Sikka Health Indicators and Risk Scores

Sikka Software, Sikka AI Lab (SAIL)

Dec, 2022

Sikka.ai is the leading API Platform in the retail healthcare industry that includes opt-in dentistry, veterinary, audiology, optometry, chiropractic, orthodontics, oral surgery, and other medical practices. Sikka.ai API Platform seamlessly connects to 96% of the retail healthcare market throughout the US and Canada and is processing billions of transactions a day. The Sikka.ai API Platform provides the base to help practices optimize their business, profitability, patient communication, revenue cycle management, patient satisfaction, and patient medical history analysis, by enabling over 50 marketplace applications on its platform.

Sikka has over 28,000 opt-in dental practices installed in the US and Canada through its market-leading API integration platform. Sikka leverages this rich dataset to determine if a life insurance applicant has one or more of fourteen pre-existing conditions or habits with an outsized impact on underwriting. These indicators are based on actual clinical notes from the licensed providers in the practice or patient-reported conditions on health history forms. Moreover, Sikka has been working on two mortality risk scores, LECA (Life Expectancy at the Current Age) and the Periodontal Disease to Mortality Score, based on a correlational analysis that measures patients' mortality risk based on their dental procedure history. Sikka's health indicators and risk scores can help determine the appropriate risk class for a policyholder. This can help ensure that "less risky" life insurance policyholders don't end up subsidizing "risky" ones, regardless of whether or not conditions are mischaracterized accidentally or intentionally.

1. Risk Scores

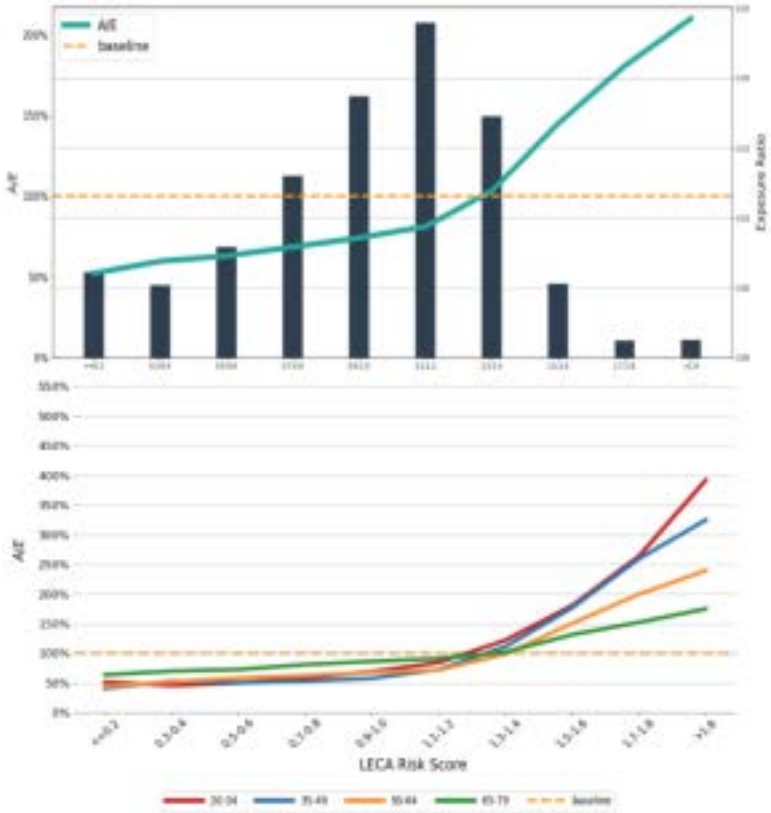
Within the past few decades, many clinical studies[1] indicate that there is a significant correlation between oral care and mortality rate. Studies show that the severity of gum diseases impacts the risk of other health issues such as cardiovascular disease because the harmful bacteria and infections from gum diseases can spread into the blood which affects overall health. When the harmful bacteria and infections reach a susceptible heart, they can cause inflammation called "Endocarditis". These studies were the initial trigger to perform a survival analysis study on Sikka's uniquely large dataset of dental records (currently at over 150M US patients[2]) to assess the possible mortality risk correlation with an individual's dental procedure history. So far, Sikka has released two mortality risk scores, LECA (Life Expectancy at Current Age) and the Periodontal Disease to Mortality Score to calibrate insurance carrier's pricing paradigms.

Currently, most of the existing mortality risk scores used in insurance processes are built based on Rx history data, credit score data, etc. However, there is no mortality risk score built from dental data until now. Based on a large number of studies that demonstrate a substantial link between dental health and overall health, Sikka built two new risk scores using sufficient dental data that are comparable to the existing risk scores in the market.

1.1 LECA (Life Expectancy at the Current Age)

Since 2019, Sikka has been working with life insurance and reinsurance companies to perform retrospective analysis on millions of policyholders and has refined the Life Expectancy at Current Age (LECA) score, a predictive risk score based on a correlational analysis that measures a patient's mortality risk based on their dental procedure history. LECA Risk Score has a range of 0 to 2 with 0.1 intervals. Higher scores are associated with higher mortality risk.

[1]Relationship between oral health and mortality rate
https://pubmed.ncbi.nlm.nih.gov/12472996/#:~:text=Results%3A%20A%20significant%20correlation%20was,indicator%20of%20all%2Dcause%20mortality.
[2]Of all the 'patients'/ 'patient' in the whitepaper, we mean consent-based patients.



2008 Entry, 6-Year Follow-Up, Entry Age 20-79

Figure 1. Actual to Expected Deaths by LECA Risk Score, Exposure Ratio shows Sample Distribution.

The LECA Risk Score was validated on around 5 million patients using a 6-year time period. We observed that the LECA Risk Score has great power in stratifying the relative mortality risk. The relative mortality values have a semi-linear relationship with the LECA Risk Score. Populations with a higher LECA Risk Score exhibit a relatively higher mortality risk. For each age group, the LECA Risk Score shows the power to differentiate the A/E across the different LECA Risk Scores. The LECA risk scores provide significant predictive information regarding future mortality across all age groups studied. In the age group 20-34 and 35- 49, the highest LECA Risk Score has an A/E more than 7x that of the lowest LECA Risk Score (Figure 1).

In the entire study, we also found that the LECA Risk Score provided great power in differentiating the mortality risk of patients with pre-existing conditions. We combined the disease data, age, and gender at the time of death with the LECA Risk Score to prove that the LECA Risk Score provides incremental information to predict mortality. It does so without duplicating additional health data a life insurance underwriter would typically have from other sources such as the Rx database or EHRs.

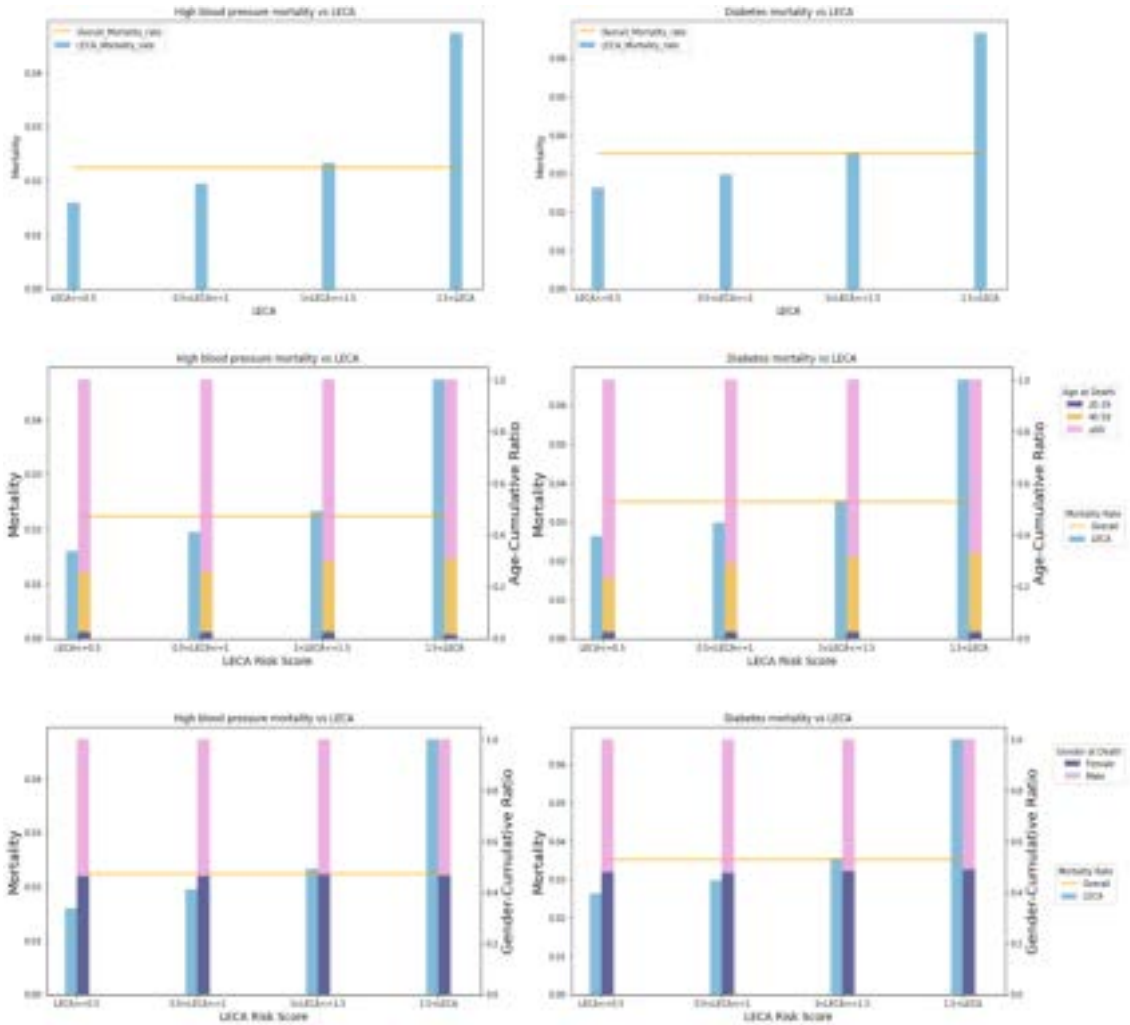


Figure 2. Mortality rate in the HBP and Diabetes patients with respect to LECA Risk Score, Blue bars show the mortality rate of the HBP and Diabetes patients within different LECA cohorts.

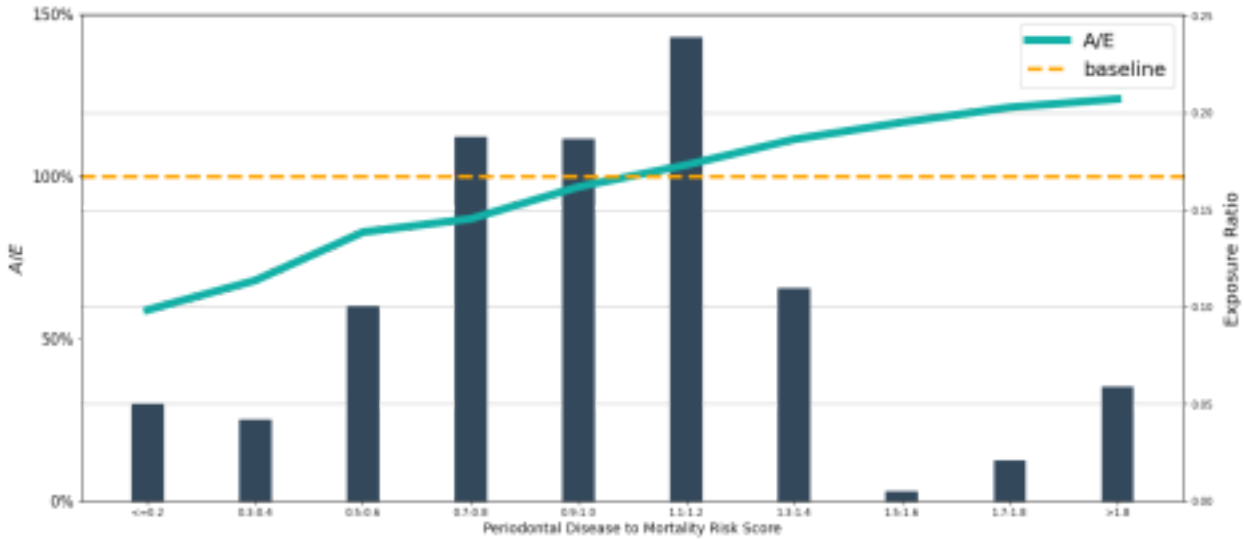
Figure 2 shows a positive correlation between the LECA Risk Score and the mortality rate for high blood pressure & diabetes patients. As the LECA Risk Score increases, the mortality rate in high blood pressure patients and diabetes patients also increases. Age and Gender, which are commonly thought to be highly correlated with mortality, do not interact with the LECA Risk Score's differentiation power. The same applies to other health diseases such as **cardiovascular disease**, **cancer**, and **tobacco usage** (see the LECA Risk Score white paper for details). The analysis illustrates that a combination of the LECA Risk Score with HIPAA-compliant use of oral care data[3] provides a lift in predicting mortality and improving the efficacy of the underwriting process. Since the LECA Risk Score is available (with the appropriate authorization of an applicant) within seconds, it's easier to use than a traditional EHR/EMR and suitable for a Simplified Issue Term product, other simplified issue products, and for any automated or accelerated underwriting process with a high level of efficacy in terms of predicting relative mortality.

[3] Terms of Service | sikka.ai

1.2 Periodontal Mortality Risk Score

The Periodontal Disease to Mortality Score is a scoring algorithm that uses periodontal disease history to predict the mortality risk of individuals relative to other individuals of the same age and gender.

Similar to the analysis Sikka completed on the Life Expectancy at Current Age (LECA) score algorithm using dental procedure history, Sikka's goal was to understand the relationship of periodontal disease history to mortality outcomes. Since periodontal diseases are proven to be correlated with cardiovascular disease[4], contribute to increased mortality associated with diabetes[5], and can impact cognitive functioning[6], it appeared to be an ideal data point to explore for the life insurance market.



2008 Entry, 6 Year Follow Up, Entry Age 20-79

Figure 3. Periodontal Disease to Mortality Score analysis and validation

The Periodontal Disease to Mortality Score algorithm consists of 64 periodontics procedures and periodontal disease indicators. About 5 million individuals were validated by Sikka over a 6-year follow-up to understand the long-term mortality implications of the score. As shown in Figure 3, the Periodontal Disease to Mortality Score achieves 2.11X segmentation between the best and worst risks (0-0.2 and 1.9-2.0, respectively). Since age and gender are accounted for within the calculation of the expected deaths, the increase in A/E is due to the differences in periodontal treatment history that are associated with the Periodontal Disease to Mortality Score.

[4]Beck J, Garcia R, Heiss G, et al (1996) Periodontal disease and cardiovascular disease. Journal of Periodontology 67:1123–1137. doi: 10.1902/jop.1996.67.10s.1123
[5]Saremi A, Nelson RG, Tulloch-Reid M, et al (2005) Periodontal disease and mortality in type 2 diabetes. Diabetes Care 28:27–32. doi: 10.2337/diacare.28.1.27
[6]Asher S, Stephen R, Mäntylä P, et al (2022) Periodontal Health, cognitive decline, and dementia: A systematic review and meta analysis of longitudinal studies. Journal of the American Geriatrics Society. doi: 10.1111/jgs.17978

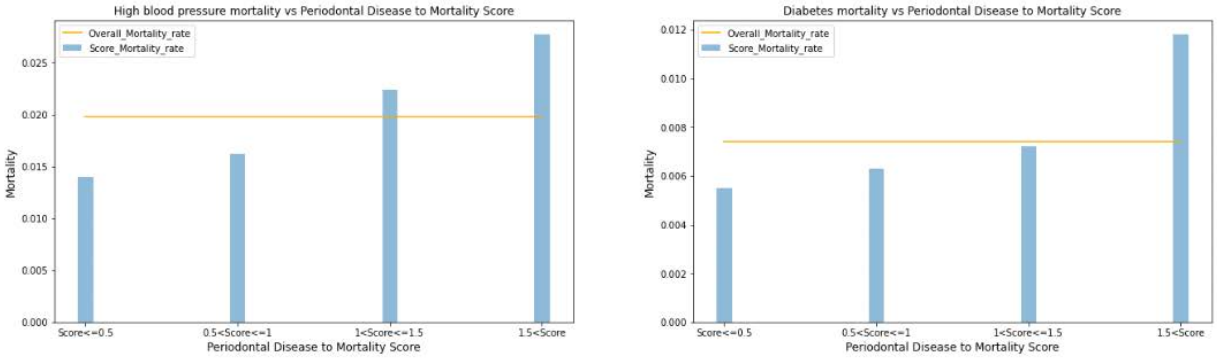


Figure 4. Mortality rate in high blood pressure and diabetes patients with respect to Periodontal Disease to Mortality Score

Figure 4 shows as the score increases, the mortality rates related to Cardiovascular and Cancer patients also increase. The same applies to patients who have high blood pressure or diabetes. The population study demonstrates the risk stratification power of the Periodontal Disease to Mortality Score for applicants with pre-existing conditions.

The above study affirms that Periodontal Disease to Mortality Score is an effective and long-term predictor of mortality. The score continues to provide mortality segmentation to applicants who have pre-existing conditions. Sikka will continue to explore how insurance underwriters can incorporate the Periodontal Disease to Mortality Score into their existing mortality risk modeling to improve the efficacy of underwriting and help make life insurance more affordable, especially for impaired applicants.

2. Sikka Health Indicators

Health Indicator	Description	Indicator Values
Sikka_tobacco_ indicator	Indicates if a patient is an active user of tobacco.	❖ 'T' = patient is a tobacco user ❖ 'NE' = no evidence of tobacco use
Sikka_ cardiovascular_ indicator	Indicates whether a patient is suffering from a cardiovascular disease (CVD) condition.	❖ '1' = patient has/had a heart disease, stroke, heart murmur, heart attack or valve prolapse ❖ 'NE' = no evidence of heart related disease condition
Sikka_cancer_ indicator	Determines if a patient has any non oral cancer.	❖ '1' = patient has/had cancer or is/was undergoing chemotherapy, radiation treatment, or other cancer treatments ❖ 'NE' = patient has no evidence of cancer
Sikka_diabetes_ indicator	Score indicates if a patient has diabetes.	❖ '1' = patient is diabetic or taking diabetic medications ❖ 'NE' = patient has no evidence of diabetes

Sikka_marijuana_ indicator	Indicates use of marijuana.	❖ '1' = the individual is an active user of marijuana or drugs ❖ 'NE' = no evidence of marijuana usage
Sikka_kidney_ indicator	Determines whether or not a patient has any kidney related disease.	❖ '1'= the patient has/had kidney disease, kidney stones, kidney failure or bladder disease ❖ 'NE' = No evidence of kidney related issues
Sikka_liver_ indicator	Identifies whether a patient has jaundice, hepatitis, or liver disease.	❖ '1' = the patient has/had liver disease ❖ 'NE' = No evidence of liver disease
Sikka_hbp_ indicator	Determines whether a person has hypertension or high blood pressure.	❖ '1' = the patient has high blood pressure or is taking hbp medications ❖ 'NE' = No evidence of high blood pressure
Sikka_hld_ indicator	Indicates Hyperlipidemia condition, which is characterized by an excess of lipids (or fats) in the blood, such as cholesterol and triglycerides.	❖ '1' = the patient has high cholesterol or is taking medicine to treat hyperlipidemia ❖ 'NE' = No evidence of high cholesterol
Sikka_blood_ disorder_ indicator	Indicates blood disorder and blood thinning medication use.	❖ '1'= denotes the presence of a specific blood illness such as hemophilia, von Willebrand disease, platelet disorders, etc. or the use of blood thinners such as aspirin, heparin, warfarin, etc. ❖ 'NE'= No evidence of blood disorders
Sikka_Alzheimer's_ indicator	Indicates Alzheimer's or early onset dementia.	❖ '1'= Indicates alzheimer's or early onset dementia ❖ 'NE'= No evidence of alzheimer's or early onset dementia
Sikka_sleep_apnea_ indicator	Determines whether a person has sleep apnea.	❖ '1'= Indicates the person has sleep apnea ❖ 'NE'= No evidence of sleep apnea
Sikka_COVID_ indicator	Determines whether a person had an instance of COVID in their dental clinical notes.	❖ '1'= Indicates the person had COVID at some point ❖ 'NE'= No evidence of COVID
Sikka_perio_ indicator	Determines whether a person has periodontal disease and determines the severity of the condition.	❖ '1'= Indicates the person has periodontal disease ❖ 'NE'= No evidence of periodontal disease

2.1 Text Classification Model

From Sikka's vast database, patients with clinical notes that contain specific keywords are identified for each health indicator. These clinical notes were preprocessed using various NLP preprocessing techniques. The **Word2vec** algorithm was used to generate a distributed representation of words from clinical notes as numerical vectors, capturing the semantics and relationships between words.

The embedded words are fed into the **Long Short Term Memory (LSTM)** model Figure 6, which is a type of recurrent neural network capable of learning order dependence in sequence prediction problems. The LSTM model is effective in memorizing important information; furthermore, unlike traditional classification algorithms, LSTM can use a multiple-word string to find out the class to which it belongs. The LSTM model was trained on a 400,000 balanced dataset with an accuracy of 99%.

As part of enhancing the text categorization, a **rules engine** was developed to incorporate any incorrect classifications found in the retrospective studies.

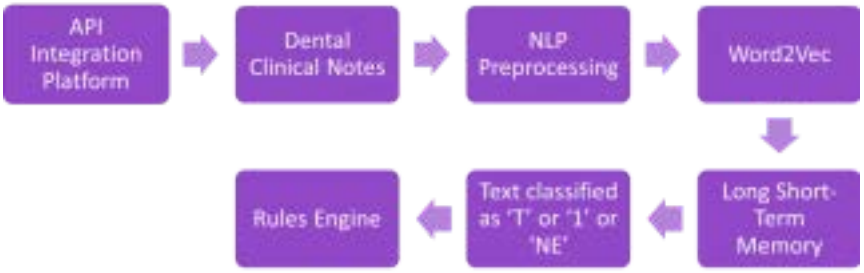


Figure 5. Text Classification Process

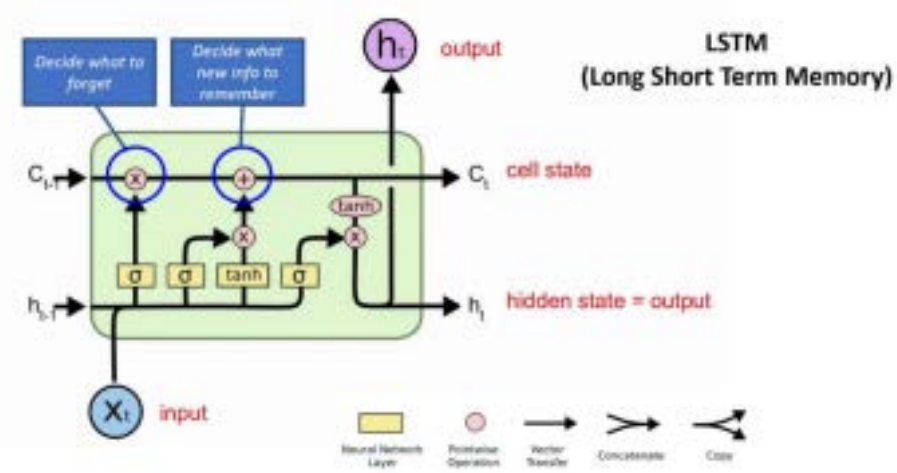


Figure 6. LSTM Architecture

2.2 Retrospective Studies and Hit Rate

Sikka's data has been validated in studies conducted by three major reinsurance companies, three leading data providers, several carriers and MGAs in both the US and Canada, and a leading life settlement company. These studies range from a select 2,500 to an expansive 8,000,000 records and have match rates of up to 54%. The Tobacco Indicator has identified significant numbers of "smoking non-disclosers" that cost carriers as much as $23,000 per traditional term policyholder in lost premiums due to misclassification, based on a cost-benefit analysis from ExamOne. Separately, Sikka Indicators are now in production with multiple carriers and have been helpful in identifying missing underlying conditions that influence underwriting. Studies of Sikka's Tobacco Indicator have been completed to identify a gross protective value of almost 10x the cost.[7] In 2019, Munich Re performed a validation of Sikka's Tobacco Indicator using insured records. Research confirms that information about dental health can be informative about overall health. Munich Re recommends each carrier perform a retrospective study to best assess the value and application of the Sikka Tobacco Indicator on its company-specific insured population.[8]

3. Summary

In the past few years, Sikka has been working on using machine learning to extract useful information from dental data which can help the underwriting process become more efficient. The usefulness of Sikka's indicators to give a "lift" in the insurance risk assessment is further demonstrated by retrospective studies with reinsurance companies. In addition, Sikka has found an interesting correlation between the low-risk score and improved mortality despite chronic and habitual conditions. For example, a high blood pressure patient or tobacco user with a low LECA Risk Score has a lower mortality risk than the average high blood pressure or tobacco user. The same can be seen for diabetes, cardiovascular disease, and cancer. This use of oral care data can be very useful for a variety of underwriting situations, particularly in combination with other data sources as well as traditional age and gender tables. Given that Sikka's Indicators and risk scores can be delivered in real-time via an API feed (applicable in both batch and single search), they can help with automated or accelerated underwriting to better risk differentiate impaired applicants. Sikka's indicators and risk scores can also be used for post-issue analysis to verify that an approved applicant has been put into the appropriate risk classification.

[7]ExamOne Cost Benefit Analysis, June, 2021, Brian Lanzrath
[8]https://www.munichre.com/us-life/en/perspectives/alternatives-for-stratifying-mortality-risk/oral-health-mortality-and-smoker-detection.html