March 22, 2024

SuperLifeStyle

A Health Incentive Program for SuperLife Life Insurance

Team Name:
Apex Analytics

Dharmadasa, Helitha
Gao, Edmond
Kirubakaran, Krishanth
Narayan, Viswesh
Verghese, Nikhil

SOCIETY OF ACTUARIES

UNSW SYDNEY

Contents

List of Tables

List of plots

1 Background and Objectives

We at Apex Analytics have been tasked with creating a health incentives program for SuperLife to implement alongside their long-term life-insurance products in Lumaria. We have approached this problem by combining several of the individual health incentives researched and considered by SuperLife's product development team into a program we have named SuperLifeStyle. Policyholders of SuperLife's long-term products can choose to participate in this program by engaging with all the health incentives in exchange for a discount on their premiums.

Here is a list of target objectives SuperLife has specified for the program, a summary of how SuperLifeStyle will aim to address them, and our proposed approach to monitoring the program's success in achieving each one:

1. **Incentivize Healthy Behaviors Through Participation in the Program**
   Participation in SuperLifeStyle involves engaging in healthy behaviors, and policyholders will be encouraged to participate through discounts to their premiums. Advertising the health benefits of participation will also encourage participation. The percentage of viable policyholders who choose to engage with each separate health incentive (engagement rates) can be used to measure and track the success of each incentive and SuperLifeStyle as a whole in achieving this objective.

2. **Decrease Expected Mortality**
   SuperLife has provided us with data indicating the health incentives which we have incorporated into SuperLifeStyle will decrease expected mortality for those who engage with them, and we have done our own external research to validate these numbers. Thus, the implementation of this program will lead to the decrease in expected mortality for participating policyholders. A comparison of mortality rates for participating policyholders with those of the general population of Lumaria can be used to measure the success of the program in decreased mortality. This metric will only be accurate long-term and thus success may be unclear for the first few years after implementation.

3. **Increase Life Insurance Sales**
   The discounts offered by SuperLifeStyle will aim to increase sales by attracting customers to SuperLife's long-term insurance products; Lumarians who may not have considered purchasing these products before due to the price will now be able to access them at a price they would accept in exchange for participation. A comparison of sales before and after SuperLife's implementation can be used to measure the success of this objective. The number of new policyholders who were not previously purchasing these products but are now doing so and participating in the scheme can also be used to measure the instantaneous impact on sales.

4. **Improve Product Marketability and Competitiveness**
   Similarly to the previous objective, the discounts offered by SuperLifeStyle will lower the effective price of long-term insurance for those willing to participate in the program, which will improve the marketability and competitiveness. Furthermore, being seen to encourage healthier lifestyles will improve SuperLife's brand reputation, which

will also contribute to the competitiveness of all their products. A comparison of the insurance sales metrics mentioned above to those of competitors will give an indication of SuperLifeStyle's success in improving marketability and competitiveness. The number of new policyholders who were previously insured by competitors can also be used.

5. **Add Economic Value to SuperLife**
   The reduction in mortality for participating policyholders will reduce their expected mortality and thus reduce the expected claims costs of these insurance products. The discounts offered by the scheme can be in the form of reduced premiums properly repriced in the hopes of increasing sales volume, or SuperLife can choose not to reprice premiums and hold on to additional cost savings themselves. SuperLifeStyle will provide economic value through the increase in insurance sales and competitiveness of SuperLife's products, as well as providing unique value in the form of health interventions that will separate SuperLife from its competitors. A comparison of the average profit margin per policy of participating and non-participating policyholders can be used to verify the consistency of the profits at any point in time after implementation, and the insurance sales metrics mentioned previously can be used to measure the success of the program under this objective.

2 Program Design

SuperLifeStyle consists of four of the health incentives provided to us by SuperLife's product development team. Policyholders can participate in SuperLifeStyle by engaging with all four incentives and will receive a discount on their premiums.

Participation in the program will be encouraged through:

- **Discounts:** Participation in SuperLifeStyle comes with a reduction in the premiums for the policyholder's long-term insurance products. This reduction will be advertised so that policyholders know of the financial benefits of participation.
- **Health benefits:** Engaging with the program will lead to a decrease in mortality. The statistics on these mortality impacts will be made clear in the marketing of the scheme so that consumers are aware of the health benefits of participation.

These are the selected four incentives and how we plan to implement them:

2.1 Safety Campaigns

Safety campaigns play a pivotal role in SuperLifeStyle, offering educational opportunities to policyholders on a variety of safety topics. These campaigns are designed to empower individuals with awareness, encouraging them to make more informed decisions. From fire safety and home security to road safety and health awareness, safety campaigns cover a broad spectrum of topics aimed at minimizing risks and protecting lives. SuperLife allocates approximately Č10 - Č35 per participant towards content development, including educational materials, infographics, videos, online modules, and in-person demonstrations.

These resources are disseminated through various channels such as emails, social media, and websites, incentivizing participation with discounts to premiums ([1],[10],[6]).

2.2 Community Fitness Challenges

SuperLife's SuperLifeStyle scheme can provide incentives such as premium discounts, gift cards, or other rewards to stimulate engagement in fitness challenges. Beyond mere motivation for individual health priorities, SuperLife can integrate community fitness challenges with its term and life assurance offerings. This strategic bundling not only aims to reduce mortality rates but also to enhance sales, revenue, foster brand loyalty and engagement, and help SuperLife balance expenses and mortality reduction.

2.3 Preventative Screening

SuperLife can offer rewards to policyholders that undergo preventive screening programs, specifically those focused on cancers. Rewards include covering an annual screening session (Č32.5), and a reduction in yearly premiums for those that engage with the intervention (Č32.5). As most cancer related deaths occur beyond the age of 50, SuperLife can offer this intervention to policyholders above this age. Research supported mortality reduction figures roughly in-line with SuperLife's own findings ([2],[8]).

2.4 Smoking Cessation

This incentive encourages the quitting of smoking. Participants will be directed to facilities and services known to help with the cessation of smoking, such as individual and group counselling sessions, and nicotine replacement therapy. Successfully abstaining from smoking is required for the policyholder to be considered as engaging with this incentive ([9]).

While short-term participation in these initiatives is enough to begin accessing the discounts, the impact to mortality from engagement will only become apparent over a much longer period, thus consistent long-term participation is a key requirement of the program. Participants who cease to engage will no longer receive discounts and may be required to pay back a portion of the premiums previously discounted.

The selection process of these four incentives was as follows: we set up a model to rank the extensive list of incentives provided to us by SuperLife's product development team based off the impact to mortality balanced by the cost of implementation and identified Safety Campaigns and Community Fitness Challenges as the two most effective. We then selected two more incentives that also ranked highly that addressed specific health concerns that we noticed had a large impact on SuperLife's policyholder demographic; cancer is a leading cause of death for policyholders and thus Preventative Screening was selected, and a large portion of policyholders are smokers, so Smoking Cessation Programs was selected. Details on the ranking methodology can be found in A.1.

These incentives were selected through external research, we were able to find details on how these health initiatives worked which aided in the construction of SuperLifeStyle's logistics. We were also able to broadly validate the mortality impact and cost figures provided by SuperLife; this will be discussed further in the Data and Data Limitations section of the report. All sources are listed in the bibliography.

3 Pricing and Costs

This section contains the approach and results of analysis on the economic and mortality impacts of implementing SuperLifeStyle.

We have used the Lumarian life table provided by SuperLife's product development team as the basis for pricing non-participating policyholders. We then applied the mortality impact numbers of each incentive to produce an adjusted life table representing the mortality patterns of a participating policyholder. Details on the mortality adjustment modelling are included in A.2.

A standard life insurance model was then used to generate total expense numbers for non-participating policyholders. This same model was subsequently used to calculate costs for participating policyholders, with the adjusted life table used as an input instead, and with the implementation costs for each incentive factored in. Technical details of the economic model are included in A.3.

Health Incentive	Impact on Mortality	Cost and Cost Pattern
Safety Campaigns	3% to 5% for all age groups	Č12.5: Yearly (Start of year)
Community Fitness Challenges	4% to 5% for 25-44 year olds 3% to 4% for 45-59 year olds 2% to 3% for 60+ year olds	Č18: Yearly
Preventative Screening	8% to 10% for 25-44 year olds 6% to 8% for 45-59 year olds 5% to 7% for 60+ year olds (A screening loading is applied to all rates)	Č65: Yearly (Start of year for those 50+)
Smoking Cessation Programs	12.5% to 50% for 25-44 year old smokers 11.5% to 25% for 45-59 year old smokers 7.5% to 11.5% for 60+ year old smokers (18% of adult Lumarians are smokers)	Č2065: One-time payment (at the lowest of issue age or 27, up to a max of age 88)

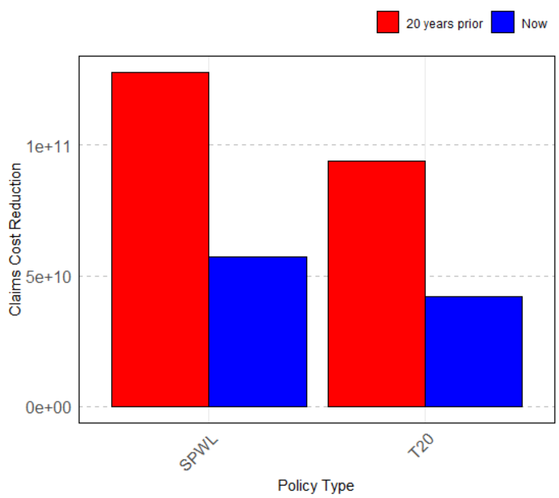
Table 1: Mortality Impact and Cost Figures Used as Inputs to the Models

Prices for the baseline scenario without SuperLifeStyle and prices with policyholders participating with the program at a 25% engagement rate are displayed together to allow for easy visualization of the reduction in costs that SuperLife can expect. SuperLifeStyle will provide economic benefit to SuperLife for term insurance products issued to those between the ages of 23-84, and whole life products issued to those above the age of 23.

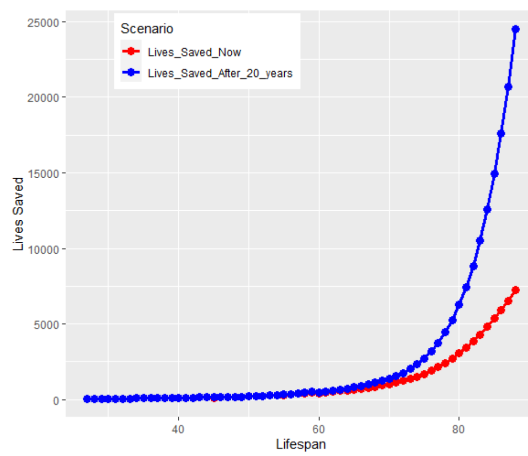
Along with other expenses, these figures can be used to generate premiums for new policyholders at different issue ages. By only providing access to SuperLifeStyle to profitable age brackets, SuperLife can benefit by either passing on cost reductions onto policyholders to increase sales volume, and/or leave premiums priced without SuperLifeStyle while holding on to the additional cost savings generated by the program. SuperLife should actively monitor

Age Group	20 Year Term Insurance (Č)		Whole Life Insurance (Č)	
	Baseline	25% Participation	Baseline	25% Participation
1-22	5,221	5,525	319,738	319,317
23-44	25,143	24,237	422,769	421,896
45-59	116,220	111,143	528,334	527,197
60-84	428,183	422,632	647,445	646,206
85+	582,268	582,503	743,699	743,920

Table 2: EPVs of Benefits and Intervention-Related Expenses for each of SuperLife's long-term Insurance Products by Age Group



(a) Claims Cost Reduction



(b) Comparison: Lives Saved Now vs. 20 Years Prior

the market and consumer response to adjust pricing and premiums dynamically to optimize both profit, value to customers and value to the business.

The provided data highlights the projected number of lives saved at different lifespans under two scenarios: The lives saved after the implementation of the program which represents the current situation, and the other depicts the anticipated outcomes 20 years prior. Overall, there is a trend of increasing lives saved with advancing lifespans in both scenarios. Additionally, the analysis of claims cost reduction further highlights the efficiency of these interventions with significant difference between the policy types. For T20 policies, the current claims reduction stands at Č11,491,341,272.57, expected to increase to Č14,409,136,415.29 after 20 years. In contrast, SPWL policies demonstrate a higher current claims cost reduction of Č58,051,148,810.82, with an anticipated future reduction of Č131,818,490,081.97. These findings underscore the importance of interventions and policies aimed at improving mortality rates, particularly in older age groups, to sustain and enhance public health outcomes over time.

4 Data and Data Limitations

The primary source of quantitative data used for our analysis was that provided by SuperLife's product development team. This data includes figures on the impact of mortality and costs of various health incentives, the policy in force data for SuperLife's policyholders, and a Lumarian life table.

For the four specific incentives we have selected, there are no significantly unreasonable figures provided, and we have managed to find external sources and studies which indicate the mortality impact and costs of such programs roughly align with the figures provided ([4],[9]), however it must be acknowledged that definitive validation of these numbers was difficult to come across. While we have not found evidence that contradicts these numbers or suggests they are inappropriate, this is still a significant data limitation, as variances in these figures will have a direct impact on the results of our models.

Another key data input for our modelling is the Lumarian life table; we have used the table provided by SuperLife as this is likely the same version, which they currently use to price their insurance products and thus would be most appropriate.

In order to check the reasonability of this life table, we have cross-validated it against two other data sources:

1. SuperLife's policy in force data: We have constructed an alternative life table using the data provided on SuperLife's policyholders to check there are no major differences. While the rates are not entirely consistent, this is to be expected as the alternative table is biased as it only considers SuperLife's current customers and not the wider Lumarian population. The comparable figures are reasonably similar, indicating that the sets of data are not glaringly inconsistent. Details on how the alternative life table was constructed from the policy in force data are included in Figure 5.

2. Australian Life Table: As an additional reasonability check, we have compared the provided life table to that of another country which is demographically similar. The rates in the Australian life table ([3]), a country that shares a very similar economic and cultural background to Lumaria, are close to those of the one provided by SuperLife, and they appear structurally similar.

While our checks indicate that the life table, we have used is reasonably and not clearly erroneous, it is important to acknowledge that it has not been perfectly validated. Issues with this life table will have major impacts on our results as variances will impact our mortality and economic analysis.

5 Assumptions

This section details the assumptions we have used in our analysis and their significance.

Qualitative assumptions include:

- **Independence of Incentives:** Our current modelling approach assumes that the four individual incentives do not interact, and a policyholder will benefit fully from the expected mortality reduction from each one. In reality, there is likely to be at least some interaction between some of the four incentives; if this is the case, our analysis will be overestimating the reduction to mortality from participation, and thus our projections of lives saved will also be overestimated. This will also lead to the premiums for participating policyholders to be underpriced, which could cause major variances in our economic analysis. In our research, we were unable to find evidence of interactions between these four different health schemes; while this does not mean there are no overlaps, it implies that any that do exist are likely to be minimal. Additionally, the grouping of policyholders into age groups (e.g., "25-44," "45-59," "60+") assumes that mortality risk within each age group is homogeneous. However, within-group variations in mortality risk factors or health conditions may exist but are not explicitly considered in the analysis.

- **Static Life Table:** In our calculation of SuperLife's mortality savings if SuperLifeStyle was implemented twenty years ago, the projection of future mortality rates assumes a constant mortality reduction rate over the 20-year period. This inherently assumes the life table has not changed over the last twenty years, which may not be the case. Changes in population demographics, healthcare advancements, or other external factors that may influence mortality trends are not accounted for in this analysis. If this assumption is incorrect and the mortality rates have shifted, our calculation would be incorrect, however it is unlikely that the mortality of Lumarian citizens has changed to a degree that would have a significant impact.

- **Pricing and Costs:** We assume all expenses aside from cost of implementation, profit margins and other factors outside of mortality are constant when performing the economic benefit analysis. This simplifies the pricing model into a simple cost comparison between new and old scenarios. While the listed implementation costs should be inclusive of all factors associated with the program, there may be other expenses that vary that would impact the findings of the cost analysis.

Quantitative assumptions include:

- **Engagement Rate**: we have selected a static engagement rate of 25% for the entire program based off figures suggested by Swiss Re ([5]). This selection is quite arbitrary as more in-depth indications of engagement rates were difficult to find. We also assume equal engagement across all interventions by eligible policyholders. The true engagement rate is likely to differ and may also not remain static but vary over time. These differences will not impact the pricing of premiums, however our projections on SuperLife's profits and mortality savings will be incorrect.

- **Interest Rate:** Based on the average of Lumeria's 1 Year Spot rate over the period spanning 1962 – 2023. A variance in this assumption would lead to possibly significant differences in the present value figures in our expense analysis.

- **Inflation Rate:** Based on the average rate of Inflation over the period spanning 1962 – 2023. We believe the average value of the 1 year spot and inflation rate over as much data as was available provides the most accurate indicator of Lumeria's future interest rates for valuing the product over the timeframes considered. A variance in this assumption would lead to differences in our expense analysis figures.

We have sensitivity tested our quantitative assumptions and the findings are discussed in the following section.

6 Risk and Risk Mitigation Considerations

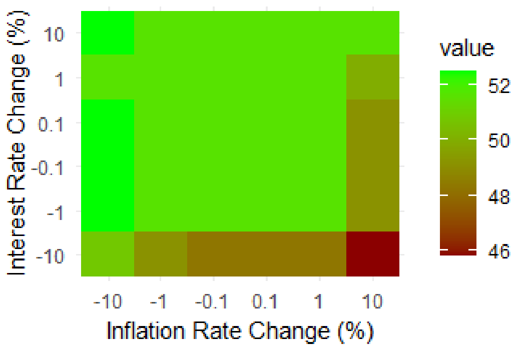
The implementation of SuperLifeStyle is accompanied by several key risks which this section will outline and discuss mitigation strategies for. As discussed in the previous section, a key quantitative risk is the incorrectness of assumptions used. To address this risk, we have conducted sensitivity testing to identify the severity of the consequences should this occur. Sensitivity analysis was conducted at three levels: 0.1%, 1% and 10% difference for the interest rate, inflation rate, and engagement rate simultaneously and was used to measure expected number of opportunities for profit for each of 120 ages rather than actual profit, given how dynamic we have designed our model to be. The inflation rate usually is tied to the interest rate; hence we detail their collective effect on profit opportunities. We have considered a modest engagement rate of 25% but see no apparent impact by even a change of ±10%. In all instances, T20 showed overall little to no fluctuation in profit opportunities with a maximum of 7% deviation from the standard case of 62 out of 120, or 52%, and behaves as a milder variant of the SPWL case.

Input Variables	Effect on Profit Counts for SPWL
Interest (and Inflation) Rate	SPWL meanwhile had varying results: • -10%: We clearly see no opportunity for profit when the inflation rate reciprocates and rises by 10%, and maximum benefits when both fall by 10%. It otherwise is close to 20% for lower impact changes to the inflation rate. • +10%: We achieve peak chance of profits at 72%. • -1% to +1%: As long as inflation does not rise by 10%, we see a peak chance of profits at close to 72%. It otherwise varies from 25 to 50%.
Engagement Rate	This had a much less pronounced effect on our results and showed little to no difference when fixing interest and inflation rates.

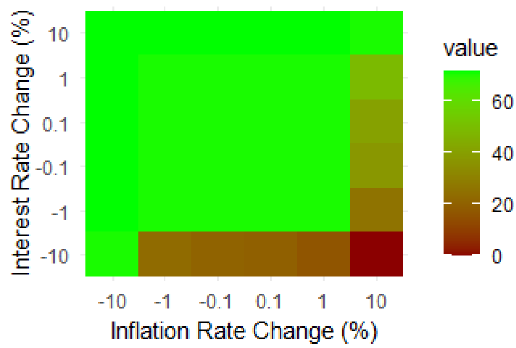
Table 3: Key Takeaway from Sensitivity Analysis with respect to Profit Count (out of 120)

Some qualitative risks of the program along with mitigating strategies include:

- There is a good chance of policyholders pretending to engage with the health incentives to access discounts, while not truly doing so, leading to no change in mortality and the underpricing of their premiums. Depending on the frequency of this occurring, this could lead to a significantly lower number of lives saved and a loss from mispricing. This



(a) Heatmap of
T20 Profit Percentage



(b) Heatmap of
SPWL Profit Percentage

risk can be easily controlled by setting up detailed checks for each health incentive to ensure the participating policyholders are genuinely engaging with the program. These checks can include lending digital devices to track policyholders' fitness activities and requiring them to provide signed attendance logs of their smoking cessation therapy sessions; the implementation of these checks is already considered in the input costs of our models.

- There is a moderate possibility of policyholders engaging with the health incentives short-term and then opting out. As long-term engagement is required for a noticeable impact on mortality, this would lead to the underpricing of premiums as these policyholders would be priced with lower mortality assumptions. A mitigation strategy for this risk is to define specific conditions for opting out, such as returning to a pre-discount regular premium for whole-life policies, and repayment of a portion of the discount depending on the length of remaining cover for term insurance.

Given the comprehensive nature of the program design, including diverse interventions targeting multiple risk factors, there's a moderate degree of certainty that the proposed program could have contributed to lowering mortality rates over the past 20 years. However, the actual impact may be influenced by various factors such as the effectiveness of interventions and the independence of variables not accounted for in the analysis. Therefore, while there's a reasonable expectation of mortality reduction, the degree of certainty is tempered by the complexity of real-world health interventions and the potential limitations of the analysis.

We are confident that the proposed SuperLifeStyle program will surpass the value of policies sold without it. Our sensitivity analysis, examining extreme variations in key variables, reveals insights into potential impacts. Notably, interest rates exert the most significant influence on calculated profits. A 10% decrease could eliminate profits, while a 10% increase maximizes them. Engagement rate variations have a minor effect, with no change in expected profits at a 10% sensitivity level. Despite uncertainties, our analysis underscores the program's efficacy and economic viability, substantiating its potential to reduce mortality and generate profits for SuperLife.

7 Ethical Considerations and Frameworks

The integration of health incentives with life insurance products introduces complexities that require ethical scrutiny.

- Incentivizing Healthy Behaviors:
  - Offering discounts for engaging in healthy behaviors seems beneficial on the surface, promoting positive lifestyle changes.
  - However, ethical considerations arise regarding the potential coercion involved. While discounts may incentivize participation, they may also pressure individuals to engage in behaviors they may not be ready or willing to undertake.
  - SuperLife must ensure transparency and respect individuals' autonomy by clearly communicating the implications of participation and offering opt-out options ([11]).
- Decreasing Expected Mortality:
  - The aim to reduce mortality rates through health incentives is laudable from a public health perspective, but the ethical implications of using mortality reduction as a metric for success should be carefully examined.
  - There's a risk of prioritizing financial gains over genuine improvements in individuals' well-being and SuperLife should ensure that the pursuit of reduced mortality is aligned with genuine health benefits for policyholders, rather than solely focusing on financial gains.
- Adding Economic Value to SuperLife:
  - The pursuit of economic value through mortality reduction and increased sales is a common business objective.
  - However, ethical concerns arise if economic gains come at the expense of fairness, transparency, or the well-being of policyholders.
  - SuperLife must ensure that economic gains are ethically obtained, balancing financial interests with social responsibility and ethical considerations.

Considering these ethical considerations, SuperLife should adopt ethical frameworks such as:

- *Utilitarianism*: Focusing on maximizing overall well-being, ensuring that the benefits of the SuperLifeStyle program outweigh any potential harms.
- *Deontology*: Emphasizing principles of autonomy, honesty, and respect for individuals' rights, guiding SuperLife to prioritize informed consent and transparency.
- *Virtue Ethics*: Encouraging SuperLife to cultivate virtues such as honesty, integrity, and compassion in its business practices, fostering long-term trust with customers ([13]).

Appendices

A Methodology

A.1 Incentive Ranking

Using the interventions dataset provided, the lower and upper bounds were extracted for both the impact on mortality rates and the per capita costs from the provided data. Subsequently, the code calculates the impact per cost for each intervention, which is determined by averaging the lower and upper bounds of both the impact on mortality and per capita cost and then dividing the average impact by the average cost. This metric provides a measure of the effectiveness of each intervention relative to its cost. Finally, the interventions are arranged in descending order based on their impact per cost, allowing for easy identification of interventions that offer the most favorable balance between effectiveness and cost-efficiency. We then selected interventions that were catered to different demographics based on the findings.

A.2 Mortality Model

The SuperLife inforce dataset was used to determine the mortality impacts based on the age groups of current policyholders who are still alive. The corresponding mortality rates were then matched using the Mortality Table of Lumaria. Neoplasms were identified as the largest case of death among policyholders. Thus, a loading was then computed to accurately weigh the effects of neoplasm across different ages. We observed that deaths due to neoplasms primarily occurred between the ages of 50-86. Similarly, it was also observed that smokers were a minority group with a mere 18% of the population being smokers. This information was then used to design the intervention program. This leads to several conditions and assumptions:

- All policyholders participate in the safety campaigns and fitness challenges.
- Policyholders over the age of 50 participate in the screening program.
- All smokers participate in the smoking cessation program .
- An engagement rate of 100% is targeted for each program.

Policyholders who fit into the criteria were given binary flags to signal their participation in the program. Then it is assumed that the impact on mortality is uniformly distributed for people of the same age group and a value is assigned based on their participation in the program.

The impact on mortality is then calculated as the cumulative impact of all interventions.

A.3 Economic Benefit Analysis Model

- **Modelling Process:** EPV's were calculated for each possible issue age (defined in the life table provided where the maximum attainable age is 120 years old). For each

Health Incentive	Impact on Mortality
Safety Campaigns	3% to 5% for all age groups
Community Fitness Challenges	4% to 5% for 25-44 year olds 3% to 4% for 45-59 year olds 2% to 3% for 60+ year olds
Preventative Screening	8% to 10% for 25-44 year olds 6% to 8% for 45-59 year olds 5% to 7% for 60+ year olds (A screening loading is applied to all rates)
Smoking Cessation Programs	12.5% to 50% for 25-44 year old smokers 11.5% to 25% for 45-59 year old smokers 7.5% to 11.5% for 60+ year old smokers (18% of adult Lumarians are smokers)

Table 4: Mortality Impact Used as Inputs to the Models

starting age we calculate the probability of survival and probability of death in the next year for each sequential age and discount Č1 for the number of years since policy issue.

- **Probability of Survival and Death:** Calculated using the provided Lumeria life table and an adjusted mortality table based on mortality reductions through our interventions.

- **Insurance Term:** Either 20 years since policy issue up to a maximum age of 120 for a 20-year term insurance product, or all years since policy issue up to a maximum age of 120 for a whole life product.

- **Discounting:** Interest rates based on historical data were used and adjusted for inflation to arrive at the real rate, which was used to discount all values in the calculations.

- **Single Payment EPV:** Probability of Survival up to current age x Probability of death in the next year x Č1 discounted for each year survived since issue. Sum across all ages in the term.

$$A_X = E(Z) = \sum_{k=0}^{\infty} v^{k+1} \cdot {}_tp_x \cdot q_{x+k} = \sum_{k=0}^{\infty} v^{k+1} \cdot {}_{k|}q_x$$

- **Annuity Due EPV:** Probability of Survival up to current age x EPV of Č1 payment. Sum across all years in the term.

$$\ddot{a}_x = \sum_{k=0}^{\infty} v^k \cdot {}_kp_x$$

- **Special EPV Cases:** For accurate pricing, intervention specific age limits were imposed on EPV calculations, for example this included an annuity only beginning at age 50, an annuity beginning at age 27 up to age 88, and annuity beginning at age 18.

- **Policy Baseline Benefits:** The standard Single Payment EPV value was used to calculate the EPV of a single payment on death and was multiplied by the average Face Value of each policy type (derived from the supplied Inforce dataset).
- **Expenses:** The respective Special Case EPVs were then multiplied by the respective costs associated with each intervention to get a value for the total discounted expenses associated with all interventions.
- **Engagement Rate:** We considered that not all policy holders would engage with SuperLifeStyle, as such for arriving at final values of economic benefit in Figure 3, we took 75% x Old EPV of Policy Benefit Payments + 25% New EPV of Policy Benefit Payments + Expenses. We considered that only 18% of the population counts as eligible for Smoking hence its engagement rate was 25% x 18%. In principle this means that 75% of customers do not experience any mortality reduction and can derive their cost EPV from the old mortality table, while the remaining 25% see a mortality reduction and will use the new mortality table but will have to consider extra expenses due to the interventions.
- **Other Considerations:** Combining policy EPVs in this way significantly simplified the Sensitivity Analysis process by eliminating the need for new life tables to be calculated at different engagement rates. A single new life table was calculated at 100% engagement and sum weighted by a 25% engagement rate of the new and old EPVs was taken as this would provide the same result.

We assume other product expenses and profit margins remain constant and can therefore be ignored in the EPV / Economic benefit calculation, as a result we only focus on a cost comparison that arises from the implementation of the SuperLifeStyle.

By calculating a comparison at each potential policy issue age, we could identify how profitable the interventions were and around which bounds profitability occurred.

Table 3 contains EPVs of the baseline case where Superlife makes a face value payment on death based on the old mortality table, and the case where policyholders engage with SuperLifeStyle at a 25% rate factoring in associated mortality reductions and extra expenses.

B Code

All of our coding was done exclusively on R, owing to it being open-source and can be found below or at https://www.github.com/VishNar007/ACTL5100.

B.1 Mortality Modelling

```
library(psych)
library('KMsurv')
library("survival")
library(tidyverse).
library(lubridate)
library(ggplot2)
library("survminer")
library("lifecontingencies")
library("wordcloud")
library("openxlsx")
```

```r
library(gridExtra)
library(dplyr)
library("pracma")
library("reshape2")
# Read the data from a CSV file
data <- read_csv("../Data/Processed Data/CLEANED_2024-srcsc-superlife-inforce-dataset.csv")
# Create a new column 'Lifespan' using the provided formula
data$Lifespan <- data$Issue.age + (2024 - data$Issue.year)
# Create a function to categorize age groups into 10-year intervals
get_age_group <- function(age) {
if (age >= 25 && age <= 44) {
return("25-44")
} else if (age >= 45 && age <= 59) {
return("45-59")
} else if (age >= 60) {
return("60+")
} else {
return("Other")
}
}
# Apply the function to create a new column 'Age_Group' based on 'Lifespan'
data$Age_Group <- sapply(data$Lifespan, get_age_group)
# Separate deceased and censored policyholders
deceased_data <- data[!is.na(data$Death.indicator), ]
censored_data <- data[is.na(data$Death.indicator), ]
policyholder_data <- censored_data[is.na(censored_data$Lapse.Indicator), ]
# Print the distribution of ages in each age group
age_group_distribution <- prop.table(table(data$Age_Group))
print(age_group_distribution)
#load incentives data
# Create a dataset in R based on the provided table. Cranked the table in chatgpt to hardcode it into a table
intervention_data <- data.frame(
Intervention_Name = c("Wellness Programs", "Fitness Tracking Incentives", "Smoking Cessation Programs",
"Annual Health Check-ups", "Telemedicine Services", "Healthy Eating Campaigns",
"Weight Management Programs", "Mental Health Support","Financial Planning Assistance",
"Educational Workshops", "Incentives for Vaccinations",
"Regular Dental Check-ups", "Vision Care Programs", "Safety Campaigns",
"Driving Safety Courses", "Heart Health Screenings", "Chronic Disease Management",
"Sleep Hygiene Programs", "Community Fitness Challenges", "Discounted Gym Memberships",
"Online Health Resources", "Personalized Health Plans", "Well-being Apps",
"Hydration Campaigns", "Sun Safety Awareness", "Emergency Preparedness Training",
"Social Connection Initiatives", "Holistic Stress Reduction",
"Financial Incentives for Healthy Behavior", "Genetic Testing",
"Alcohol Moderation Programs", "Environmental Wellness",
"Employee Assistance Programs", "Holistic Nutrition Education",
"Incentives for Preventive Screenings", "Holistic Health Assessments",
"Cancer Prevention Initiatives", "Community Gardens",
"Active Aging Programs", "Home Safety Inspections",
"Mindfulness Programs", "Parenting Support Services",
"Travel Safety Tips", "Financial Literacy Workshops","Hiking and Outdoor Activities Groups",
"Cognitive Health Programs",
"Art and Creativity Classes",
"Mind-Body Wellness Retreats",
"Incentives for Regular Medication Adherence",
"Ergonomic Workstation Assessments"),
Impact_on_Mortality = c("2-5% reduction", "3-6% reduction", "0-50% reduction",
"5-10% reduction", "3-5% reduction", "2-4% reduction",
"5-10% reduction", "3-8% reduction","2-4% reduction", "2-8% reduction",
"2-4% reduction", "2-3% reduction", "3-5% reduction",
"2-4% reduction", "5-10% reduction", "5-10% reduction",
"3-5% reduction", "2-5% reduction", "3-6% reduction",
"2-4% reduction", "3-6% reduction", "2-4% reduction",
"2-3% reduction", "2-4% reduction", "2-4% reduction",
"3-5% reduction", "3-8% reduction", "2-5% reduction",
"2-4% reduction", "3-6% reduction", "2-4% reduction",
"2-4% reduction", "3-5% reduction", "5-10% reduction",
"3-6% reduction", "5-10% reduction", "2-4% reduction",
"3-6% reduction", "3-5% reduction", "3-8% reduction",
"2-4% reduction", "2-4% reduction", "2-4% reduction",
"3-6% reduction", "3-6% reduction", "2-4% reduction",
"3-6% reduction", "2-5% reduction", "2-4% reduction"
),
Per_Capita_Cost = c("C90-C345 per year", "C35-C175 per tracker", "C870-C3485 per participant",
"C175-C870 per check-up", "C50-C175 per consultation", "C10-C35 per participant",
"C175-C870 per program", "C90-C345 per counseling session", "C90-C345 per session",
"C20-C85 per workshop", "C20-C85 per incentive", "C90-C345 per check-up",
"C90-C345 per participant", "C10-C35 per participant", "C85-C175 per course",
"C90-C345 per screening", "C175-C870 per program", "C20-C85 per program",
"C10-C35 per participant", "C175-C870 per membership", "C10-C35 per participant",
"C90-C345 per plan", "C10-C35 per app", "C10-C35 per campaign",
"C10-C35 per campaign", "C20-C85 per training session", "C10-C35 per social event",
"C20-C85 per session", "C20-C85 per incentive", "C90-C345 per test",
"C90-C345 per program", "C10-C35 per campaign", "C90-C345 per counseling session",
"C20-C85 per session", "C20-C85 per incentive", "C90-C345 per assessment",
"C20-C85 per initiative", "C10-C35 per garden plot", "C20-C85 per program",
"C20-C85 per inspection", "C20-C85 per session", "C10-C35 per session",
"C10-C35 per campaign", "C20-C85 per workshop", "C20-C85 per group",
"C20-C85 per program", "C10-C35 per class", "C90-C345 per retreat",
"C20-C85 per incentive", "C20-C85 per assessment"))
# Extracting lower_bound and upper_bound
mort_lower_bound <- as.numeric(gsub("%.*", "", sapply(strsplit(intervention_data$Impact_on_Mortality, "-"), `[`, 1)))
mort_upper_bound <- as.numeric(gsub("%.*", "", sapply(strsplit(intervention_data$Impact_on_Mortality, "-"), `[`, 2)))
```

```
# Extracting lower_bound and upper_bound
c_lower_bound <- as.numeric(gsub("C| per.*", "", sapply(strsplit(intervention_data$Per_Capita_Cost, "-"), `[`, 1)))
c_upper_bound <- as.numeric(gsub("C| per.*", "", sapply(strsplit(intervention_data$Per_Capita_Cost, "-"), `[`, 2)))
# Apply the function to extract lower and upper bounds for Per_Capita_Cost
# Impact per cost is determined as the average impact / average cost
intervention_data <- intervention_data %>%
mutate(Mort_Lower_Bound = mort_lower_bound,
Mort_Upper_Bound = mort_upper_bound,
Cost_Lower_Bound = c_lower_bound,
Cost_Upper_Bound = c_upper_bound,
impact_per_cost = ((Mort_Lower_Bound + Mort_Upper_Bound) / 2) / ((Cost_Lower_Bound + Cost_Upper_Bound) / 2))%>%
arrange(desc(impact_per_cost))
############################################
# Age group percentages
age_group_percentages <- c("25-34" = 0.20, "35-54" = 0.5, "55-64" = 0.75, "65+" = 0.9)
# Create a function to calculate impact per cost for each age group
calculate_impact_per_cost_age <- function(intervention_data, age_group_percentage) {
impact_per_cost_age <- (intervention_data$Mort_Upper_Bound - age_group_percentage *
(intervention_data$Mort_Upper_Bound - intervention_data$Mort_Lower_Bound)) /
(intervention_data$Cost_Lower_Bound + (1-age_group_percentage) *
(intervention_data$Cost_Upper_Bound - intervention_data$Cost_Lower_Bound))
return(impact_per_cost_age)
}
# Create a list to store results for each age group
result_data_list <- list()
# Calculate impact per cost for each age group
for (age_group in names(age_group_percentages)) {
age_group_percentage <- age_group_percentages[age_group]
intervention_data_age <- intervention_data %>%
mutate(impact_per_cost_age = calculate_impact_per_cost_age(., age_group_percentage),
Age_Group = age_group) %>%
arrange(desc(impact_per_cost_age))
# Store the result for the current age group in the list
result_data_list[[age_group]] <- intervention_data_age %>% select(Intervention_Name, Age_Group, impact_per_cost_age)
}
################## mortality modelling ########################
mortality_table <- read_excel("../Data/Case Study Data/srcsc-2024-lumaria-mortality-table.xlsx", skip = 13)
mortality_table <- mortality_table[, 1:2]
# Determine the chosen interventions
chosen_interventions <- intervention_data %>%
filter(Intervention_Name %in% c("Safety Campaigns", "Community Fitness Challenges", "Screening Aging Programs",
"Smoking Cessation Programs"))
# Merge the data frames based on Age_Group
policyholder_mortality <- merge(policyholder_data, mortality_table, by.x = "Lifespan", by.y = "Age", all.x = TRUE)
# Load neoplasms mortality loading data
neoplasms_loading <- read.csv("../Processed Data/Case Study Data/Neoplasm_Mortality_Loading.csv")
# Create a sequence of ages from 27 to 88
age_range <- data.frame(Age.at.Death = 27:88)
# Merge with original data, replacing missing values with zeros
neoplasms_loading <- merge(age_range, neoplasms_loading, by = "Age.at.Death", all.x = TRUE)
neoplasms_loading[is.na(neoplasms_loading)] <- 0
policyholder_mortality <- merge(policyholder_mortality, neoplasms_loading, by.x = "Lifespan",
by.y = "Age.at.Death", all.x = TRUE)
# Rename the new column
colnames(policyholder_mortality)[ncol(policyholder_mortality)] <- "Screening.Weight"
# Program design
# All policyholders participate in the safety campaigns
policyholder_mortality$Safety_flag <- 1
# All policyholders participate in fitness challenges
policyholder_mortality$Fitness_flag <- 1
# All policyholders over the age of 40 participate in the screening program
policyholder_mortality$Screening_flag <- ifelse(policyholder_mortality$Lifespan >= 40, 1, 0)
# All smokers participate in the smoking cessation program
policyholder_mortality$Smoking_flag <- ifelse(policyholder_mortality$Smoker.Status == "S", 1, 0)
# Set an engagement rate of 25%
engagement_rate <- 1
# Function to adjust flags based on engagement rate
adjust_flag <- function(flag) {
num_engaged <- sum(flag == 1)
num_to_engage <- round(num_engaged * engagement_rate)
indices <- sample(which(flag == 1), num_to_engage)
new_flag <- rep(0, length(flag))
new_flag[indices] <- 1
return(new_flag)
}
# Adjust flags based on engagement rate
policyholder_mortality$Safety_flag <- adjust_flag(policyholder_mortality$Safety_flag)
policyholder_mortality$Fitness_flag <- adjust_flag(policyholder_mortality$Fitness_flag)
policyholder_mortality$Screening_flag <- adjust_flag(policyholder_mortality$Screening_flag)
policyholder_mortality$Smoking_flag <- adjust_flag(policyholder_mortality$Smoking_flag)
# Calculate the sum of all flags for each policyholder
policyholder_mortality$Total_flags <- rowSums(policyholder_mortality[, c("Safety_flag", "Fitness_flag",
"Screening_flag", "Smoking_flag")])
# Count the number of policyholders with zero total flags
num_not_engaged <- sum(policyholder_mortality$Total_flags == 0)
# Calculate the proportion of policyholders that do not engage in any flag
proportion_not_engaged <- num_not_engaged / nrow(policyholder_mortality)
# For reproducibility
set.seed(42)
# Generate random values for the dummy flags
policyholder_mortality$safety_dummy <- ifelse(policyholder_mortality$Safety_flag == 1,
runif(nrow(policyholder_mortality), 0.03, 0.05), 0)
```

```
# Setting mortality rates for different age groups
policyholder_mortality$fitness_dummy <- ifelse(policyholder_mortality$Fitness_flag == 1,
#4% to 5% for 25-44 year olds
ifelse(policyholder_mortality$Age_Group == "25-44",
runif(nrow(policyholder_mortality), 0.04, 0.05),
#3% to 4% for 45-59 year olds
ifelse(policyholder_mortality$Age_Group == "45-59",
runif(nrow(policyholder_mortality), 0.03, 0.04),
#2% to 3% for 60+ year olds
ifelse(policyholder_mortality$Age_Group == "60+",
runif(nrow(policyholder_mortality), 0.02, 0.03),
0))),
0)
# Setting mortality rates for screenings of different age groups
policyholder_mortality$screening_dummy <- ifelse(policyholder_mortality$Screening_flag == 1,
# 8% to 10% for 25-44 year olds
ifelse(policyholder_mortality$Age_Group == "25-44",
runif(nrow(policyholder_mortality), 0.08, 0.1) * policyholder_mortality$Screening.Weight,
# 6% to 8% for 45-59 year olds
ifelse(policyholder_mortality$Age_Group == "45-59",
runif(nrow(policyholder_mortality), 0.06, 0.08) * policyholder_mortality$Screening.Weight,
# 5% to 7% for 60+ year olds
ifelse(policyholder_mortality$Age_Group == "60+",
runif(nrow(policyholder_mortality), 0.05, 0.07) * policyholder_mortality$Screening.Weight,
0))),
0)
# Setting mortality rates for smokers of different age groups
policyholder_mortality$smoking_dummy <- ifelse(policyholder_mortality$Smoking_flag == 1,
#12.5% to 50% for 25-44 year olds
ifelse(policyholder_mortality$Age_Group == "25-44",
runif(nrow(policyholder_mortality), 0.125, 0.5),
#11.5% to 25% for 45-59 year olds
ifelse(policyholder_mortality$Age_Group == "45-59",
runif(nrow(policyholder_mortality), 0.115, 0.25),
#7.5% to 11.5% for 60+ year olds
ifelse(policyholder_mortality$Age_Group == "60+",
runif(nrow(policyholder_mortality), 0.075, 0.115),
0))),
0)
# Calculate the product of (1 - dummy) values for mortality_impact
policyholder_mortality$mortality_impact <- (1 - policyholder_mortality$safety_dummy) *
(1 - policyholder_mortality$fitness_dummy) *
(1 - policyholder_mortality$screening_dummy) *
(1 - policyholder_mortality$smoking_dummy)
policyholder_mortality$improved_mortality <- policyholder_mortality$mortality_impact *
policyholder_mortality$Mortality.Rate
policyholder_mortality$mortality_reduction <- policyholder_mortality$Mortality.Rate -
policyholder_mortality$improved_mortality
# Group by Lifespan and calculate the averages
average_mortality_table <- policyholder_mortality %>%
group_by(Lifespan) %>%
summarise(new_mortality = mean(improved_mortality, na.rm = TRUE),
average_mortality_reduction = mean(mortality_reduction, na.rm = TRUE),
old_mortality= mean(Mortality.Rate) )
# Calculate new mortality 20 years into the future
average_mortality_table$new_mortality_20_years_future <- average_mortality_table$new_mortality *
(1 - average_mortality_table$average_mortality_reduction) ^ 20
# Generate a sequence of ages from 1 to 120
all_ages <- data.frame(Age = 1:120)
# Expand the mortality table to include ages 1 to 120
average_mortality_table_expanded <- merge(all_ages, average_mortality_table,
by.x = "Age", by.y = "Lifespan", all.x = TRUE)
# Replace new_mortality with mortality.rate from mortality_table
average_mortality_table_expanded$old_mortality
<- mortality_table$Mortality.Rate[match(average_mortality_table_expanded$Age, mortality_table$Age)]
# Replace new_mortality with old_mortality for ages 1 to 26 and 89 to 120
average_mortality_table_expanded$new_mortality
<- ifelse(average_mortality_table_expanded$Age <= 26 |
average_mortality_table_expanded$Age >= 89,
mortality_table$Mortality.Rate[match(average_mortality_table_expanded$Age, mortality_table$Age)],
mortality_table_expanded$new_mortality)
write.xlsx(average_mortality_table_expanded, file = "average_mortality_table.xlsx", rowNames = FALSE)
# Calculate lives saved before running the program
average_mortality_table$old_deaths <- average_mortality_table$old_mortality * nrow(policyholder_mortality)
average_mortality_table$new_deaths <- average_mortality_table$new_mortality * nrow(policyholder_mortality)
average_mortality_table$lives_saved_now <- average_mortality_table$old_deaths - average_mortality_table$new_deaths
# Calculate lives saved 20 years into the future
average_mortality_table$new_mortality_future <- average_mortality_table$new_mortality *
(1 - average_mortality_table$average_mortality_reduction) ^ 20
average_mortality_table$new_deaths_future <-
average_mortality_table$new_mortality_future * nrow(policyholder_mortality)
average_mortality_table$lives_saved_future <-
average_mortality_table$old_deaths - average_mortality_table$new_deaths_future
# Create a new dataframe for plotting
lives_saved_plot <- data.frame(
Lifespan = average_mortality_table$Lifespan,
Lives_Saved_Now = average_mortality_table$lives_saved_now,
Lives_Saved_Future = average_mortality_table$lives_saved_future
)
# Melt the dataframe for easier plotting
lives_saved_plot <- melt(lives_saved_plot, id.vars = "Lifespan", variable.name = "Scenario", value.name = "Lives_Saved")
# Plot both curves on one plot
```

```
ggplot(lives_saved_plot, aes(x = Lifespan, y = Lives_Saved, color = Scenario)) +
geom_line() +
geom_point() +
labs(title = "Comparison of Lives Saved Now implementation and 20 Years Prior",
x = "Lifespan",
y = "Number of Lives Saved",
color = "Scenario") +
scale_color_manual(values = c("red", "blue")) # Adjust colors as needed
# Group by Lifespan and calculate the averages
cost_table_T20 <- policyholder_mortality %>%
filter(Policy.type == "T20")%>%
group_by(Lifespan) %>%
summarise(average_cost = mean(Face.amount, na.rm = TRUE) )
# Group by Lifespan and calculate the averages
cost_table_SPWL <- policyholder_mortality %>%
filter(Policy.type == "SPWL")%>%
group_by(Lifespan) %>%
summarise(average_cost = mean(Face.amount, na.rm = TRUE) )
# Merge average_mortality_table with cost_table_T20 based on Lifespan
merged_data_T20 <- merge(average_mortality_table, cost_table_T20, by = "Lifespan")
# Merge average_mortality_table with cost_table_SPWL based on Lifespan
merged_data_SPWL <- merge(average_mortality_table, cost_table_SPWL, by = "Lifespan")
# Calculate the claims cost reduction for T20 policies
cost_reduction_now_T20 <- sum(merged_data_T20$lives_saved_now * merged_data_T20$average_cost)
cost_reduction_future_T20 <- sum(merged_data_T20$lives_saved_future * merged_data_T20$average_cost)
# Calculate the claims cost reduction for SPWL policies
cost_reduction_now_SPWL <- sum(merged_data_SPWL$lives_saved_now * merged_data_SPWL$average_cost)
cost_reduction_future_SPWL <- sum(merged_data_SPWL$lives_saved_future * merged_data_SPWL$average_cost)
# Print the claims cost reduction for T20 policies
print(paste("Claims Cost Reduction Now (T20):", cost_reduction_now_T20))
print(paste("Claims Cost Reduction Future (T20):", cost_reduction_future_T20))
# Print the claims cost reduction for SPWL policies
print(paste("Claims Cost Reduction Now (SPWL):", cost_reduction_now_SPWL))
print(paste("Claims Cost Reduction Future (SPWL):", cost_reduction_future_SPWL))
# Create a data frame for plotting
cost_reduction_data <- data.frame(
Policy_Type = c("T20", "T20", "SPWL", "SPWL"),
Scenario = c("Now", "20 years prior", "Now", "20 years prior"),
Claims_Cost_Reduction = c(cost_reduction_now_T20, cost_reduction_future_T20,
cost_reduction_now_SPWL, cost_reduction_future_SPWL)
)
# Create a ggplot
ggplot(cost_reduction_data, aes(x = Policy_Type, y = Claims_Cost_Reduction, fill = Scenario)) +
geom_bar(stat = "identity", position = "dodge") +
labs(title = "Claims Cost Reduction for T20 and SPWL Policies",
x = "Policy Type",
y = "Claims Cost Reduction") +
scale_fill_manual(values = c("Now" = "blue", "20 years prior" = "red")) +
theme_minimal()
```

C Tables

Age	T20_baseline	T20_intervention	SPWL_baseline	SPWL_intervention
1	2674.10	2697.83	278293.77	278205.69
2	2752.80	2782.68	281828.19	281738.94
3	2891.68	2927.87	285457.76	285367.34
4	3069.47	3111.99	289165.17	289073.58
5	3267.97	3316.85	292933.37	292840.51
6	3480.78	3528.43	296757.10	296663.04
7	3708.57	3825.84	300636.57	300541.26
8	3952.19	4068.65	304572.58	304475.95
9	4209.64	4325.07	308562.49	308464.59
10	4485.02	4599.33	312606.51	312507.30
11	4773.46	4886.17	316699.39	316598.85
12	5077.28	5187.77	320843.54	320741.66
13	5396.14	5503.83	325036.27	324933.05
14	5728.88	5833.38	329276.71	329172.09
15	6069.83	6170.66	333558.62	333452.58
16	6422.44	6518.61	337878.69	337771.17
17	6788.71	6879.41	342235.13	342126.11
18	7164.13	7248.28	346624.81	346514.31
19	7559.11	7627.92	351054.70	350934.92
20	7987.75	8040.03	355530.53	355401.41
21	8449.77	8483.96	360053.03	359914.38
22	8959.29	8973.93	364627.82	364479.50
23	9524.24	9517.65	369257.95	369099.89
24	10161.10	10135.07	373951.32	373783.30
25	10874.46	10827.65	378708.56	378530.52
26	11672.58	11603.22	383527.61	383339.35
27	12567.37	12483.69	388408.70	388216.70
28	13563.94	13462.44	393351.96	393154.73
29	14671.97	14550.49	398357.95	398155.70
30	15897.40	15766.10	403426.00	403218.82
31	17252.76	17108.82	408553.56	408341.30
32	18754.16	18595.20	413739.55	413522.39
33	20418.04	20241.62	418985.19	418763.48
34	22263.13	22066.30	424288.62	424062.57
35	24313.69	24092.82	429650.40	429420.10
36	26574.80	26326.18	435069.07	434834.64
37	29050.54	28770.69	440539.41	440301.29
38	31757.36	31442.34	446058.92	445817.46
39	34744.68	34399.80	451628.54	451384.31
40	38029.59	37650.62	457248.77	457002.16
41	41599.84	41183.36	462916.58	462668.13
42	45486.35	45028.48	468634.19	468384.50
43	49700.03	49196.51	474399.01	474148.65

44	54273.67	53720.69	480210.42	479960.12
45	59258.52	58647.21	486067.24	485815.31
46	64673.97	63999.42	491969.21	491716.15
47	70528.23	69785.80	497911.57	497657.97
48	76861.22	76045.93	503889.83	503636.38
49	83726.02	82831.85	509903.82	509651.36
50	91186.88	90209.85	515951.35	515700.99
51	99322.09	98239.68	522031.58	521767.66
52	108206.75	107012.39	528138.62	527862.00
53	117956.15	116643.16	534266.53	533978.04
54	128664.26	127225.59	540411.75	540112.32
55	140415.80	138844.45	546567.44	546258.08
56	153303.82	151592.46	552726.75	552408.61
57	167381.38	165525.40	558891.24	558565.45
58	182673.15	180669.42	565060.00	564727.76
59	199141.52	196989.55	571228.97	570891.53
60	216715.20	214409.17	577382.51	577036.98
61	235375.28	232921.87	583520.33	583167.96
62	255088.05	252498.75	589652.36	589294.10
63	275760.67	273052.18	595774.37	595411.33
64	297168.90	294365.97	601887.45	601520.75
65	318972.67	316104.57	607985.91	607616.55
66	340906.59	338009.10	614057.47	613686.53
67	362690.32	359798.36	620103.76	619732.29
68	384077.26	381580.46	626127.79	625756.75
69	404802.43	402680.31	632121.99	631752.45
70	424428.47	422651.68	638078.30	637711.73
71	442744.65	441278.89	643984.73	643622.04
72	459547.90	458356.01	649826.83	649469.40
73	474700.82	473744.16	655589.06	655238.54
74	488175.18	487416.05	661244.30	660902.44
75	499979.75	499382.42	666773.93	666442.80
76	510212.54	509745.11	672159.50	671841.46
77	519006.04	518641.39	677378.55	677076.54
78	526518.32	526233.64	682421.78	682138.61
79	532937.69	532715.46	687281.96	687020.83
80	538450.84	538278.33	691961.62	691726.09
81	543231.55	543100.45	696460.28	696254.30
82	547428.00	547334.09	700764.67	700592.95
83	551157.34	551099.68	704860.61	704728.56
84	554503.48	554484.26	708728.95	708643.45
85	557543.81	557567.34	712375.61	712344.53
86	560350.43	560423.46	715827.97	715861.16
87	562956.01	563087.73	719085.06	719194.85

88	565385.12	565589.32	722151.41	722355.62
89	567641.17	567745.68	725015.33	725119.84
90	569715.22	569813.48	727656.16	727754.42
91	571653.77	571746.18	730128.54	730220.95
92	573454.74	573541.71	732427.33	732514.29
93	575129.30	575211.20	734565.56	734647.46
94	576695.44	576772.62	736565.68	736642.86
95	578148.48	578221.26	738421.47	738494.25
96	579505.20	579573.88	740154.28	740222.96
97	580764.99	580829.86	741763.30	741828.17
98	581922.47	581983.85	743241.66	743303.03
99	582990.27	583048.42	744605.47	744663.61
100	583962.57	584017.78	745847.31	745902.52
101	584848.75	584901.28	746979.15	747031.67
102	585671.12	585721.17	748029.50	748079.54
103	586432.42	586480.17	749001.84	749049.59
104	587135.89	587181.50	749900.32	749945.94
105	587785.80	587829.46	750730.41	750774.06
106	588388.48	588430.31	751500.15	751541.98
107	588949.68	588989.81	752216.93	752257.06
108	589474.69	589513.24	752887.49	752926.04
109	589975.11	590012.15	753526.63	753563.67
110	590481.92	590517.43	754173.94	754209.44
111	590947.96	590982.06	754769.17	754803.27
112	591377.33	591410.13	755317.57	755350.36
113	591773.68	591805.28	755823.79	755855.39
114	592140.49	592170.98	756292.29	756322.78
115	592482.26	592511.72	756728.80	756758.26
116	592803.57	592832.05	757139.18	757167.67
117	593112.45	593140.01	757533.70	757561.25
118	593459.18	593485.68	757976.54	758003.04
119	594329.39	594353.27	759087.99	759111.87
120	0.00	0.00	0.00	0.00

Figure 3: Baseline and Intervention-Loaded Policy Death Claims

Age_bracket	T20_baseline	T20_intervention	SPWL_baseline	SPWL_intervention
01-22	5221	5297	319738	319633
23-44	25143	24917	422769	422550
45-59	116220	114951	528334	528050
60-84	428183	426795	647445	647135
85+	582268	582327	743699	743754

Figure 4: Benefits by Age Group

Age	new_mortality	Age	new_mortality	Age	new_mortality	Age	new_mortality
1	0.003547	31	5.75E-04	61	0.007706137	91	0.184358
2	3.37E-04	32	6.08E-04	62	0.008425002	92	0.201508
3	2.40E-04	33	6.39E-04	63	0.009214067	93	0.219478
4	1.80E-04	34	6.79E-04	64	0.010052216	94	0.238045
5	1.58E-04	35	7.24E-04	65	0.010990504	95	0.257967
6	1.47E-04	36	7.77E-04	66	0.012074901	96	0.278455
7	1.38E-04	37	8.45E-04	67	0.013232364	97	0.300155
8	1.29E-04	38	9.27E-04	68	0.014464646	98	0.323127
9	1.26E-04	39	0.001008566	69	0.015850388	99	0.346523
10	1.25E-04	40	0.001096023	70	0.017403031	100	0.370973
11	1.37E-04	41	0.001196916	71	0.019203825	101	0.395588
12	1.45E-04	42	0.001295947	72	0.021268988	102	0.4198
13	1.61E-04	43	0.001410663	73	0.023647656	103	0.444591
14	1.81E-04	44	0.00153215	74	0.026488957	104	0.469816
15	2.17E-04	45	0.001699073	75	0.029761824	105	0.495312
16	2.63E-04	46	0.001845127	76	0.033518483	106	0.520918
17	3.15E-04	47	0.002016109	77	0.037847203	107	0.54666
18	3.76E-04	48	0.002212545	78	0.042705082	108	0.572585
19	4.24E-04	49	0.002418628	79	0.048102506	109	0.598546
20	4.59E-04	50	0.002641412	80	0.053989503	110	0.624392
21	4.96E-04	51	0.002884964	81	0.060457979	111	0.654399
22	5.21E-04	52	0.003162655	82	0.067734641	112	0.684243
23	5.40E-04	53	0.003482268	83	0.075925129	113	0.713887
24	5.43E-04	54	0.003834359	84	0.085157308	114	0.743294
25	5.45E-04	55	0.004235977	85	0.095163417	115	0.772428
26	5.56E-04	56	0.004690341	86	0.105601516	116	0.801414
27	5.00E-04	57	0.005164575	87	0.117030164	117	0.830399
28	5.15E-04	58	0.005670861	88	0.129267724	118	0.859385
29	5.27E-04	59	0.006228851	89	0.152489	119	0.888371
30	5.45E-04	60	0.006987103	90	0.168478	120	1

Figure 5: Post Interventions Life Table

Data Cleaning

Helitha Dharmadasa - z5451805

2024-02-23

```
library(tidyverse)
```

Read Data

```
superlife_df <- read_csv("../Data/Case Study Data/2024-srcsc-superlife-inforce-dataset.csv",
    skip = 3)
```

```
## Rows: 978582 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): Policy.number, Policy.type, Sex, Smoker.Status, Underwriting.Class,...
## dbl (7): Issue.year, Issue.age, Face.amount, Region, Death.indicator, Year.o...
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
cause_of_death_map <- read_csv("../Data/External Data/superlife_inforce_causes_of_death.csv")
```

```
## Rows: 17 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Unique.Cause.of.Death, Description
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
cause_of_death_map
```

```
## # A tibble: 17 x 2
##    Unique.Cause.of.Death Description
##    <chr>                 <chr>
##  1 A00-B99               Certain infectious and parasitic diseases
##  2 C00-D48               Neoplasms
##  3 D50-D89               Diseases of the blood and blood-forming organs and cer~
##  4 E00-E88               Endocrine, nutritional and metabolic diseases
##  5 F01-F99               Mental and behavioural disorders
##  6 G00-G98               Diseases of the nervous system
```

```
##  7 I00-I99              Diseases of the circulatory system
##  8 J00-J98              Diseases of the respiratory system
##  9 K00-K92              Diseases of the digestive system
## 10 L00-L98              Diseases of the skin and subcutaneous tissue
## 11 M00-M99              Diseases of the musculoskeletal system and connective ~
## 12 N00-N98              Diseases of the genitourinary system
## 13 <NA>                 <NA>
## 14 O00-O99              Pregnancy, childbirth and the puerperium
## 15 Q00-Q99              Congenital malformations, deformations and chromosomal~
## 16 R00-R99              Symptoms, signs and abnormal clinical and laboratory f~
## 17 V01-Y89              External causes of morbidity and mortality
```

```
head(superlife_df)
```

```
## # A tibble: 6 x 16
##   Policy.number Policy.type Issue.year Issue.age Sex   Face.amount Smoker.Status
##   <chr>         <chr>            <dbl>     <dbl> <chr>       <dbl> <chr>
## 1 O8FN6OR4KXIS  T20               2001        54 F          100000 NS
## 2 K0JK2XD81ZNI  SPWL              2001        54 M         1000000 NS
## 3 AH3A98MHTO8H  T20               2001        27 F           50000 NS
## 4 C9QPJMIH8H9Y  T20               2001        55 F         2000000 NS
## 5 2C1HL2XQOWME  T20               2001        39 F          250000 NS
## 6 LKW7MA7BPAV1  SPWL              2001        41 M         2000000 NS
## # i 9 more variables: Underwriting.Class <chr>, Urban.vs.Rural <chr>,
## #   Region <dbl>, Distribution.Channel <chr>, Death.indicator <dbl>,
## #   Year.of.Death <dbl>, Lapse.Indicator <chr>, Year.of.Lapse <dbl>,
## #   Cause.of.Death <chr>
```

```
summary(superlife_df)
```

```
##  Policy.number      Policy.type          Issue.year     Issue.age
##  Length:978582      Length:978582      Min.   :2001   Min.   :26.0
##  Class :character   Class :character   1st Qu.:2009   1st Qu.:36.0
##  Mode  :character   Mode  :character   Median :2015   Median :44.0
##                                        Mean   :2014   Mean   :44.1
##                                        3rd Qu.:2020   3rd Qu.:52.0
##                                        Max.   :2023   Max.   :65.0
##
##      Sex             Face.amount      Smoker.Status      Underwriting.Class
##  Length:978582      Min.   :  50000   Length:978582      Length:978582
##  Class :character   1st Qu.: 100000   Class :character   Class :character
##  Mode  :character   Median : 500000   Mode  :character   Mode  :character
##                     Mean   : 665574
##                     3rd Qu.:1000000
##                     Max.   :2000000
##
##  Urban.vs.Rural         Region      Distribution.Channel Death.indicator
##  Length:978582      Min.   :1.000   Length:978582        Min.   :1
##  Class :character   1st Qu.:1.000   Class :character     1st Qu.:1
##  Mode  :character   Median :2.000   Mode  :character     Median :1
##                     Mean   :2.748                        Mean   :1
##                     3rd Qu.:4.000                        3rd Qu.:1
##                     Max.   :6.000                        Max.   :1
```

```
##                                                          NA's   :938206
##   Year.of.Death    Lapse.Indicator    Year.of.Lapse    Cause.of.Death
##   Min.   :2001     Length:978582      Min.   :2001     Length:978582
##   1st Qu.:2015     Class :character   1st Qu.:2017     Class :character
##   Median :2019     Mode  :character   Median :2021     Mode  :character
##   Mean   :2018                        Mean   :2019
##   3rd Qu.:2021                        3rd Qu.:2022
##   Max.   :2023                        Max.   :2023
##   NA's   :938206                      NA's   :867693
```

Clean & Transform Data

```
superlife_df <- superlife_df %>%
    mutate(Lapse.Indicator = ifelse(Lapse.Indicator == "Y", 1, Lapse.Indicator),
        Age.at.Death = Year.of.Death - Issue.year + Issue.age)
```

```
# Join Cause of death desc. with main dataset
superlife_df <- left_join(superlife_df, cause_of_death_map, by = c(Cause.of.Death = "Unique.Cause.of.Dea

superlife_df <- superlife_df %>%
    rename(Cause.of.Death.Description = "Description")
```

```
# Write cleaned data
write_csv(superlife_df, "../Data/Processed Data/CLEANED_2024-srcsc-superlife-inforce-dataset.csv")
```

Data Visualisation and Neoplasm Loading

Helitha Dharmadasa - z5451805

2024-02-23

```r
library(tidyverse)
```

```r
superlife_df <- read_csv("../Data/Processed Data/CLEANED_2024-srcsc-superlife-inforce-dataset.csv")
```

```
## Rows: 978582 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): Policy.number, Policy.type, Sex, Smoker.Status, Underwriting.Class,...
## dbl (9): Issue.year, Issue.age, Face.amount, Region, Death.indicator, Year.o...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
head(superlife_df)
```

```
## # A tibble: 6 x 18
##   Policy.number Policy.type Issue.year Issue.age Sex   Face.amount Smoker.Status
##   <chr>         <chr>            <dbl>     <dbl> <chr>       <dbl> <chr>
## 1 O8FN60R4KXIS  T20               2001        54 F          100000 NS
## 2 K0JK2XD81ZNI  SPWL              2001        54 M         1000000 NS
## 3 AH3A98MHTO8H  T20               2001        27 F           50000 NS
## 4 C9QPJMIH8H9Y  T20               2001        55 F         2000000 NS
## 5 2C1HL2XQOWME  T20               2001        39 F          250000 NS
## 6 LKW7MA7BPAV1  SPWL              2001        41 M         2000000 NS
## # i 11 more variables: Underwriting.Class <chr>, Urban.vs.Rural <chr>,
## #   Region <dbl>, Distribution.Channel <chr>, Death.indicator <dbl>,
## #   Year.of.Death <dbl>, Lapse.Indicator <dbl>, Year.of.Lapse <dbl>,
## #   Cause.of.Death <chr>, Age.at.Death <dbl>, Cause.of.Death.Description <chr>
```

```r
summary(superlife_df)
```

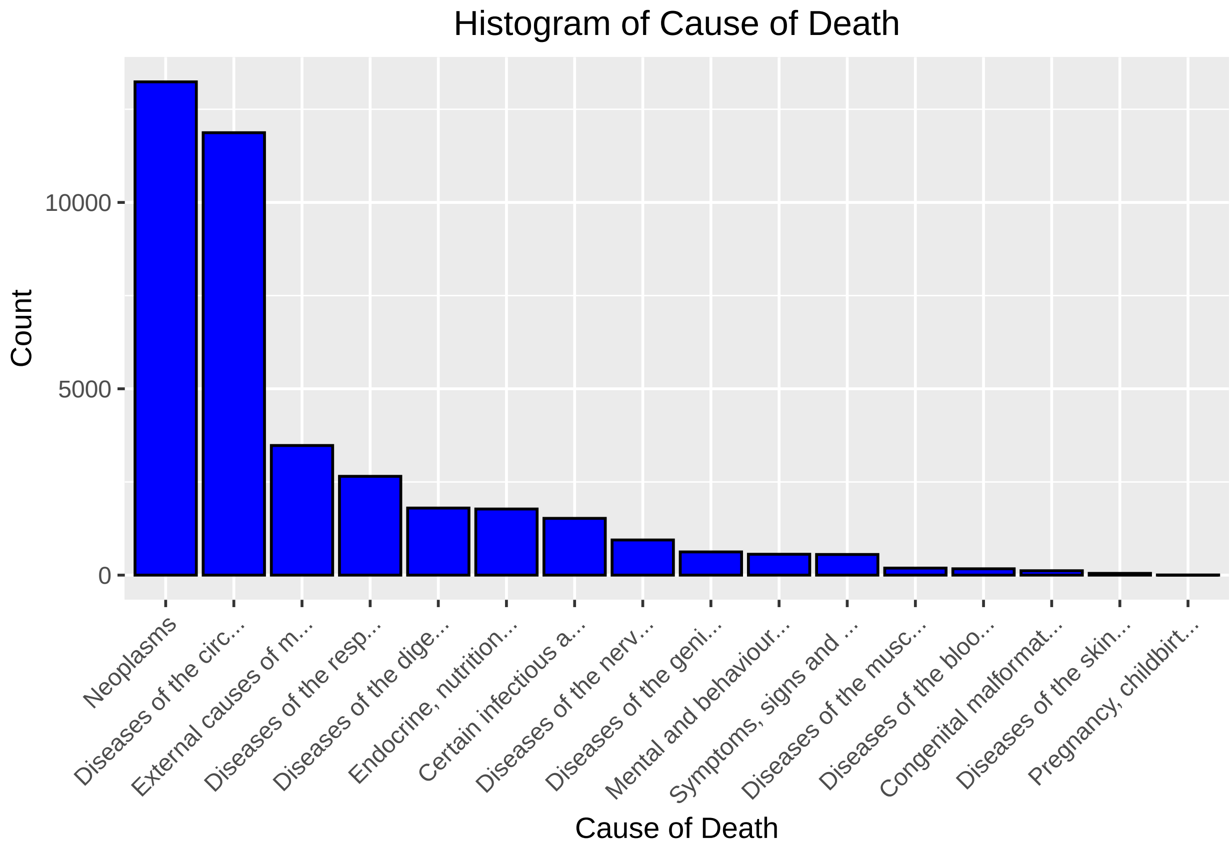
```
##  Policy.number      Policy.type          Issue.year     Issue.age
##  Length:978582      Length:978582      Min.   :2001   Min.   :26.0
##  Class :character   Class :character   1st Qu.:2009   1st Qu.:36.0
##  Mode  :character   Mode  :character   Median :2015   Median :44.0
##                                        Mean   :2014   Mean   :44.1
##                                        3rd Qu.:2020   3rd Qu.:52.0
##                                        Max.   :2023   Max.   :65.0
##
```

```
##       Sex              Face.amount      Smoker.Status      Underwriting.Class
##  Length:978582      Min.   :  50000   Length:978582      Length:978582
##  Class :character   1st Qu.: 100000   Class :character   Class :character
##  Mode  :character   Median : 500000   Mode  :character   Mode  :character
##                     Mean   : 665574
##                     3rd Qu.:1000000
##                     Max.   :2000000
##
##  Urban.vs.Rural         Region      Distribution.Channel Death.indicator
##  Length:978582      Min.   :1.000   Length:978582        Min.   :1
##  Class :character   1st Qu.:1.000   Class :character     1st Qu.:1
##  Mode  :character   Median :2.000   Mode  :character     Median :1
##                     Mean   :2.748                        Mean   :1
##                     3rd Qu.:4.000                        3rd Qu.:1
##                     Max.   :6.000                        Max.   :1
##                                                          NA's   :938206
##  Year.of.Death    Lapse.Indicator  Year.of.Lapse    Cause.of.Death
##  Min.   :2001     Min.   :1        Min.   :2001     Length:978582
##  1st Qu.:2015     1st Qu.:1        1st Qu.:2017     Class :character
##  Median :2019     Median :1        Median :2021     Mode  :character
##  Mean   :2018     Mean   :1        Mean   :2019
##  3rd Qu.:2021     3rd Qu.:1        3rd Qu.:2022
##  Max.   :2023     Max.   :1        Max.   :2023
##  NA's   :938206   NA's   :867693   NA's   :867693
##   Age.at.Death    Cause.of.Death.Description
##  Min.   :26.0     Length:978582
##  1st Qu.:52.0     Class :character
##  Median :59.0     Mode  :character
##  Mean   :58.6
##  3rd Qu.:66.0
##  Max.   :87.0
##  NA's   :938206
```

Misc Plots for Initial Analysis

```
plot_df <- superlife_df %>%
    filter(Cause.of.Death.Description != "NA") %>%
    mutate(Cause.of.Death.Description = ifelse(nchar(Cause.of.Death.Description) >
        20, paste0(str_sub(Cause.of.Death.Description, end = 20), "..."), Cause.of.Death.Description))
    group_by(Cause.of.Death.Description) %>%
    summarise(count = n())

ggplot(plot_df, aes(x = reorder(Cause.of.Death.Description, desc(count)), y = count)) +
    geom_col(fill = "blue", color = "black") + labs(title = "Histogram of Cause of Death",
    x = "Cause of Death", y = "Count") + theme(axis.text.x = element_text(angle = 45,
    hjust = 1), plot.title = element_text(hjust = 0.5))
```

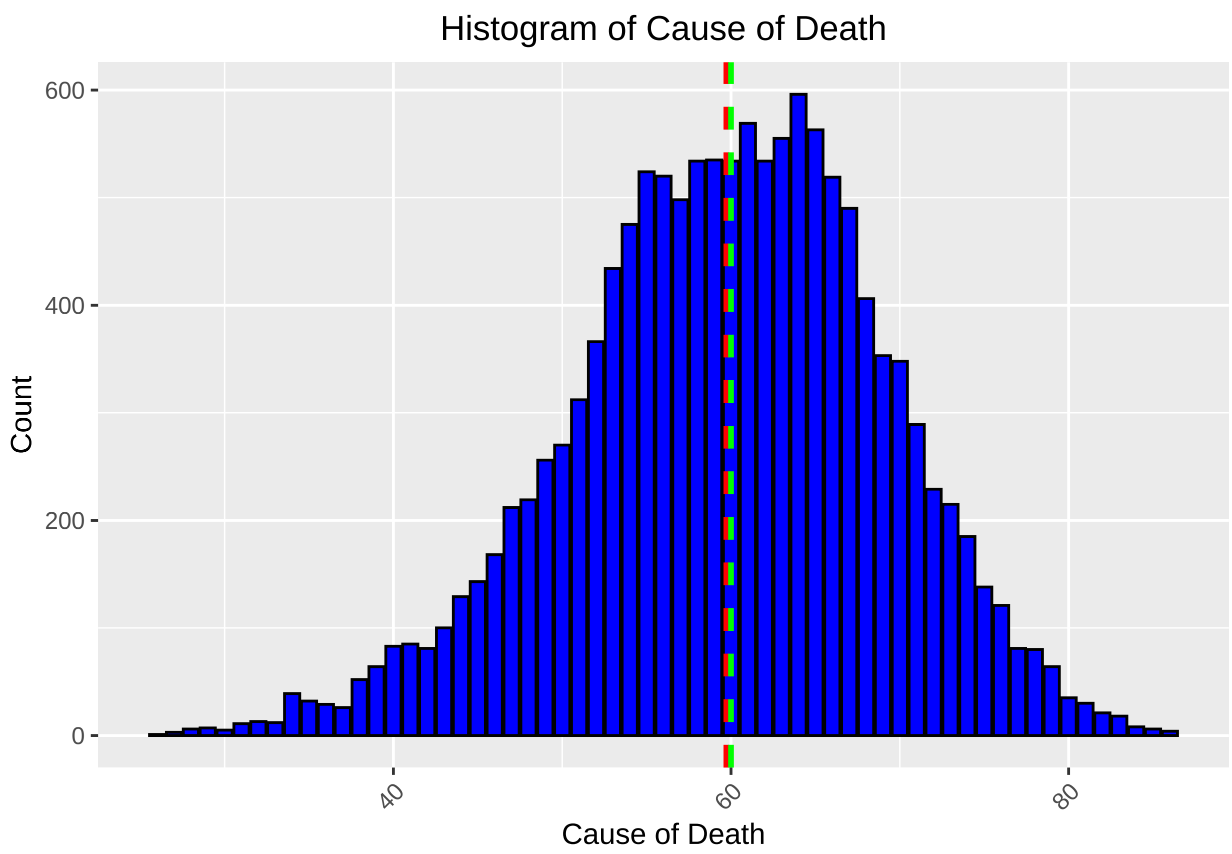



```
neoplasm_df <- superlife_df %>%
    filter(Cause.of.Death.Description == "Neoplasms") %>%
    group_by(Age.at.Death) %>%
    summarise(count = n())

mean <- weighted.mean(neoplasm_df$Age.at.Death, neoplasm_df$count)
median <- median(rep(neoplasm_df$Age.at.Death, times = neoplasm_df$count))

hist <- ggplot(neoplasm_df, aes(x = Age.at.Death, y = count)) + geom_col(fill = "blue",
    color = "black") + labs(title = "Histogram of Cause of Death", x = "Cause of Death",
    y = "Count") + theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(h

hist + geom_vline(xintercept = mean, color = "red", linetype = "dashed", size = 1) +
    geom_vline(xintercept = median, color = "green", linetype = "dashed", size = 1)
```



Generate and write Neoplasm loading based on cancer death rates

```
neoplasm_df <- neoplasm_df %>%
    filter(Age.at.Death >= 50) %>%
    mutate(Weight = count/sum(count)) %>%
    select(Age.at.Death, Weight)

write_csv(neoplasm_df, "../Data/Processed Data/Neoplasm_Mortality_Loading.csv")
```

```
sum(neoplasm_df$Weight)
```

```
## [1] 1
```

```
neoplasm_df <- superlife_df %>%
    filter(Cause.of.Death.Description == "Neoplasms") %>%
    group_by(Sex, Age.at.Death) %>%
    summarise(count = n())
```

```
## `summarise()` has grouped output by 'Sex'. You can override using the `.groups`
## argument.
```

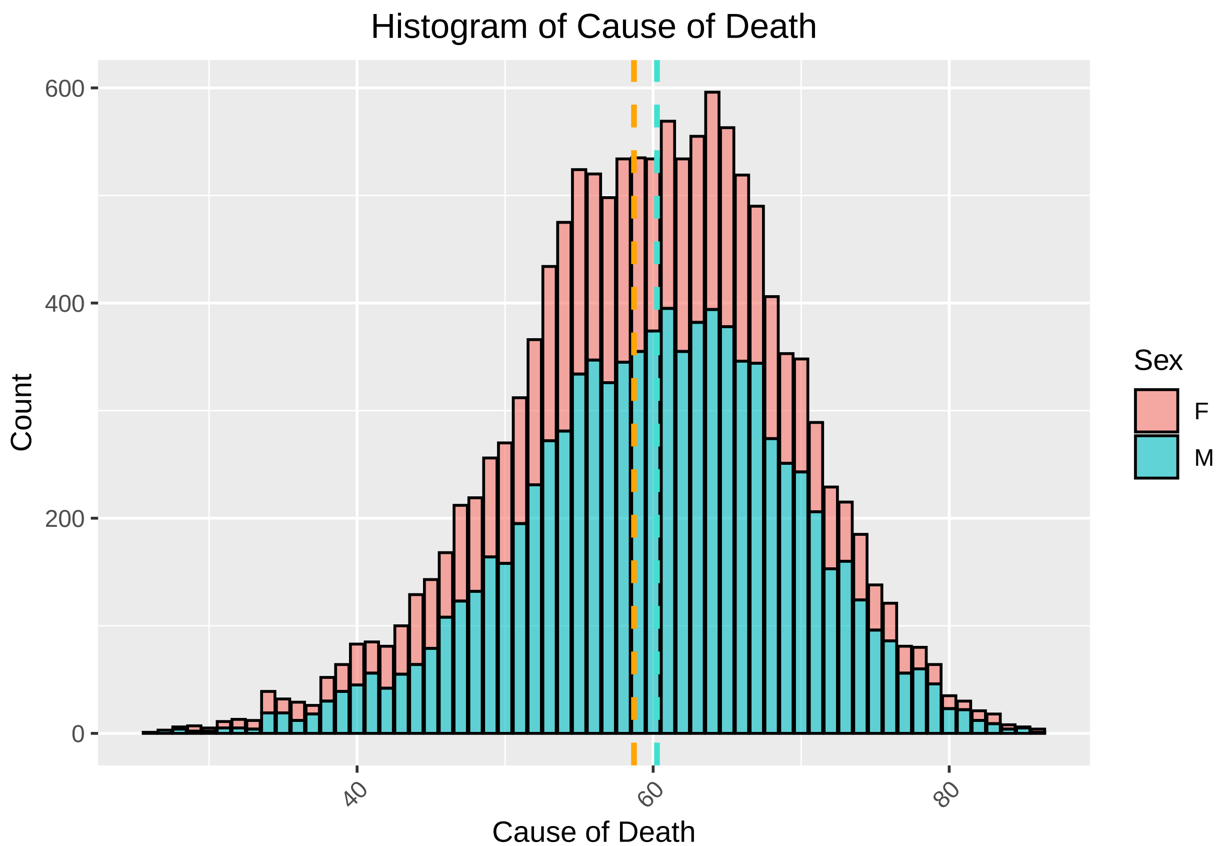
```
means <- neoplasm_df %>%
    group_by(Sex) %>%
    summarise(mean = weighted.mean(Age.at.Death, count))

mean_f <- means %>%
    filter(Sex == "F") %>%
    pull(mean)


mean_m <- means %>%
    filter(Sex == "M") %>%
    pull(mean)

hist <- ggplot(neoplasm_df, aes(x = Age.at.Death, y = count, fill = Sex)) + geom_col(color = "black",
    alpha = 0.6) + labs(title = "Histogram of Cause of Death", x = "Cause of Death",
    y = "Count") + theme(axis.text.x = element_text(angle = 45, hjust = 1), plot.title = element_text(h

hist + geom_vline(xintercept = mean_f, color = "orange", linetype = "dashed", size = 1) +
    geom_vline(xintercept = mean_m, color = "turquoise", linetype = "dashed", size = 1)
```




```
# Smoking rate in inforce data
smokers <- superlife_df %>%
    filter(Smoker.Status == "S") %>%
    nrow()
```

```r
smokers/nrow(superlife_df)
```

```
## [1] 0.06309129
```

Inforce Mortality Modelling

Helitha Dharmadasa - z5451805

2024-02-23

```r
library(tidyverse)
```

Read Data

```r
superlife_df <- read_csv("../Data/Processed Data/CLEANED_2024-srcsc-superlife-inforce-dataset.csv")
```

```
## Rows: 978582 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): Policy.number, Policy.type, Sex, Smoker.Status, Underwriting.Class,...
## dbl (9): Issue.year, Issue.age, Face.amount, Region, Death.indicator, Year.o...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
head(superlife_df)
```

```
## # A tibble: 6 x 18
##   Policy.number Policy.type Issue.year Issue.age Sex   Face.amount Smoker.Status
##   <chr>         <chr>            <dbl>     <dbl> <chr>       <dbl> <chr>
## 1 O8FN60R4KXIS  T20               2001        54 F          100000 NS
## 2 K0JK2XD81ZNI  SPWL              2001        54 M         1000000 NS
## 3 AH3A98MHTO8H  T20               2001        27 F           50000 NS
## 4 C9QPJMIH8H9Y  T20               2001        55 F         2000000 NS
## 5 2C1HL2XQOWME  T20               2001        39 F          250000 NS
## 6 LKW7MA7BPAV1  SPWL              2001        41 M         2000000 NS
## # i 11 more variables: Underwriting.Class <chr>, Urban.vs.Rural <chr>,
## #   Region <dbl>, Distribution.Channel <chr>, Death.indicator <dbl>,
## #   Year.of.Death <dbl>, Lapse.Indicator <dbl>, Year.of.Lapse <dbl>,
## #   Cause.of.Death <chr>, Age.at.Death <dbl>, Cause.of.Death.Description <chr>
```

```r
summary(superlife_df)
```

```
##  Policy.number      Policy.type          Issue.year     Issue.age
##  Length:978582      Length:978582      Min.   :2001   Min.   :26.0
##  Class :character   Class :character   1st Qu.:2009   1st Qu.:36.0
##  Mode  :character   Mode  :character   Median :2015   Median :44.0
```

```
##                                        Mean   :2014   Mean   :44.1  
##                                        3rd Qu.:2020   3rd Qu.:52.0  
##                                        Max.   :2023   Max.   :65.0  
##                                                                      
##      Sex             Face.amount      Smoker.Status      Underwriting.Class
##  Length:978582      Min.   :  50000   Length:978582      Length:978582     
##  Class :character   1st Qu.: 100000   Class :character   Class :character  
##  Mode  :character   Median : 500000   Mode  :character   Mode  :character  
##                     Mean   : 665574                                        
##                     3rd Qu.:1000000                                        
##                     Max.   :2000000                                        
##                                                                            
##  Urban.vs.Rural         Region      Distribution.Channel Death.indicator
##  Length:978582      Min.   :1.000   Length:978582        Min.   :1      
##  Class :character   1st Qu.:1.000   Class :character     1st Qu.:1      
##  Mode  :character   Median :2.000   Mode  :character     Median :1      
##                     Mean   :2.748                        Mean   :1      
##                     3rd Qu.:4.000                        3rd Qu.:1      
##                     Max.   :6.000                        Max.   :1      
##                                                          NA's   :938206 
##  Year.of.Death    Lapse.Indicator  Year.of.Lapse    Cause.of.Death    
##  Min.   :2001     Min.   :1        Min.   :2001     Length:978582     
##  1st Qu.:2015     1st Qu.:1        1st Qu.:2017     Class :character  
##  Median :2019     Median :1        Median :2021     Mode  :character  
##  Mean   :2018     Mean   :1        Mean   :2019                       
##  3rd Qu.:2021     3rd Qu.:1        3rd Qu.:2022                       
##  Max.   :2023     Max.   :1        Max.   :2023                       
##  NA's   :938206   NA's   :867693   NA's   :867693                     
##   Age.at.Death    Cause.of.Death.Description
##  Min.   :26.0     Length:978582             
##  1st Qu.:52.0     Class :character          
##  Median :59.0     Mode  :character          
##  Mean   :58.6                               
##  3rd Qu.:66.0                               
##  Max.   :87.0                               
##  NA's   :938206
```

```r
max_year <- max(superlife_df$Issue.year)

superlife_df <- superlife_df %>%
    filter(is.na(Lapse.Indicator)) %>%
    mutate(Max.age = coalesce(Age.at.Death, max_year - Issue.year + Issue.age))

max_obs <- nrow(superlife_df)

superlife_df
```

```
## # A tibble: 867,693 x 19
##    Policy.number Policy.type Issue.year Issue.age Sex   Face.amount
##    <chr>         <chr>            <dbl>     <dbl> <chr>       <dbl>
##  1 K0JK2XD81ZNI  SPWL              2001        54 M         1000000
##  2 LKW7MA7BPAV1  SPWL              2001        41 M         2000000
##  3 MWUNTLGLE8NR  SPWL              2001        37 F          100000
##  4 BJJ1U7SIJUCS  SPWL              2001        48 F         1000000
```

```
##   5 JTFR6CAODMLQ  T20               2001        46 M           50000
##   6 CHBTT2PBPQYC  SPWL              2001        50 M         1000000
##   7 K3H8WN6O2QMJ  SPWL              2001        50 M          100000
##   8 HSITVHDV2XTJ  T20               2001        48 F          250000
##   9 KN7X1NLMWUIN  T20               2001        52 M         1000000
##  10 ISEEQXTXIIV4  SPWL              2001        42 F         2000000
## # i 867,683 more rows
## # i 13 more variables: Smoker.Status <chr>, Underwriting.Class <chr>,
## #   Urban.vs.Rural <chr>, Region <dbl>, Distribution.Channel <chr>,
## #   Death.indicator <dbl>, Year.of.Death <dbl>, Lapse.Indicator <dbl>,
## #   Year.of.Lapse <dbl>, Cause.of.Death <chr>, Age.at.Death <dbl>,
## #   Cause.of.Death.Description <chr>, Max.age <dbl>
```

Calculate Inforce Mortality

```
# Calculate mortality rate of inforce dataset
mortality_df <- superlife_df %>%
    select(Max.age) %>%
    rowwise() %>%
    mutate(Age = list(seq(1, Max.age))) %>%
    unnest(c(Age)) %>%
    group_by(Age) %>%
    summarise(lx = n()) %>%
    mutate(mortality_rate = 1 - ifelse(is.na(lead(lx)), 0, (lead(lx)/lx)))

mortality_df
```

```
## # A tibble: 87 x 3
##      Age     lx mortality_rate
##    <int>  <int>          <dbl>
##  1     1 867693              0
##  2     2 867693              0
##  3     3 867693              0
##  4     4 867693              0
##  5     5 867693              0
##  6     6 867693              0
##  7     7 867693              0
##  8     8 867693              0
##  9     9 867693              0
## 10    10 867693              0
## # i 77 more rows
```

```
write_csv(mortality_df, "../Data/Processed Data/Superlife-inforce-mortality-table.csv")
```

Benefit Modelling

Helitha Dharmadasa - z5451805

2024-02-23

```r
library(tidyverse)
library(readxl)
```

Read Data

```r
superlife_df <- read_csv("../Data/Processed Data/CLEANED_2024-srcsc-superlife-inforce-dataset.csv")
```

```
## Rows: 978582 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): Policy.number, Policy.type, Sex, Smoker.Status, Underwriting.Class,...
## dbl (9): Issue.year, Issue.age, Face.amount, Region, Death.indicator, Year.o...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
rate_df <- read_excel("../Data/Case Study Data/srcsc-2024-lumaria-economic-data.xlsx",
    skip = 10)
```

```r
mortality_df <- read_excel("../Data/Case Study Data/srcsc-2024-lumaria-mortality-table.xlsx",
    skip = 13)

adj_mortality_df <- read_excel("../Data/Processed Data/average_mortality_table.xlsx")
```

Transform Mortality tables

```r
# Create lx (Number of people surviving) to life tables
mortality_df <- mortality_df %>%
    mutate(survival_rate = 1 - `Mortality Rate`, lx = cumprod(survival_rate))

adj_mortality_df <- adj_mortality_df %>%
    mutate(survival_rate = 1 - new_mortality, lx = cumprod(survival_rate))

write_csv(mortality_df, "../Data/Processed Data/mortality_baseline.csv")

write_csv(adj_mortality_df, "../Data/Processed Data/mortality_adjusted.csv")
```

Create benefit modelling function

```
benefit_model <- function(policy_duration, mortality_table, interest_rate, inflation_rate) {
    # Function that takes generates $1 EPVs for different payment structures.
    # Two types of EPV's are calculated, single payout - paying out $1 on
    # death, and annuity_dues paying $1 yearly until death or artifical cap.
    # Inputs: policy_duration - (20 Years Term or 120 Years whole life)
    # mortality_table - Table containing lx values interest_rate - interest
    # rate to use for discounting inflation_rate - inflation rate to calculate
    # real rate Outputs: output_df$EPV_single - EPV of a insurance polcy paying
    # out death output_df$EPV_annuity_due - EPV on an annuity paying until
    # death output_df$EPV_single_18 - EPV of an insurance policy that starts
    # 18+ paying $1 on death output_df$EPV_single_50 - EPV of an insurance
    # policy that starts at 18+ up to 88 output_df$EPV_annuity_due_18 - EPV of
    # an annuity starting at 18+ output_df$EPV_annuity_due_50 - EPV of an
    # annuity starting 50+

    real_rate = ((interest_rate - inflation_rate)/(1 + inflation_rate))

    v = 1/(1 + real_rate)

    output_df <- tibble(age = 1:120, EPV_single = rep(0, 120), EPV_annuity_due = rep(0,
        120), EPV_single_18 = rep(0, 120), EPV_single_50 = rep(0, 120), EPV_annuity_due_18 = rep(0,
        120), EPV_annuity_due_50 = rep(0, 120))

    # Ages 1-119 Note: 119 because we don't have Age 121 in life table for 120
    # calc.
    for (starting_age in 1:119) {

        single = 0
        annuity_due = 0
        single_18 = 0
        single_50 = 0
        annuity_due_18 = 0
        annuity_due_50 = 0

        # Rolling window of policy dur or capped at life table limits
        # (truncation error?)
        age_max <- min(starting_age + policy_duration, 119)

        for (death_age in starting_age:age_max) {

            t = death_age - starting_age

            # P(starting age is alive until death age)
            tpx = mortality_table$lx[death_age]/mortality_table$lx[starting_age]

            # P(death age dies in the next year)
            qxt = 1 - (mortality_table$lx[death_age + 1]/mortality_table$lx[death_age])

            single = single + v^(t + 1) * tpx * qxt  #Paid out EOY of Death
            annuity_due = annuity_due + v^(t) * tpx  #Paid out SOY of every year alive
```

```r
            if (death_age >= 18) {
                # Paid on death if older than 18
                single_18 = single_18 + v^(t + 1) * tpx * qxt

                # Paid SOY yearly starting at 18
                annuity_due_18 = annuity_due_18 + v^(t) * tpx
            }

            if (death_age >= 50) {
                # Paid SOY yearly starting at 50
                annuity_due_50 = annuity_due_50 + v^(t) * tpx
            }

            if (death_age == 27) {
                # Paid at 50 if alive at 50
                single_50 = v^(t) * tpx
            } else if (starting_age > 27 & starting_age <= 88) {
                single_50 = 1
            }
        }

        output_df$EPV_single[starting_age] = single
        output_df$EPV_annuity_due[starting_age] = annuity_due
        output_df$EPV_single_18[starting_age] = single_18
        output_df$EPV_single_50[starting_age] = single_50
        output_df$EPV_annuity_due_18[starting_age] = annuity_due_18
        output_df$EPV_annuity_due_50[starting_age] = annuity_due_50

    }

    return(output_df)
}
```

```r
# Call benefit_model for different terms
interest_rate <- mean(rate_df$`1-yr Risk Free Annual Spot Rate`)
inflation_rate <- mean(rate_df$Inflation)

T20_EPV_df <- benefit_model(20, mortality_df, interest_rate, inflation_rate) %>%
    mutate(EPV_single_adj = benefit_model(20, adj_mortality_df, interest_rate, inflation_rate)$EPV_singl

SPWL_EPV_df <- benefit_model(120, mortality_df, interest_rate, inflation_rate) %>%
    mutate(EPV_single_adj = benefit_model(120, adj_mortality_df, interest_rate, inflation_rate)$EPV_sing
```

```r
# Average face values for the two policy types
T20_FV <- superlife_df %>%
    filter(Policy.type == "T20") %>%
    summarise(Mean = mean(Face.amount)) %>%
    pull(Mean)

SPWL_FV <- superlife_df %>%
    filter(Policy.type == "SPWL") %>%
    summarise(Mean = mean(Face.amount)) %>%
    pull(Mean)
```

```r
# Engagement rate (same for all interventions)
engagement_rate = 0.25

# Calculate expense dataframes for the two policies
T20_expense_df <- tibble(Age = T20_EPV_df$age,
                         smoking = T20_EPV_df$EPV_single_50 * 2065 * 0.18,
                         screening = T20_EPV_df$EPV_annuity_due_50 * 65,
                         fitness =  T20_EPV_df$EPV_annuity_due_18 * 18,
                         safety = T20_EPV_df$EPV_annuity_due_18 * 12.5)

SPWL_expense_df <- tibble(Age = SPWL_EPV_df$age,
                         smoking = SPWL_EPV_df$EPV_single_50 * 2065 * 0.18,
                         screening = SPWL_EPV_df$EPV_annuity_due_50 * 65,
                         fitness =  SPWL_EPV_df$EPV_annuity_due_18 * 18,
                         safety = SPWL_EPV_df$EPV_annuity_due_18 * 12.5)

# Calculate total EPVs to compare the different programs
benefit_df <- tibble(Age = SPWL_EPV_df$age,
                         T20_baseline = T20_EPV_df$EPV_single * T20_FV,
                         T20_intervention = ((1-engagement_rate) * T20_EPV_df$EPV_single + engagement_rate
                           engagement_rate * (T20_expense_df$smoking +
                           T20_expense_df$screening +
                           T20_expense_df$fitness +
                           T20_expense_df$safety),

                         SPWL_baseline = SPWL_EPV_df$EPV_single * SPWL_FV,
                         SPWL_intervention = ((1-engagement_rate) * SPWL_EPV_df$EPV_single + engagement_rat
                           engagement_rate * (SPWL_expense_df$smoking +
                           SPWL_expense_df$screening +
                           SPWL_expense_df$fitness +
                           SPWL_expense_df$safety))

benefit_df <- benefit_df %>%
  mutate(T20_profit_flag = factor(ifelse(T20_intervention < T20_baseline, "Profit", "Loss")),
         SPWL_profit_flag = factor(ifelse(SPWL_intervention < SPWL_baseline, "Profit", "Loss")))

write_csv(benefit_df, "../Data/Processed Data/Benefit_Modelling.csv")
```

```r
# Calculate benefit comparison dataframe aggregated to different age brackets
summary_df <- benefit_df %>%
    filter(Age < 120) %>%
    mutate(Age_bracket = cut(Age, breaks = c(1, 23, 45, 60, 85, 120), labels = c("1-22",
        "23-44", "45-59", "60-84", "85+"), right = FALSE))



summary_df <- summary_df %>%
    group_by(Age_bracket) %>%
    summarise(across(c(T20_baseline, T20_intervention, SPWL_baseline, SPWL_intervention),
        ~round(mean(.), 0)))

write_csv(summary_df, "../Data/Processed Data/Benefits_by_Age_Group.csv")
```

Sensitivity Analysis

Nikhil Alex - z5451383

2024-03-21

```
library(tidyverse)
library(readxl)
```

```
superlife_df <- read_csv("../Data/Processed Data/CLEANED_2024-srcsc-superlife-inforce-dataset.csv")
```

```
## Rows: 978582 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): Policy.number, Policy.type, Sex, Smoker.Status, Underwriting.Class,...
## dbl (9): Issue.year, Issue.age, Face.amount, Region, Death.indicator, Year.o...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
rate_df <- read_excel("../Data/Case Study Data/srcsc-2024-lumaria-economic-data.xlsx",
    skip = 10)
mortality_df <- read_csv("../Data/Processed Data/mortality_baseline.csv")
```

```
## Rows: 120 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (4): Age, Mortality Rate, survival_rate, lx
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
adj_mortality_df <- read_csv("../Data/Processed Data/mortality_adjusted.csv")
```

```
## Rows: 120 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (7): Age, new_mortality, average_mortality_reduction, old_mortality, new...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
interest_rate <- mean(rate_df$`1-yr Risk Free Annual Spot Rate`)
inflation_rate <- mean(rate_df$Inflation)
engagement_rate = 0.25
```

```
# Benefit Model from Inforce Benefit modelling.Rmd

benefit_model <- function(policy_duration, mortality_table, interest_rate, inflation_rate) {

    real_rate = ((interest_rate - inflation_rate)/(1 + inflation_rate))
```

```r
v = 1/(1 + real_rate)

output_df <- tibble(age = 1:120, EPV_single = rep(0, 120), EPV_annuity_due = rep(0,
    120), EPV_single_18 = rep(0, 120), EPV_single_50 = rep(0, 120), EPV_annuity_due_18 = rep(0,
    120), EPV_annuity_due_50 = rep(0, 120))

# debug_mat <- matrix(0, nrow = 120, ncol = 120)

# Ages 1-119 Note: 119 because we don't have Age 121 in life table for 120
# calc.
for (starting_age in 1:119) {

    single = 0
    annuity_due = 0
    single_18 = 0
    single_50 = 0
    annuity_due_18 = 0
    annuity_due_50 = 0

    # Rolling window of policy dur or capped at life table limits
    # (truncation error?)
    age_max <- min(starting_age + policy_duration, 119)

    for (death_age in starting_age:age_max) {

        t = death_age - starting_age

        # P(starting age is alive until death age)
        tpx = mortality_table$lx[death_age]/mortality_table$lx[starting_age]

        # P(death age dies in the next year)
        qxt = 1 - (mortality_table$lx[death_age + 1]/mortality_table$lx[death_age])

        # debug debug_mat[starting_age, death_age] = v^(t+1) * tpx * qxt

        single = single + v^(t + 1) * tpx * qxt  #Paid out EOY of Death
        annuity_due = annuity_due + v^(t) * tpx  #Paid out SOY of every year alive

        if (death_age >= 18) {
            # Paid on death if older than 18
            single_18 = single_18 + v^(t + 1) * tpx * qxt

            # Paid SOY yearly starting at 18
            annuity_due_18 = annuity_due_18 + v^(t) * tpx
        }

        if (death_age >= 50) {
            # Paid SOY yearly starting at 50
            annuity_due_50 = annuity_due_50 + v^(t) * tpx
        }

        if (death_age == 27) {
            # Paid at 50 if alive at 50
```

```
            single_50 = v^(t) * tpx
        } else if (starting_age > 27 & starting_age <= 88) {
            single_50 = 1
        }
    }

    output_df$EPV_single[starting_age] = single
    output_df$EPV_annuity_due[starting_age] = annuity_due
    output_df$EPV_single_18[starting_age] = single_18
    output_df$EPV_single_50[starting_age] = single_50
    output_df$EPV_annuity_due_18[starting_age] = annuity_due_18
    output_df$EPV_annuity_due_50[starting_age] = annuity_due_50

  }

  return(output_df)
}
```

```
T20_FV <- superlife_df %>%
  filter(Policy.type == "T20") %>%
  summarise(Mean = mean(Face.amount)) %>%
  pull(Mean)

SPWL_FV <- superlife_df %>%
  filter(Policy.type == "SPWL") %>%
  summarise(Mean = mean(Face.amount)) %>%
  pull(Mean)
```

```
# Sensitivity levels
sens_levels <- c(0.001, 0.01, 0.1, -0.001, -0.01, -0.1)  # 0.1%, 1%, and 10% differences


sens_results <- tibble(Inflation_Rate_Change = numeric(), Interest_Rate_Change = numeric(),
  Engagement_Rate_Change = numeric(), T20_Profit_Count = numeric(), SPWL_Profit_Count = numeric())
```

```
for (inflation_change in sens_levels) {
  for (interest_change in sens_levels) {
    for (engagement_change in sens_levels) {

      # New parameter values
      new_inflation_rate <- inflation_rate * (1 + inflation_change)
      new_interest_rate <- interest_rate * (1 + interest_change)
      new_engagement_rate <- engagement_rate * (1 + engagement_change)

      # Updating EPVs
      T20_EPV_df_sens <- benefit_model(20, mortality_df, new_interest_rate,
        new_inflation_rate) %>%
        mutate(EPV_single_adj = benefit_model(20, adj_mortality_df, new_interest_rate,
          new_inflation_rate)$EPV_single)

      SPWL_EPV_df_sens <- benefit_model(120, mortality_df, new_interest_rate,
        new_inflation_rate) %>%
        mutate(EPV_single_adj = benefit_model(120, adj_mortality_df, new_interest_rate,
          new_inflation_rate)$EPV_single)
```

```
            T20_expense_df_sens <- tibble(Age = T20_EPV_df_sens$age, smoking = T20_EPV_df_sens$EPV_singl
                2065 * 0.18, screening = T20_EPV_df_sens$EPV_annuity_due_50 * 65,
                fitness = T20_EPV_df_sens$EPV_annuity_due_18 * 18, safety = T20_EPV_df_sens$EPV_annuity
                  12.5)

            SPWL_expense_df_sens <- tibble(Age = SPWL_EPV_df_sens$age, smoking = SPWL_EPV_df_sens$EPV_s
                2065 * 0.18, screening = SPWL_EPV_df_sens$EPV_annuity_due_50 * 65,
                fitness = SPWL_EPV_df_sens$EPV_annuity_due_18 * 18, safety = SPWL_EPV_df_sens$EPV_annui
                  12.5)

            benefit_df_sens <- tibble(Age = SPWL_EPV_df_sens$age, T20_baseline_sens = T20_EPV_df_sens$E
                T20_FV, T20_intervention_sens = ((1 - new_engagement_rate) * T20_EPV_df_sens$EPV_single
                new_engagement_rate * T20_EPV_df_sens$EPV_single_adj) * T20_FV +
                new_engagement_rate * (T20_expense_df_sens$smoking + T20_expense_df_sens$screening +
                  T20_expense_df_sens$fitness + T20_expense_df_sens$safety), SPWL_baseline_sens = SPWL_
                SPWL_FV, SPWL_intervention_sens = ((1 - new_engagement_rate) * SPWL_EPV_df_sens$EPV_sin
                new_engagement_rate * SPWL_EPV_df_sens$EPV_single_adj) * SPWL_FV +
                new_engagement_rate * (SPWL_expense_df_sens$smoking + SPWL_expense_df_sens$screening +
                  SPWL_expense_df_sens$fitness + SPWL_expense_df_sens$safety))

            benefit_df_sens <- benefit_df_sens %>%
                mutate(T20_profit_flag = factor(ifelse(T20_intervention_sens < T20_baseline_sens,
                  "Profit", "Loss")), SPWL_profit_flag = factor(ifelse(SPWL_intervention_sens <
                  SPWL_baseline_sens, "Profit", "Loss")))

            # Count profit occurrences
            T20_profit_count <- sum(benefit_df_sens$T20_profit_flag == "Profit")
            SPWL_profit_count <- sum(benefit_df_sens$SPWL_profit_flag == "Profit")

            # Append results to sensitivity results df
            sens_results <- sens_results %>%
                add_row(Inflation_Rate_Change = inflation_change, Interest_Rate_Change = interest_chang
                  Engagement_Rate_Change = engagement_change, T20_Profit_Count = T20_profit_count,
                  SPWL_Profit_Count = SPWL_profit_count)
        }
    }
}
```

```
# Results
View(sens_results)
summary(sens_results)
```

```
##  Inflation_Rate_Change Interest_Rate_Change Engagement_Rate_Change
##  Min.   :-0.10         Min.   :-0.10        Min.   :-0.10
##  1st Qu.:-0.01         1st Qu.:-0.01        1st Qu.:-0.01
##  Median : 0.00         Median : 0.00        Median : 0.00
##  Mean   : 0.00         Mean   : 0.00        Mean   : 0.00
##  3rd Qu.: 0.01         3rd Qu.: 0.01        3rd Qu.: 0.01
##  Max.   : 0.10         Max.   : 0.10        Max.   : 0.10
##  T20_Profit_Count SPWL_Profit_Count
##  Min.   :55.00    Min.   : 0.00
##  1st Qu.:60.75    1st Qu.:78.50
##  Median :62.00    Median :85.00
##  Mean   :61.17    Mean   :71.81
```

```
## 3rd Qu.:62.00    3rd Qu.:85.25
## Max.   :63.00    Max.   :86.00
```

```r
write_csv(sens_results, "../Data/Processed Data/Sensitivity_Analysis.csv")
```