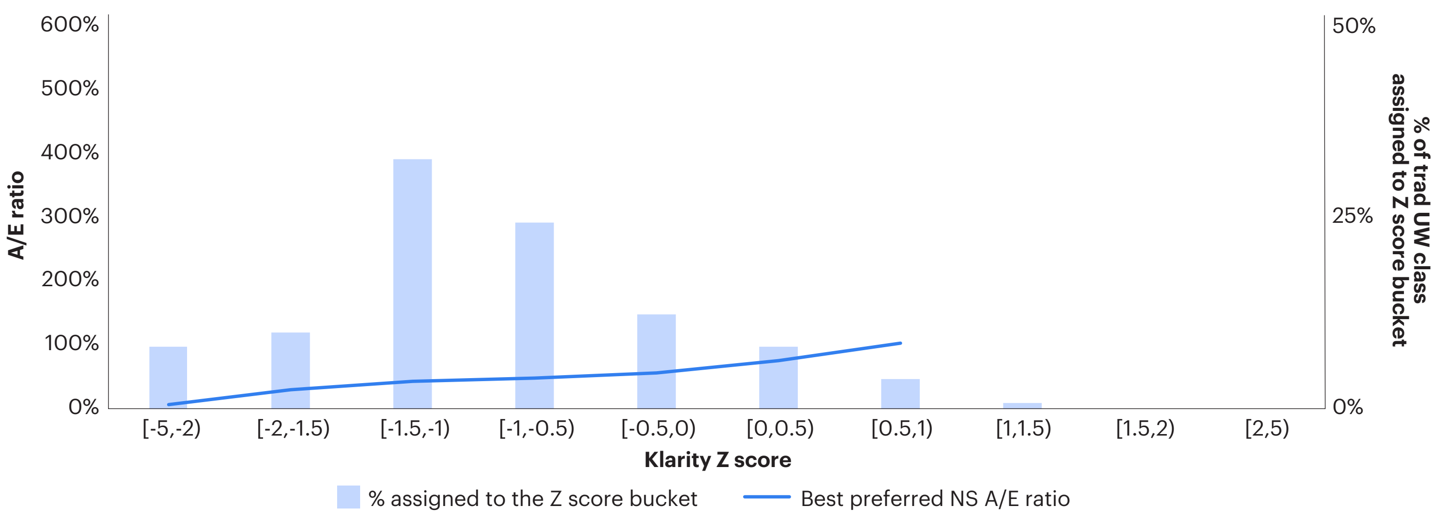
Risk scores of 0.5 and higher represent one-third of the exposure and have a relative A/E ratio ranging from 128% (14% of the exposure) to over 300% (7% of the exposure). These higher scores represent 57% of the deaths.

The following three graphs (*Figures 5, 6* and *7)* demonstrate the Klarity model's ability to better classify risks, even among those who would be classified under the same risk classification using traditional underwriting. In *Figure 5* below, we show how the best preferred class under Company 1's underwriting criteria would be reassigned using the Klarity model. The blue bars show the percentage of the exposure that fall into each range of Klarity scores, while the line graph shows the A/E ratio among each group of Klarity risk scores. The line graph is not shown for the higher positive Klarity Z score buckets due to too small a number of deaths to have credibility.

This highlights two things:

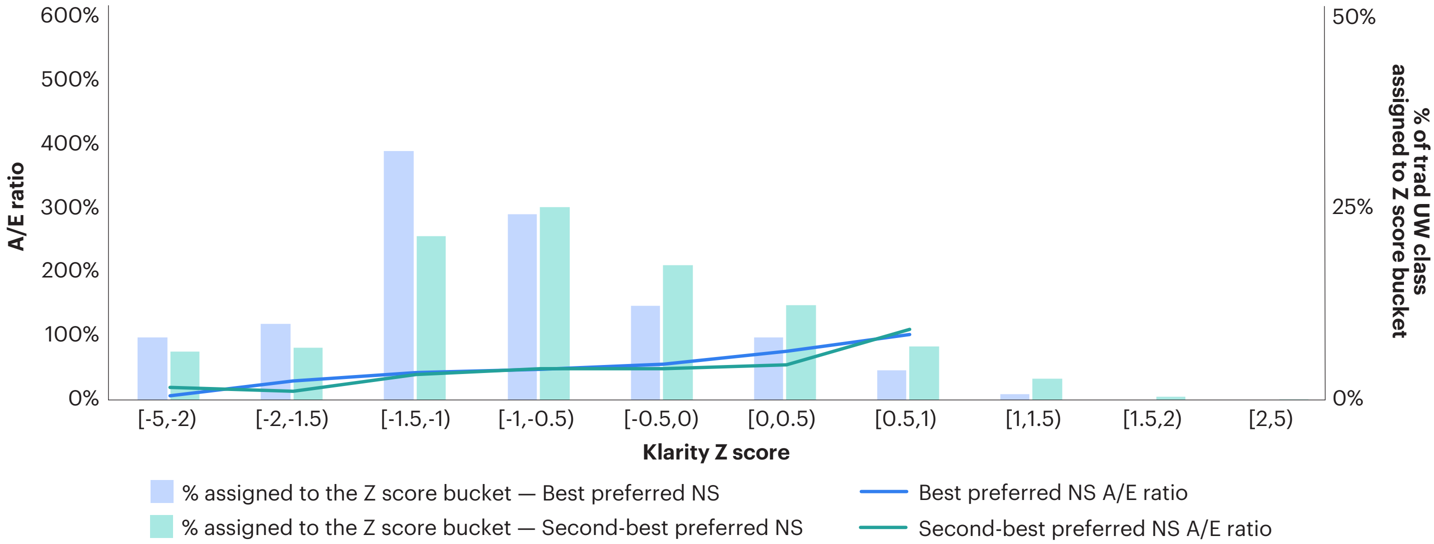
1. The Klarity model is in agreement that the majority of these risks currently classified as best preferred exhibit more favorable mortality relative to others in the same age, gender, smoking status category.
2. Greater segmentation and more accurate risk assignment is possible, as demonstrated by the clear gradation in the A/E ratios by risk score. Among these lives, about 5% were assigned risk scores representing over 100% A/E ratio from the Klarity model, and it demonstrates a right-tailed distribution of risks. These higher risks tended to be driven by high blood pressure and personal history of disease according to the SHAP values.

Figure. 5: **Klarity score movement from Company 1 UW — Best preferred nonsmoker**



Source: WTW analysis

Figure. 6: **Klarity score movement from Company 1 UW — Best and second-best preferred nonsmoker**



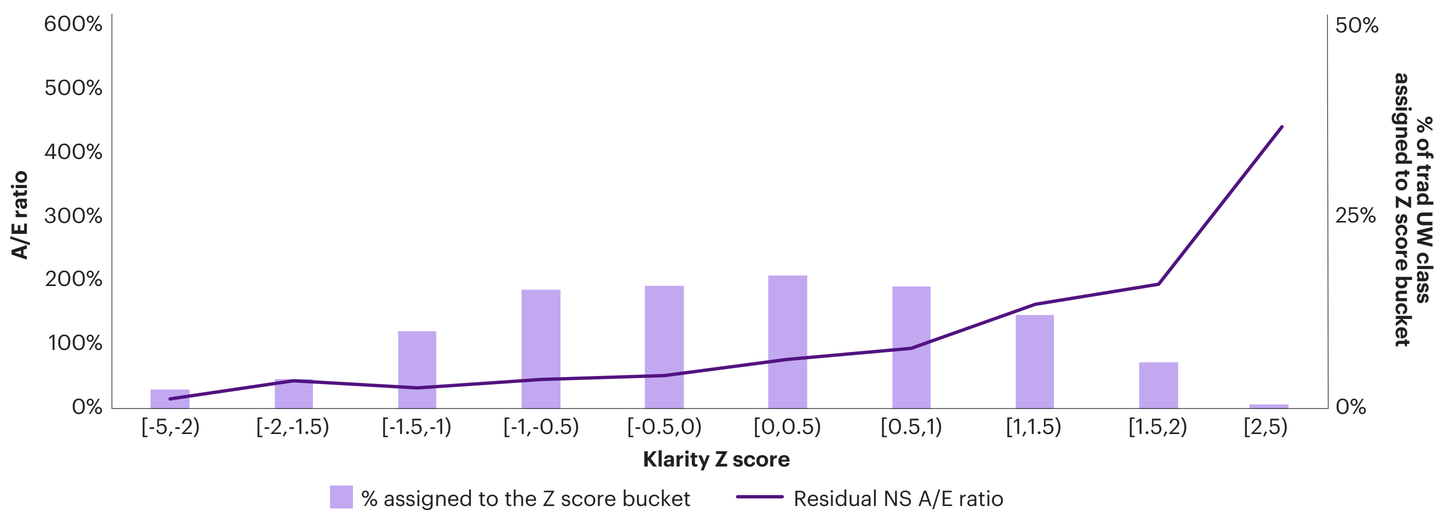
Source: WTW analysis

Figure 6 layers the second-best preferred nonsmoker class on top of the best preferred nonsmoker class from the previous graph. There is some differentiation as the best preferred class contains more risks grouped in lower risk groups and the second best preferred class contains more risks in the higher mortality risk groups, but ultimately these classes overlap significantly, and the A/Es are very similar. This is an area where traditional underwriting has room for improvement.

As another example, we present the Klarity reassignment of non-preferred nonsmokers under Company 1's underwriting criteria (*Figure 7*, below).

As expected, this group of lives is assigned more average mortality under the Klarity model compared with the best preferred class above. However, among these, **roughly 17% of lives are assigned to risk scores below one standard deviation relative to their demographic baseline Z score, meaning they likely are a better risk than standard**. The main drivers of these assignments according to SHAP values are activity duration, inactive time and step count. The A/E graph again highlights that the Klarity model can stratify risks, leading to a better understanding of risk and the possibility of making better decisions given this knowledge.

Figure. 7: **Klarity score movement from Company 1 UW — Residual nonsmoker**



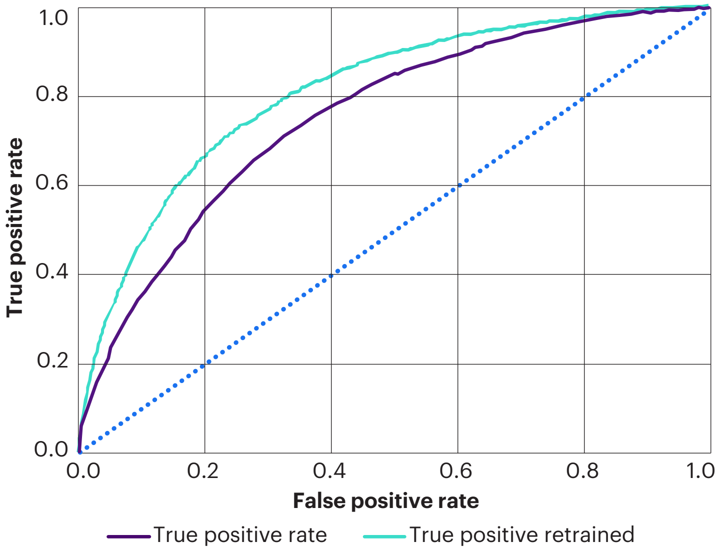
Source: WTW analysis

As a further step in training the Klarity model for use in the U.S., Klarity retrained the model combining 20% of NHANES data into the training data. The updated model was tested on the remaining 80% of the NHANES data. The retrained model provides further stratification of risks with improved differentiation of some of the best risks (i.e., low scores).

The model's predictive statistics are notably improved as can be seen in *Figure 8*. The area under the curve (AUC) refers to the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate[8] in the y-axis against the false positive rate[9] on the x-axis for every threshold[10] between 0 and 1. AUC can be thought of as measuring how well the model can distinguish between deaths and survivals. A higher AUC indicates a model is better able to classify observations into classes or the various thresholds.

These improvements are a result of only including 20% of the NHANES data. We expect continued improvement in the results when the model is trained on the entire NHANES data, which would be a starting point to be leveraged by companies in the U.S. Additionally, the models could be further enhanced by using actual U.S. life insurance data — potentially during pilot testing programs. This opens the potential for even greater gains in accurately allocating risks across different risk buckets.

Figure. 8: **ROC curve for model 1** — Even the U.K.-only trained model shows a well-balanced ROC curve when tested on U.S. data with high AUC



Source: Klarity analysis

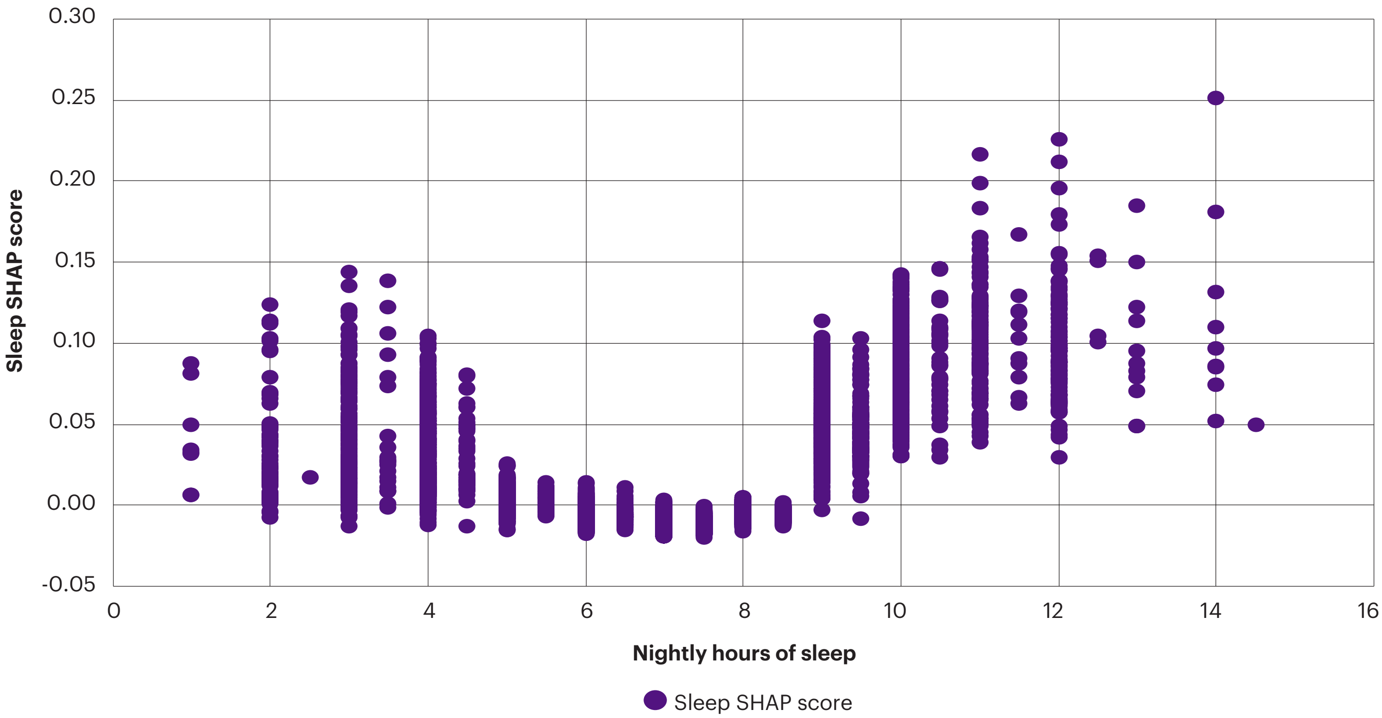
Model	Training data set	Test data set	AUC
Model 1: U.K. model	90% U.K. data	100% of NHANES	75.6%
Model 2: U.K. model retrained with NHANES	90% U.K. data + 20% NHANES data	80% of NHANES	78.8%

8. Defined as the number of positives (deaths) correctly identified by the model divided by the total number of positives.
9. Defined as the number of positives (deaths) incorrectly identified by the model divided by the total number of negatives (survivors).
10. To make a classification between death or survival, the model compares the predicted score for each individual with the threshold. If the prediction is higher than the threshold, the model classifies that as a death, otherwise a survival.

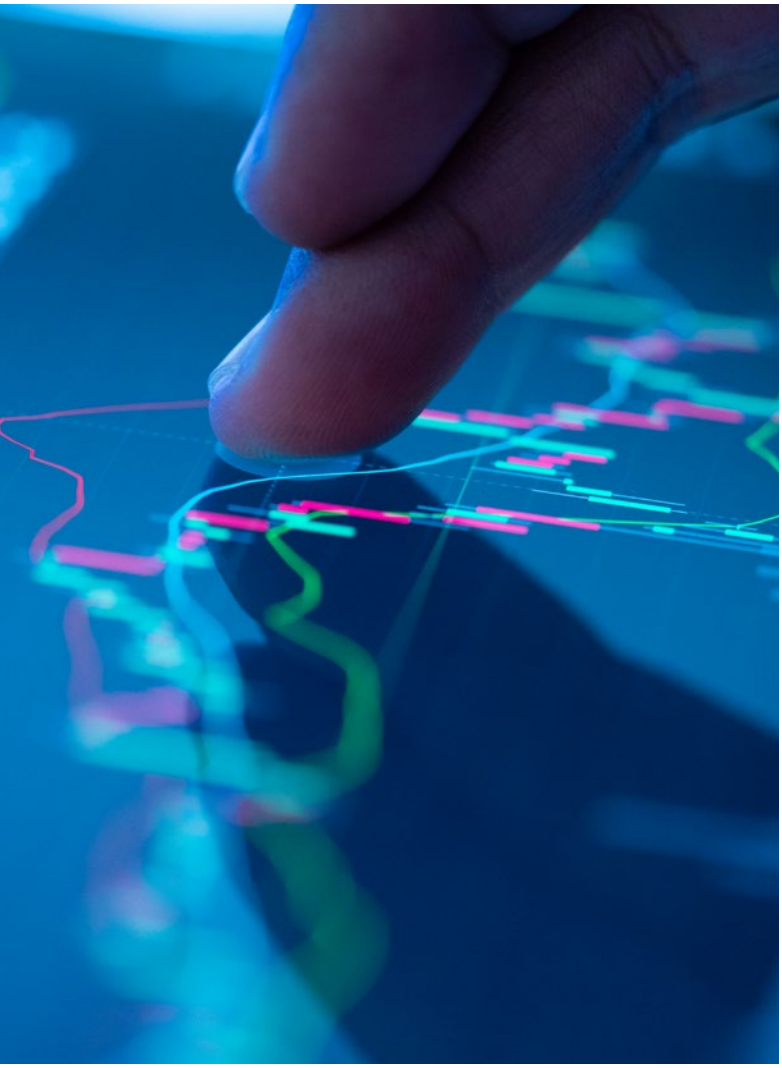
We also reviewed a model that used hyperparameter optimization. The results appeared to improve the AUC, but this model was not ready in time to complete our in-depth analysis. This does, however, suggest the possibility for continued refinements and improvements in the accuracy of the model.

The SHAP graph below (*Figure 9*), demonstrates how the model can provide insight regarding how it takes each factor into account when making a mortality prediction. Having this capability allows the model to be able to explain why a given risk received its risk score as well as demonstrate general relationships between the variable and mortality. For example, in the graph below, the model assigns the lowest risk to nightly hours of sleep between 7 and 8.5. There is not much variation in score in this range. On the other hand, for someone who gets 10–12 hours of sleep per night, the model has both higher SHAP scores and high variability in SHAP scores. This indicates that the model identifies 10–12 hours of sleep as generally associated with higher mortality risk but that this can be significantly impacted by other factors. While SHAP values do not directly capture specific combinations of variables, their results can be compared to identify the most influential factors in mortality prediction. The Klarity model can determine the relative importance of each input variable at an individual level and rank them accordingly.

Figure. 9: **Sleep SHAP Score**



Source: WTW analysis



Benefit gained from activity data

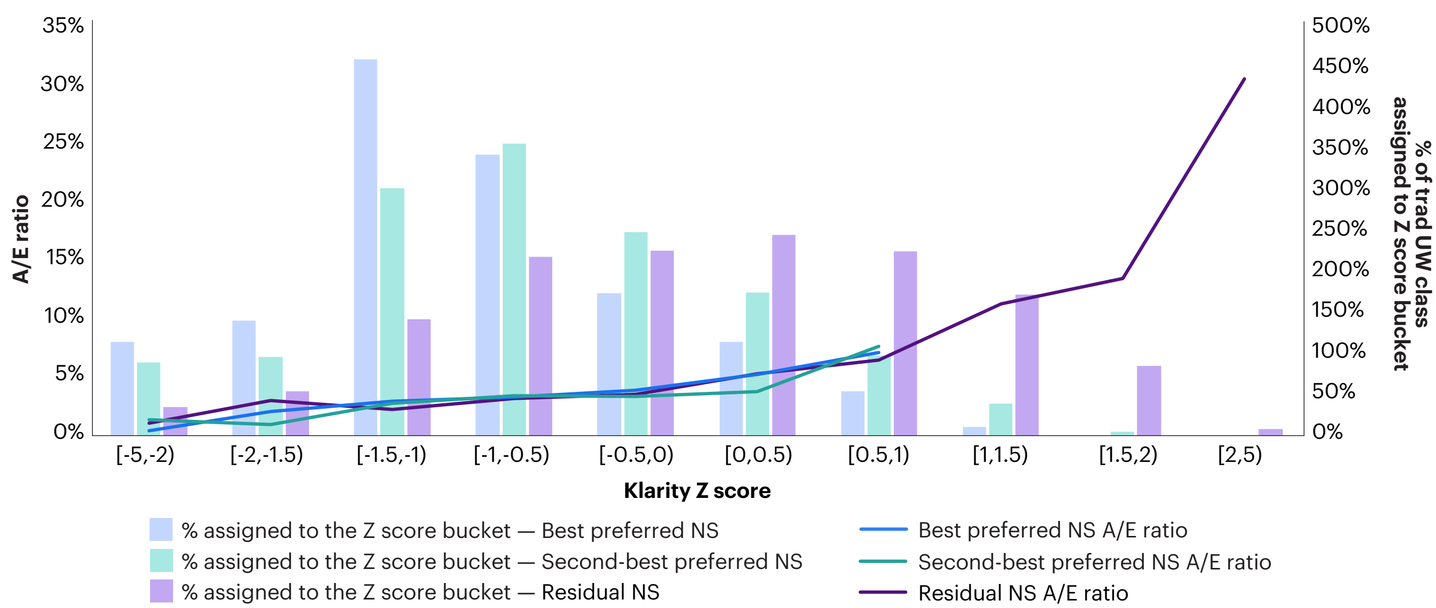
As mentioned above, the Klarity model includes activity data from NHANES as a component of the risk classification. A portion of the improvement in risk stratification observed with the Klarity model can be attributed to the use of activity data. To quantify the improvement gained by including activity data, we ran the Klarity model without any activity data in the data set by removing items such as sleep duration, active/inactive time, step count and average heart rate (model 3).

Model	Training data set	Test data set	Activity data	AUC
Model 1: U.K. model	90% U.K. data	100% of NHANES	Included	75.6%
Model 3: U.K. model with no activity data	90% U.K. data	100% of NHANES	Excluded	74.0%

As can be seen in *Figures 10* and *11*, the Klarity model provides a material improvement to risk stratification even in the absence of activity information.

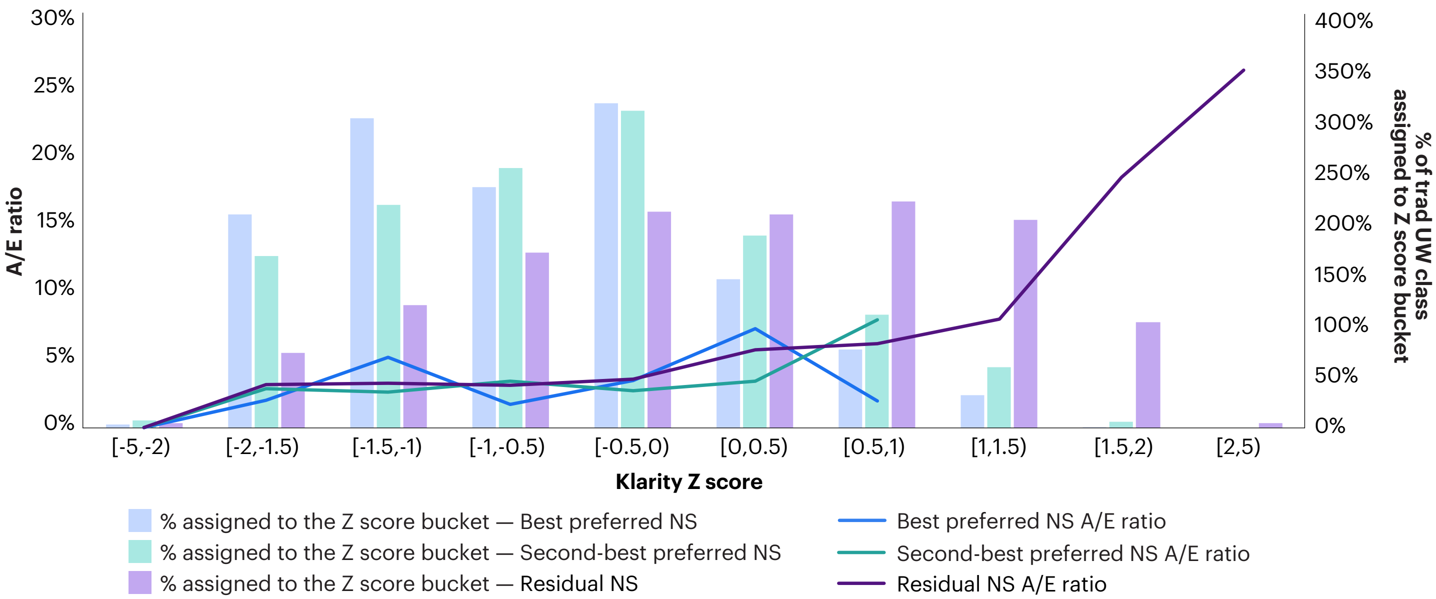
The Klarity model demonstrates a significant improvement in risk stratification, even when activity data are unavailable, showcasing its robustness and adaptability to varying data conditions.

Figure. 10: **Klarity score movement from Company 1 UW — Nontobacco, with activity data**



Source: WTW analysis

Figure. 11: **Klarity score movement from Company 1 UW — Nontobacco, no activity data**



Source: WTW analysis

The analysis in *Figure 11* clearly demonstrates challenges in correctly classifying the risks with traditional data tools and the significant overlap of risks among the risk classes. In this model, a Klarity Z score below –1 includes 50% of the best preferred nonsmoker risks from traditional underwriting. The inclusion of activity data also suggests **34% of the second-best nonsmoker risks and 16% of the residual standard class risks could also qualify based on their mortality risk profile when activity level is considered. Similarly, risks currently classified as best or second-best preferred NS exhibit A/E ratios in excess of 105%, more akin to a residual (not preferred) risk.** This suggests opportunity to better group risks to more clearly differentiate the mortality in each of the risk class buckets and reduce the mortality slippage from misclassification. The Klarity model can be calibrated to different Z score buckets to customize a company's targeted risk grouping and expected mortality levels.

The Klarity model offers advancements over traditional mortality underwriting guidelines by leveraging predictive analytics and nontraditional data sources.

While traditional methods rely on broad classifications using mostly static variables such as BMI, blood pressure and smoking status, the Klarity model incorporates richer data sets, including wearable device metrics, personal and family health histories, and lifestyle factors. This allows for a more granular and accurate stratification of risks, reducing mortality dispersion within risk classes and enabling more individualized pricing. Additionally, the model's use of machine learning improves predictive accuracy by detecting complex relationships between variables.

Another key differentiator of the Klarity model is its transparency through SHAP values, which provide clear insights into how individual risk scores are determined.

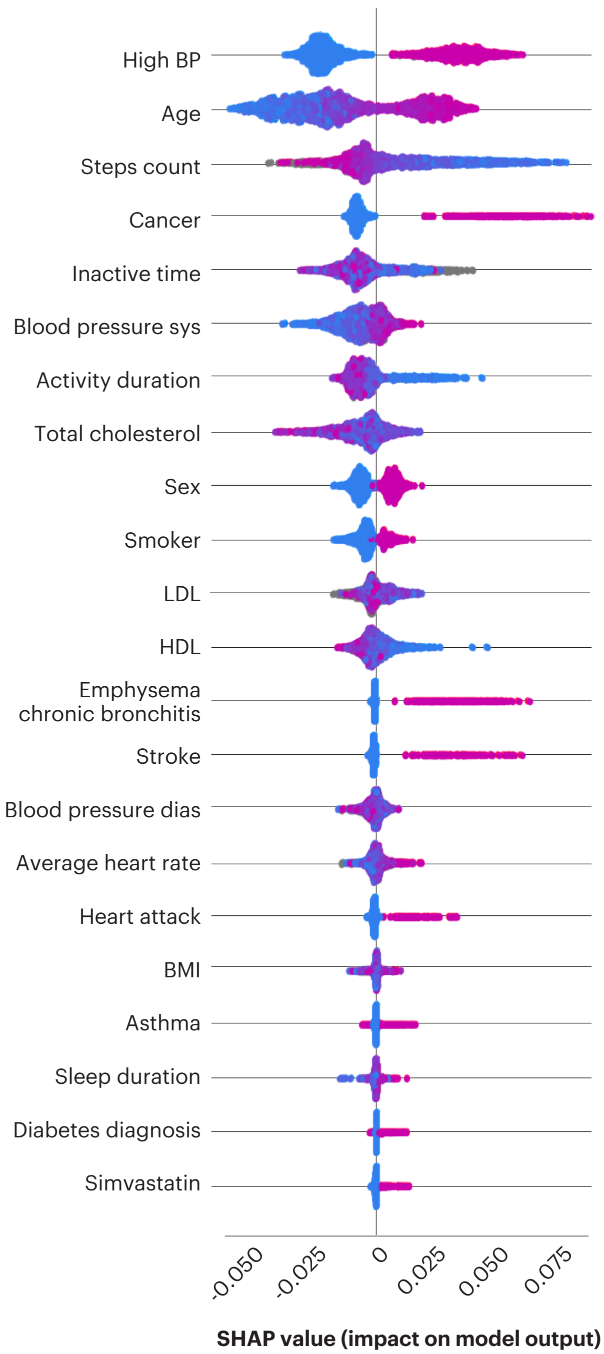
This level of detail improves the explainability with policyholders and regulators but also highlights the model's ability to adapt across markets by retraining on localized data. Together, these features enable insurers to refine underwriting practices and deliver more tailored, commensurate outcomes for customers.

Figure 13 on the following page shows the SHAP scores for a model trained on only data that included step counts. This model was tested on the 2005 NHANES cohort only due to it being the only cohort that has step count data. Each observation in the data has a SHAP value for each variable, which is plotted as a histogram. The coloration provides further information as to whether the underlying feature has a high or low value for that observation. For example, by looking at the age plot, we can tell that age is an important factor due to having a wide range of SHAP values and that lower ages tend to result in lower SHAP values while higher ages tend to result in higher SHAP values. Plots with significant overlap in the pink and blue values indicate that the model considers the relationships with other inputs to play an important role in determining that variable's impact on mortality for that person.

Level of activity, including step count and activity duration, provides high correlation to mortality outcomes, even more than some traditional markers such as cholesterol, BMI, and family history of heart disease and diabetes.

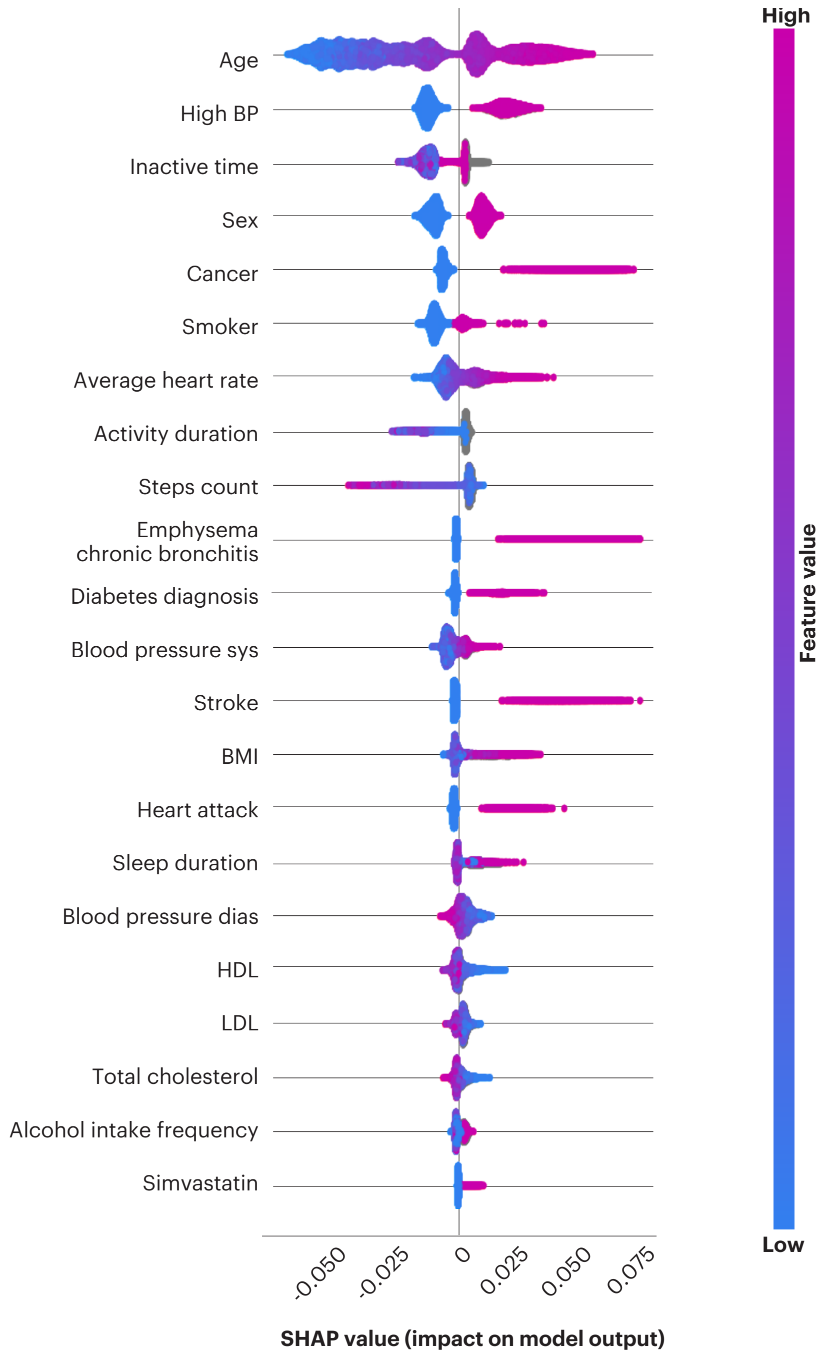
Figure 12 demonstrates that when the model has access to step counts for each person, it is the third most influential predictor in stratifying risk scores. In *Figure 13*, this predictor drops in importance ranking due to most observations being missing. Both charts agree that age, high blood pressure, inactive time and cancer are some of the most important factors.

Figure. 12: **SHAP chart for 2005 NHANES cohort**



Source: Klarity analysis

Figure. 13: **SHAP chart for all NHANES data**



Source: Klarity analysis

Section 4:

Use cases and next steps

The Klarity model presents numerous potential use cases that can significantly enhance the efficiency and accuracy of the underwriting process and ultimately enhance in-force engagement and new product opportunities.

Improved stratification of risk

One of the primary applications is improving the use of preferred underwriting. By utilizing a mortality risk score, companies can better align individual applicants into the risk classes that best fit with their health profiles. This will likely allow for companies to employ more granular preferred classification, and the additional granularity ensures that individuals are more accurately assessed and placed into appropriate risk categories, leading to more accurate pricing and improved risk management. Additionally, the Klarity model can enhance the accuracy of underwriting within existing class structures given that we have observed that it will classify risks differently than a traditional underwriting structure. We have observed that the Klarity model greatly improves the stratification of A/E ratios across various bucketing of their mortality risk score.

Enhanced triaging of risk

Another key use case is in the triaging of risks for various underwriting methods. The Klarity model can streamline this process by determining the most suitable underwriting method for each applicant. Higher risk policies can be designated for full underwriting, while lower risk policies may be assigned for automated underwriting. This triaging capability not only speeds up the underwriting process but also ensures that underwriting resources are allocated efficiently, and it allows for a human to be in the loop for reviewing the more complex cases for risk assessment.

Clearer visibility for substandard risks

The model can aid in determining more accurate pricing of substandard risks, such as individuals with diabetes, by identifying the healthier better-controlled cases within this group. This targeted approach can improve the overall health profile of the insured population, reduce underwriting expense and optimize marketing strategies.

Device proliferation and reduced friction in data collection leading to broader and more accurate analytics

The use of wearable device information has been of interest but slow to gain traction for underwriting of life insurance, even though insurers are beginning to understand the potential value of the data coupled with increased ease in accessing the data as wearable use becomes more common among Americans. Adoption of wearable devices has continued to increase, with one in three Americans now using them to track activity.[11] However, roughly nine in 10 Americans now have a smartphone capable of capturing basic activity data such as steps and exercise minutes.[12] Wearables provide access to several variables that have important correlations with mortality risk, such as daily active/inactive time, sleep duration, step counts and heart rate. The data from wearables provide a history of data rather than point-in-time measures. Additionally, the data from wearables can be synced up into doctors' records to allow for more accurate data to be used during visits as well as real-time monitoring in certain circumstances, such as when vitals get to a concerning level.

11. Source: www.nhlbi.nih.gov/news/2023/study-reveals-wearable-device-trends-among-us-adults
12. Source: www.pewresearch.org/internet/fact-sheet/mobile/

In-force management, engagement and new product design opportunities promoting healthy longevity

In the realm of in-force management and engagement, the Klarity model can use data from wearables to identify early signs of potential diseases, enabling proactive health interventions and better management of existing policies. This proactive approach not only enhances the health outcomes for policyholders but also reduces risk for insurers. Additional possibilities include the development of real-time pricing models, with discounts that remain in effect if positive healthy attributes or activity levels continue or enabling automated coverage amount increases as needs change as long as policyholders maintain a healthy lifestyle and are willing to share their data with their insurers.

With this increased ability to have additional data, new products can be created that offer discounts based on health goals or health metrics. In other insurance parallels, such as health or property & casualty, premiums are often only set or guaranteed for one year. While the life insurance product is typically guaranteed in some manner to maturity, the focus on additional discounts allows lower charges to help increase persistency of healthy lives and align insureds' future costs with their future health decisions.

WTW has only scratched the surface of validating and testing the Klarity model on U.S. data. Given the original model we tested was trained on U.K. data, we expect even more significant lift in the value the model can add by training the model on U.S. insurance data. In addition, the model can be trained to offer more customized tailored solutions to specific companies, which can be tested in pilot programs.

WTW reviewed the model's ability to optimize hyperparameters using a subset of target data, which may enhance its predictive performance. Moreover, its strong performance in stratifying risks presents an opportunity for more precise classification and rating across a range of risks, including potentially more complex cases, offering insurers a valuable tool for improved underwriting and pricing accuracy.

Regulatory considerations, such as those in Colorado and New York for fairness and bias in underwriting and use of external data sources, remain critical. While these data sets are widely acknowledged in the medical community as a driver of health outcomes, they are currently categorized as External Consumer Data and Information Sources under emerging regulations. As this field evolves, testing the Klarity model's outputs for specific populations will be essential to ensure compliance and fairness, particularly until wearable data are fully integrated into the regulatory framework as medical data.

Section 5:
About Klarity

Klarity is a U.K.-based health data analytics firm that provides preventative health solutions. Its proprietary mortality and morbidity risk prediction models leverage advanced data sources such as wearables, electronic health records and blood biomarkers. Klarity empowers insurers, brokers and employers to improve risk stratification, early detection of chronic diseases and personalized health insights for their customers.

Klarity solutions consist of:

Predictive analytics
Offering mortality and morbidity models to drive data-driven decisions

Screening solutions
Providing innovative pre-screening technology for chronic disease detection, emphasizing cancer

Integrated health insights
Combining advanced analytics with lifestyle recommendations in nutrition, exercise, emotional wellbeing and sleep, ensuring holistic health solutions

https://klarity.health

Klarity

Section 6:
About WTW

WTW's Insurance Consulting & Technology business serves the insurance industry with a powerful combination of advisory services and leading-edge technology. Our mission is to innovate and transform insurance, and we deliver solutions that help clients better select, finance and manage risk and capital.

We work with clients of all sizes globally, including most of the world's leading insurance groups. Our specialist insurance software is used by over 1,000 client companies on six continents. With over 1,700 colleagues in 35 markets, we continually strive to be a partner and employer of choice to the insurance industry.

Our ICT team:

Mary Bahna-Nolan, FSA, MAAA, CERA
Senior Director
mary.bahna-nolan@wtwco.com

John Gramelspacher, FSA, MAAA
Director
john.gramelspacher@wtwco.com

Justin Fountain, ASA, MAAA
Manager
justin.fountain@wtwco.com

James Burgess, ASA, MAAA
Senior Consultant
james.burgess@wtwco.com

www.wtwco.com

Reliances and limitations

- WTW relied on the risk scoring model results provided by Klarity and did not have access to the actual model and underlying data inputs or weights.
- The model results and analysis relied upon publicly available data. WTW selected data based on broad insurance demographic profiles. An insured population may exhibit different characteristics from the population studied in our analysis.
- The analysis was based on broad industry information and thus, individual company experience may vary or application may need to be adjusted for any individual company.
- Underwriting scoring for risk classification did not have full access to individual medical records and may not fully reflect all risk characteristics of the individuals analyzed.
- The views expressed within this paper are those of WTW. WTW did not receive renumeration for the opinions and analysis.
- This report is not to be distributed or relied upon other than in its entirety or without the express written permission of WTW.

Disclaimer

WTW hopes you found the general information provided in this publication informative and helpful. The information contained herein is not intended to constitute legal or other professional advice and should not be relied upon in lieu of consultation with your own legal advisors.

This paper may contain information or materials created or provided by third parties over whom Willis Towers Watson has no control or responsibility. These third-party information or materials are not under Willis Towers Watson's control, and Willis Towers Watson is not responsible for the accuracy, copyright compliance, legality, or any other aspect of such third-party information or materials. The inclusion of such third-party information or materials does not imply endorsement of any third parties by Willis Towers Watson or any association of Willis Towers Watson with any third parties.

About WTW

At WTW (NASDAQ: WTW), we provide data-driven, insight-led solutions in the areas of people, risk and capital. Leveraging the global view and local expertise of our colleagues serving 140 countries and markets, we help you sharpen your strategy, enhance organizational resilience, motivate your workforce and maximize performance. Working shoulder to shoulder with you, we uncover opportunities for sustainable success — and provide perspective that moves you. Learn more at wtwco.com.

wtwco.com/social-media

wtw-HP-2025-0831

wtwco.com