

8-5-2016

# Generalized Linear Model Approach to Adjusting Expected Assumptions of Long-Term Care Incidence Rates

Rozita Ramli

*University of Connecticut*, ita.aktuari@gmail.com

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

---

## Recommended Citation

Ramli, Rozita, "Generalized Linear Model Approach to Adjusting Expected Assumptions of Long-Term Care Incidence Rates" (2016). *Doctoral Dissertations*. 1174.  
<https://opencommons.uconn.edu/dissertations/1174>

# Generalized Linear Model Approach to Adjusting Expected Assumptions of Long-Term Care Incidence Rates

Rozita Ramli, PhD

University of Connecticut, 2016

Actuarial assumptions are needed in most of actuarial works, for example pricing and reserving and setting capital standards. In long-term care (LTC) insurance, three main actuarial assumptions are incidence, termination and utilization rates. These assumptions can be seen as factors affecting the financial progress in LTC insurance business. Insurance companies must make sure the assumptions they use adequately reflect actual experience to avoid adverse financial consequences. Hence, a regular systematic methodology to monitor expected assumptions is essential. This work proposes a methodology to adjusting expected assumptions of LTC incidence rates. The methodology uses a Generalized Linear Model (GLM) approach in modeling incidence rates and the adjustment is parallel to an established industry technique of stochastic deferred acquisition cost (DAC) unlocking. The methodology uses confidence intervals as a decision tool to decide if any adjustment is needed. We also bring in a credibility component into the methodology to provide rationalization for choosing the appropriate significant level of the confidence interval when applying the methodology. Lastly we study the effect of size of historical experience used in the adjustment on the efficiency of the methodology. Similar to how the stochastic DAC unlocking technique is an acceptable technique by regulators and practitioners, we foresee our GLM methodology to be accepted as a basis for adjusting assumptions under Principle-Based Reserving in the future.

Generalized Linear Model Approach to Adjusting Expected Assumptions of Long-Term Care  
Incidence Rates

Rozita Ramli

B.Sc., Universiti Kebangsaan Malaysia, 2002

M.Sc., City University London, 2005

M.Sc., University of Connecticut, 2011

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2016

Copyright by

Rozita Ramli

2016

APPROVAL PAGE

Doctor of Philosophy Dissertation

Generalized Linear Model Approach to Adjusting Expected Assumptions of Long-Term Care

Incidence Rates

Presented by

Rozita Ramli, B.Sc., M.Sc.

Major Advisor \_\_\_\_\_

Jeyaraj Vadiveloo

Associate Advisor \_\_\_\_\_

Emiliano Valdez

Associate Advisor \_\_\_\_\_

Guojun Gan

University of Connecticut

2016

# Acknowledgement

All praises due to the Almighty.

First and foremost I would like to thank my parents and who are also my best friends, my mother Zaharah Jusoh and my late father Ramli Arshad. Their unconditional love and support has given me the strength to keep me going further in life and giving back to people around me. If their sacrifice were to be summed up in short, I would say I owe my life to them.

Then my gratitude goes to my advisor Professor Jeyaraj Vadiveloo who not only acted as an advisor, but from the very beginning of my study has been very supportive. He guided, advised, assisted and inspired me. Thank you for seeing and believing in my potential, and helping to promote it to its utmost.

Thanks to Monique Roy who is first a true friend, and then a supportive staff member at UConn's Math Department, for bringing me into her beautiful family and making my life away from family feel like home, and help ease a lot of academic issues throughout my studies.

My next gratitude goes to my old officemates, Ushani Dias, Priyantha Hewa, Wenyan Zheng, Gao Niu, Shujuan Huang, and also to my current officemates who have been treating me not only as a friend, but more importantly as a sister. I completed this piece of work using the laptop they bought for me after I lost my own.

Then I would like to thank all faculty at UConn's Math Department who had significantly influenced me throughout my study at UConn, James Bridgeman, Iddo Ben Ari, Yung Choi, Patrick McKenna, Vadim Olshevsky, (late) Evarist Gine, Reed Solomon, Greg Smith, Emiliano Valdez, Brian Hartman, Guojun Gan and the rest for their advice, support and assistance.

Thank you to my friends, Hua Ren, Daniel Kelleher, Bernardo Marquez and Joshua Katz for the patience and perseverance in introducing and teaching real math concepts to me at the beginning of my study at UConn.

I would also like to express my gratitude to my siblings, friends at ICUC (Islamic Center of UConn), ex-roommates, all friends and colleagues at Universiti Kebangsaan Malaysia and UConn's Math Department, all friends and their families, all teachers and everybody whose presence had a significant impact to my life as a PhD student at UConn specifically, and to my personal life in general that I might have not mentioned here. This list would go very long because there are so many of you that I have to say thank you to, but I believe you know who you are and how you affected my life and I thank you all very much.

Last but not least I would like to thank the love of my life and who is also my best friend, Turgut Yilmaz for being a very loving, understanding and supportive husband. Thank you for cherishing my strengths and helping me overcome my weaknesses. I would not be able to do this without the firm trust in me and steadfast encouragement that you have been providing.

One more time, I humbly thank all of you for being part of this challenging knowledge-seeking journey and helping me achieve one of my biggest dreams. I thank you all so very much.

# Table of Contents

Abstract	
Title Page	i
Copyright Page	ii
Approval Page	iii
Acknowledgment	iv
Table of Contents	vi
<b>Chapter 1: Introduction</b>	<b>1</b>
1.1 SOA LTC (Claim Incidence) Intercompany Experience Study	3
1.2 Mortality Rates Adjustment	4
<b>Chapter 2: Background</b>	<b>6</b>
2.1 LTC Insurance	6
2.2 Claims Tracking and Monitoring	8
2.3 LTC Incidence Rates and Poisson Generalized Linear Models (GLM)	9
<b>Chapter 3: Methodology</b>	<b>13</b>
3.1 Stochastic DAC Unlocking for Variable Annuity Products Technique	13
3.2 Proposed Methodology	15
3.2.1 Model Assumption of Initial Expected Assumptions	16
3.2.2 Adjustment Procedure	17
3.3 Reliability of the Proposed Methodology	19
	vi



3.3.1	Type I Error	19
3.3.2	Type II Error	21
3.3.3	Type III Error	22
3.4	Proposed Methodology with Credibility Effect	25
<b>Chapter 4:</b>	<b>Analyses and Results</b>	<b>28</b>
4.1	Establishment of Initial Expected Assumptions of LTC Incidence Rates	28
4.2	Verification of the Reliability of the Proposed Methodology	31
4.2.1	Type I Error	31
4.2.2	Type II Error	32
4.2.3	Type III Error	34
4.3	Application of the Proposed Methodology	36
4.3.1	Application of the Proposed Methodology for One Simulation	36
4.3.2	Application of the Proposed Methodology for One Hundred Simulations	42
4.4	Proposed Methodology with Credibility Effect	44
4.4.1	Credibility Effect to Stable Dataset	45
4.4.2	Credibility Effect to Unstable Dataset	46
4.5	Effect of Historical Data on the Application of the Proposed Methodology	47
<b>Chapter 5:</b>	<b>Conclusion</b>	<b>49</b>
	References	54
	Appendix	55

# Chapter 1: Introduction

Actuarial assumptions are needed in most of actuarial works, particularly pricing and reserving. Traditionally, most of actuarial assumptions were built from an experience dataset by observing the number of decrements and dividing it by the number of exposures. For example, in obtaining the expected mortality rates of a specific group, the number of deaths out of the total exposure of this group was observed and the ratio of this number to the exposure gives the crude estimate. The crude rates for all groups then were smoothed via a graduation procedure to generate a consistent set of rates that could be used as a set of actuarial assumptions. The traditional approach only allows for risk evaluations for large group categories (such as male, female, smoker and non-smoker), since a small risk category (such as male non-smoker with policy face amounts exceeding \$1M and attained ages above 60) with such a refinement may result in small exposures and no claims.

Nowadays, most of actuarial assumptions are constructed through more sophisticated mathematical models that provide richer information. For example, the most prominent and widely used ones are Generalized Linear Models, also known as GLM. With the same size dataset used by the traditional approach for constructing actuarial assumptions, GLM provides expected values for more refined risk groups as long as the predictor variables (such as policy face amount, gender, attained age, duration, etc.) are captured in the dataset. It captures lots more drivers, or explanatory variables, that could impact the assumptions. This is achievable because GLM uses a mathematical

formula with a set of statistical assumptions to produce estimates of coefficients of the predictor variables used to predict the risk metric of interest, such as mortality rates.

As in the traditional approach, the actuarial estimates attained by GLM for a group of policies could be used to compare against actual experience of the same group over time. However, adjusting expected coefficients in a GLM model to reflect actual experience is not as easy as in the traditional approach when the risk classes are much broader. Some expected coefficients may remain unchanged, some could increase while others could decrease in order to reflect actual experience. Also, for reasons of credibility, some consistent and statistically sound criteria has to be developed to recognize when an expected coefficient has to be adjusted as well as the level of adjustment needed to reflect actual experience.

This is the main motivation of this work which is to develop a methodology to adjust actuarial assumptions derived by a GLM model to reflect actual experience. The proposed methodology is new but it follows an existing procedure currently adopted by companies and regulators, DAC unlocking technique<sup>4</sup>, very closely. While the DAC technique is used for adjusting current deferred acquisition costs (DAC) for variable annuity products, the proposed methodology, while applied for a different purpose, uses a similar methodology. Our methodology provides a statistically rigorous technique to adjust actuarial assumptions derived by GLM and we will apply this methodology to adjust expected assumptions for long term care (LTC) incidence rates.

## 1.1 SOA LTC (Claim Incidence) Intercompany Experience Study<sup>5</sup>

The basis of LTC expected incidence rates is the Society of Actuaries (SOA) LTC Claim Incidence Intercompany Experience Study. There were two models developed by this study. One is a Poisson GLM model based on total life exposure, which comprises the number of days during the exposure period between the policy effective date and termination date. The second is a Poisson GLM model based on active life exposure, which equals total life exposure less the period of time on claim. Exposure forms the denominator part of incidence rates which implies rates produced using total life exposure is smaller than the one produced using active life exposure.

Eleven explanatory variables were used in both of the models and they were categorized into two groups, namely product characteristic and policyholder characteristic. The table below lists the eleven explanatory variables:

Product Characteristic	Policyholder Characteristic
Underwriting Type	Issue Age
Elimination Period	Premium Class
Benefit Period	Marital Status
Coverage Type	Region
Maximum Daily Benefit	Gender
Tax Qualified Status	-

Table 1: Explanatory variables for SOA LTC claims incidence study

This study also showed the existence of interaction among these variables. Benefit period is the only variable that does not interact with other variables.

There were twenty-two LTC insurance companies that participated in the study. The models fit well for the aggregate dataset and is reasonable for individual companies. Other users are advised to use the models with caution since it may not sufficiently represent their actual

experience. This again brings the significance of the concern on how to update actuarial assumptions in a consistent and statistically rigorous manner to reflect individual company experience.

Other approaches in the literature to adjust expected assumptions are described in the next section. However, there is nothing in the actuarial literature describing how expected assumptions should be adjusted in a GLM context.

## **1.2 Mortality Rates Adjustment**

Anderson R. & Rosenberg<sup>1</sup> suggested the use of age-adjusted mortality rates to replace crude death rates to better represent mortality rates of a specific population. Moreover, age-adjusted mortality rates yield to coherent comparison should we need to study mortality rates of two or more populations. The technique that the authors presented provides not only a better reflection of actual mortality experience, but also as a means of standardizing mortality rates.

Benjamin G.<sup>2</sup> used a credibility technique to adjust mortality rates. Plots of the ratio of actual to expected deaths were used. Expected assumptions represent actual experience sufficiently if this plot is close to one. As time passes, insurance companies gain more actual experience which means more information regarding a company's own experience is gathered and hence it should not depend on expected assumptions completely. The technique that the author discussed combines expected assumptions and actual experience appropriately. The appropriateness of the combination was measured by the expected value of the random variable, which is the mortality rate. It is then compared to the actual number of deaths and exposures to determine if actual experience possessed full or partial credibility for producing the pertinent mortality rates.

Both approaches<sup>1, 2</sup> mentioned above handle mortality rates with one factor which is age. This limits their application to LTC incidence rates since it depends on several factors such as age, gender, marital status, elimination period and maximum daily benefit, to name a few. Furthermore, for the SOA Intercompany LTC Study, a Generalized Linear Model (GLM) approach was used, in particular, a Poisson GLM was used for modeling incidence rates. A new approach for updating expected assumptions of LTC insurance, and in particular incidence rates, under the framework of GLM is necessary. The new approach should conform to common business practices and provide a consistent and statistically rigorous approach to changing expected assumptions.

This work proposes a methodology to adjust expected assumptions of LTC incidence rates. The methodology uses a GLM approach in modeling incidence rates and the adjustment is parallel to an established industry technique of stochastic deferred acquisition cost (DAC) unlocking<sup>4</sup>.

## **Chapter 2: Background**

### **2.1 LTC Insurance**

LTC is a range of services and support needed by people who experience difficulties in performing activities of daily living (ADLs) such as bathing, dressing, using the toilet, transferring (to or from bed or chair), caring for incontinence and eating. Other LTC services and support also aid people who need help in instrumental activities of daily living (IADLs) such as housework, managing money, taking medication, preparing and cleaning up after meals, shopping for groceries or clothes, using the telephone or other communication devices, caring for pets and responding to emergency alerts such as fire alarms<sup>7</sup>.

Assistance for LTC is usually provided at home, nursing home or assisted living facility. MEDICARE and MEDICAID are public programs that provide financial help if one needs LTC and satisfies the requirements to qualify him for the aid. The difference between these two programs are who and what they are covering. While MEDICARE basically covers people who are age 65 and above and need LTC, MEDICAID on the other hand helps people who have financial constraints for acquiring medical assistance. However these programs are not readily available should one be in need of LTC assistance. As mentioned above, he has to satisfy the requirements to qualify him for the aid and most likely he has to show that he has exhausted his wealth before he become eligible. There is cost associated with getting LTC and one might be in financial distress if he needs it over an extended period of time. Everybody faces this risk but it is

most prevalent in older people. LTCI is an insurance product offered in the market to provide coverage for this particular risk.

The policyholder pays a predetermined LTC insurance premium. In 2007, the average premium was above \$2200 per year. In the situation where expected assumptions used for determining the premium rate experience a significant adverse change, an insurance company can increase this predetermined premium. One has to undergo an underwriting process before becoming eligible for LTC insurance. This means there are conditions that prevent an individual from obtaining LTC insurance such as if he is already using LTC or suffering from Alzheimer<sup>7</sup>.

Two aspects a policyholder must meet prior to receiving LTC insurance benefits are benefit triggers and elimination or waiting period. Benefit triggers are conditions to enable a benefit payment and once this condition takes place, it has to last for a specific period of time called the elimination period in order to receive the benefit. Benefit reimbursement is subject to maximum benefit and period coverages<sup>7</sup>.

From the insurance company viewpoint, there are three main expected assumptions in LTC insurance, namely incidence, termination and utilization rates. Incidence rates is the number of new claims by healthy policies over a specific period of time. Termination rates is the number of closed claims by existing claims over a specific period of time. Utilization rates is the amount of actual reimbursement relative to the maximum daily benefit over a specific period of time. These rates are important factors to determine the financial progress of an LTC block of business. Hence, it is very crucial to have a set of expected assumptions that sufficiently reflects future actual experience.



## **2.2 Claims Tracking and Monitoring**

Claims tracking and monitoring is driven by three key components<sup>6</sup>. The first component is a periodic analysis of actual experience to expected assumptions. Most major companies perform this component. This is done by analyzing actual to expected ratios. If expected assumptions reflects actual experience sufficiently, the plot should be close to one. However, this statement does not indicate how close to one is defined. This may mislead one's judgement on the soundness of the expected assumption in representing actual experience.

The second component is to identify the block of business whose actual experience differs significantly from expected assumptions. Authors<sup>6</sup> used confidence intervals for this purpose where actual experience of a block of business that fell outside of this confidence interval was considered departing significantly from expected assumptions. This is a good early warning indicator for an insurance company and appropriate actions could be taken to manage the risk.

The last component follows from the second component, where if there is a significant difference between actual experience and expected assumptions, it needs to be identified if it is a one-time only event or a trend. Authors<sup>6</sup> used student's t-test for this purpose. If it is the latter, the research work we are proposing develops a statistically rigorous method to adjust the initial expected assumptions to provide an adjusted set of expected assumptions that will better reflect future experience.

## 2.3 LTC Incidence Rates and Poisson Generalized Linear Models (GLM)

Our research focuses on a statistically rigorous methodology on how to adjust expected LTC incidence rates that have been created in a GLM model. LTC incidence rates is the number of new claims by healthy policies over a specific period of time. The SOA LTC study<sup>5</sup> showed that it depends on many factors or variables namely issue date, date of birth, gender, underwriting class, underwriting type, marital status, coverage, benefit period, elimination period, claim incurred date, claim type and paid amount. These many factors make it impossible to obtain the rates directly from the raw data. This is because each variable has many levels which results in a lot of combinations and hence makes it impossible to get rates for each combination. Therefore, a systematic method to produce the rates which at the same time allows the rates to depend on these many factors must be employed. Poisson GLM was used by the SOA study<sup>5</sup>, moreover it has been employed in practice, to model LTC incidence rates.

Poisson distribution satisfies three key components to be under a GLM framework. Let  $Y$  be a Poisson random variable, usually a count variable, with mean  $\mu$  with  $X$  being its respective independent variables vector. We have the following:

- (i) It is from an exponential family since its probability density function (pdf) can be written in the form of

$$f(y; \theta, \phi) = \exp \left( \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right).$$

Since  $Y$  is a Poisson random variable its pdf is as follows:

$$f(y, \mu) = \frac{\exp(-\mu) \mu^y}{y!} = \exp(y \log(\mu) - \mu - \log(y!)), \text{ with}$$

$$\theta = \log(\mu), \quad b(\theta) = \mu, \quad \phi = 1, \quad a(\phi) = 1 \quad \text{and} \quad c(y, \phi) = -\log(y!).$$

- (ii) Its variance is a function of its mean,  $\text{Var}(Y) = \mu$ .
- (iii) Its link function,  $\eta(\cdot)$ , which is a one-to-one function that connects  $\mu$  and linear predictor,  $\beta X$ , is a log function. Hence these equalities hold:

$$\eta(\mu) = \log(\mu) = \beta X = \sum_{i=1}^n \beta_i X_i \in (-\infty, \infty)$$

$$\eta^{-1}(\mu) = \mu = \exp(\beta X) = \exp\left(\sum_{i=1}^n \beta_i X_i\right) \in (0, \infty)$$

In the above equalities,  $\beta$  represents the vector of coefficients obtained from an estimation technique such as maximum likelihood or Bayesian methods. Since the estimation technique solves  $\beta$  that satisfies the optimal criteria of the employed estimation method, there is no constraint for the linear predictor value, and therefore it can take any real number. On the other hand,  $\mu$  represents the mean value of  $Y$  which only takes non-negative values, thus the transformation resulting from applying the log link function to the linear predictor guarantees  $\mu$  to stay in the appropriate domain. The number of independent variables that constructed  $X$  which is considered for a particular estimation is  $n$ . This means  $\beta_i$  could be a value or a vector depending on  $X_i$ . For example, if  $X_i$  is a continuous variable,  $\beta_i$  produced by the estimation is just a value, and if  $X_i$  is a categorical variable with  $k \geq 2$  levels,  $\beta_i$  is a vector with  $k-1$  elements since one level will be considered as a reference group while the other levels will be compared to this reference group.

A minor modification to the above composition needs to be done for modelling rates of an event such as incidence rates rather than its count. In this situation, instead of just considering count, its exposure is also brought in for modelling. As seen earlier, incidence rates is the number of new claims by healthy policies over a specific period of time. Number of new claims is the count variable and healthy policies is the exposure. Let  $Y_i$  be a random variable representing number of new claims for observation  $i$ . It follows a Poisson distribution with mean  $\mu_i$ , exposure  $t_i$  and vector of independent or explanatory variables  $X_i = (X_1^i, X_2^i, \dots, X_n^i)$ . Therefore expected incidence rates for observation  $i$ ,  $IR_i$ , is given by the following:

$$\log\left(\frac{\mu_i}{t_i}\right) = \log(IR_i) = \sum_{k=1}^n \beta_k X_k^i \quad (1)$$

$$IR_i = \exp\left(\sum_{k=1}^n \beta_k X_k^i\right) = \prod_{k=1}^n \exp(\beta_k X_k^i) = \prod_{k=1}^n f_k^i \quad (2)$$

Equation (2) implies that incidence rates for observation  $i$  can be written as a product of  $n$  factors,  $f_k^i$ . Exposure is also known as an off-set whose coefficient value is known and not to be estimated along with other coefficients on the left hand side of equation (1). This means equation (1) can also be written as

$$\log(\mu_i) = \log(t_i) + \sum_{k=1}^n \beta_k X_k^i = \beta_0^i + \sum_{k=1}^n \beta_k X_k^i. \quad (3)$$

This work uses Poisson GLM as a tool for modelling LTC incidence rates, hence estimation coefficients and other statistical properties of a GLM framework are not discussed here. There are a lot of computer packages including those which are personalized by companies for performing GLM. Given a random variable that comes from an exponential family, an appropriate link

function has to be chosen before applying GLM, and also recognizing whether an off-set is required.

Statistical package R is used for this work entirely. The simulation process for this research is very intensive. Not only each simulation requires a generation of a lot of data (a data set of 10,000 healthy policyholders followed for 5 or 7 years) but it also needs to be repeated for 100 times to calculate the various error probabilities. A mere looping routine will not help in terms of speed and efficiency of the program. A brilliant way to exploit vectors and matrices is crucial to have the extensive programs run as efficient as possible. This is precisely the coding approach we used for the whole research.

## Chapter 3: Methodology

As mentioned previously, the proposed methodology to adjust expected assumptions for LTC incidence rates is parallel to an existing method, stochastic DAC technique. The main difference is that, while stochastic DAC technique is used to adjust the DAC balance for a variable annuity product, our methodology is for adjusting expected assumptions. Both of them have the same aim which is to provide a particular estimation value that reflects actual experience sufficiently, and hence reduce the volatility of the difference between actual experience and expected assumptions.

Both stochastic DAC and our methodology are very similar in two aspect. First, they use a confidence interval approach to recognize significant deviation in actual experience from expected assumptions. Secondly, they use a gradual approach to adjust expected assumptions when actual experience is significantly different from expected by letting the adjusted expected assumptions parameter to be the closest boundary of the confidence interval of the actual experience parameter.

### 3.1 Stochastic DAC Unlocking for Variable Annuities Technique<sup>4</sup>

Variable annuities are products with volatile profit. Since they are an increasingly important share of assets under management for many insurance companies, their deferred acquisition cost (DAC) has to be administered properly. DAC is a cost which is considered an asset that is amortized in accordance to future profits (margins) generated by the variable annuity block of business. DAC at the end of a period is defined as follows:

$$\begin{aligned}
& \text{DAC at end of period} = \text{DAC at the beginning of period} \\
& \quad + \text{Interest earned on DAC} \\
& \quad + \text{New capitalizations} \\
& \quad - \text{DAC amortization} \\
& \quad \pm \text{DAC catch-up}
\end{aligned} \tag{4}$$

Revised or unlocked DAC balance is calculated at time  $t$  by multiplying the present value of future margins with the respective revised amortization rate. The difference between current DAC and unlocked DAC is what is called the DAC catch-up. Unlocked DAC is a revised DAC which is considered as what the actual DAC is supposed to be. Since DAC is an asset, positive DAC catch-up contributes to GAAP earnings while negative DAC catch-up reduces GAAP earnings.

The stochastic DAC unlocking technique provides an approach to control the DAC catch-up so it reduces earnings volatility for a company relative to market volatility to recognize the long-term nature of annuity contracts. The following are steps to perform the stochastic DAC unlocking technique:

1. Stochastically generate future equity returns.
2. For each set of generated returns, determine future profit margins and the corresponding amortization rate to amortize the current DAC balance.
3. Calculate the new unlocked DAC balance based on the new amortization rate in (2).
4. Repeat the above for each set of stochastically set generated equity returns to generate a distribution of unlocked DAC balances.

5. A confidence band (i.e. corridor) is then constructed around the mean of the distribution of the distribution of unlocked DAC balances.
6. The current DAC balance is compared to the distribution constructed in (5) with the following rules:
  - a) If current DAC balance falls within the corridor no catch-up results.
  - b) Otherwise, the catch-up equals the amount needed to bring the current DAC balance to the nearest corridor boundary.

### **3.2 Proposed Methodology**

Similar to the SOA LTC Intercompany Study<sup>5</sup>, most of the expected assumptions for LTC are presented in tabular form and are obtained from GLM model. This includes LTC incidence rates which are modeled by a Poisson GLM whose components are displayed in tables to be appropriately multiplied in order to give the final rates for the respective risk.

The proposed methodology provides an objective approach to adjust expected assumptions of LTC incidence rates, mainly for LTC insurance claims tracking and monitoring. It can be also applied to problems related to adjusting a set of initial values as long as it is under a GLM framework. This method establishes two steps for adjusting LTC incidence expected assumptions. The first step is related to the model assumption for the initial expected assumptions and the second step is the adjustment procedure.



### 3.2.1 Model Assumption of Initial Expected Assumptions

The components of LTC incidence rates expected assumptions given in tabular form is derived from a Poisson GLM whose complete model is known. If  $Y_i$  is the number of new claims for observation  $i$ , and it follows a Poisson distribution with mean  $\mu_i$ , exposure  $t_i$  and a set of explanatory variables  $X_i = (X_1^i, X_2^i, \dots, X_n^i)$ , then the corresponding incidence rates,  $IR_i$ , is given as follows:

$$\log(IR_i) = \log\left(\frac{\mu_i}{t_i}\right) = \sum_{k=1}^n \beta_k X_k^i \quad (5)$$

$$IR_i = \exp\left(\sum_{k=1}^n \beta_k X_k^i\right) = \prod_{k=1}^n \exp(\beta_k X_k^i) = \prod_{k=1}^n f_k^i \quad (6)$$

Equation (6) implies that the incidence rates can be written as a product of  $n$  factors,  $f_k^i$ , as represented by the tabular form of expected assumptions. Everything is known in equation (6) including the beta parameters. It is seen as the standard rates which might have been obtained from the whole LTC insurance industry or from the aggregate dataset of several companies that is believed to sufficiently reflect the overall LTC incidence rates. The model above not only was used by the SOA study<sup>3</sup> but also has been widely adopted by the LTC insurance industry.

This assumption is the basis of the proposed methodology which is parallel to the continuation of the use of a GLM framework to adjust the LTC incidence rates expected assumptions in the next phase. It assumes the same model with exactly the same structure except that the beta coefficients are unknown and are to be estimated through the Poisson GLM using the actual experience dataset.

The main idea of this proposed methodology is to compare the expected assumptions to the actual experience and combine these two statistically so a balance in the general industry and company specific trends can be achieved. This concept is similar to the stochastic DAC technique in that it uses a confidence interval to determine if the actual-to-expected deviation is significant and adjusts the expected value up to the confidence limits instead of average value obtained from mean of the random variable of interest and its respective exposure. Extension and variation to this approach provides a practical approach on how any expected actuarial assumption can be adjusted to reflect a company's own experience.

In the situation where only expected assumptions in tabular form is available, simulation will be used to generate a dataset from the expected assumptions. Then a Poisson GLM will be applied to model incidence rates and the model obtained through this procedure will be used as the model mentioned in equations (5) and (6) for this step.

### **3.2.2 Adjustment Procedure**

This proposed methodology is aligned to the DAC unlocking technique. Given an actual experience dataset, the following explains a step-by-step procedure to adjusting expected assumptions of LTC incidence rates:

1. Fit a Poisson GLM to the dataset to model the actual incidence rates. The fitted model should be in the same exact form as one in the expected assumptions which is represented by equations (5) and (6).
2. From step 1, a set of fitted beta parameters are produced with their respective standard error. Construct a confidence interval for each of the fitted beta.

3. For every confidence interval produced in step 2, compare it with its respective beta from the expected assumptions.
4. For each confidence interval constructed in step 3, let  $\beta_E$ ,  $\beta_L$  and  $\beta_U$  be the beta from the expected assumptions, the lower bound of the confidence interval in step 2 and the upper bound of the confidence interval in step 2, respectively. The adjustment is done based on the following rules:
  - (i) If  $\beta_E < \beta_L$ , then the adjusted beta is equal to  $\beta_L$ .
  - (ii) If  $\beta_E > \beta_U$ , then the adjusted beta is equal to  $\beta_U$ .
  - (iii) If  $\beta_L \leq \beta_E \leq \beta_U$ , then no adjustment needed, in other words, the adjusted beta is equal to  $\beta_E$ .
5. Use the adjusted expected assumptions derived from the adjusted betas as the revised expected assumptions for the incidence rates.

The wider the confidence interval, the more likely beta from the expected assumptions is accepted which means the less likely the adjustment will transpire. In other words, more credibility is put to the actual experience by a narrower confidence interval. The width of the confidence interval depends on two factors which are the level of significance,  $\alpha$ , which can be regarded as the complement of the company's threshold for deviation of the actual experience from the expected assumptions, and uncertainty in the actual experience, the standard error of beta,  $\hat{\sigma}_\beta$ .

A larger company is more likely to want to put greater reliability on its own experience and that would imply a smaller confidence interval or larger  $\alpha$ . A large  $\hat{\sigma}_\beta$  represent high

volatility in the beta parameter estimate which would widen the confidence interval and reduce the chance that the beta from expected assumptions will be adjusted.

### 3.3 Reliability of the Proposed Methodology

Significant departure from initial expected assumptions used in LTC insurance pricing or reserving could result in intense financial consequences for a company. This method produces not only an objective approach to adjusting expected assumptions of LTC incidence rates, but also an easy and efficient way to track and monitor LTC insurance claims. While the richness of the GLM framework and the statistical rigor of using confidence interval makes this methodology very applicable, its reliability needs to be verified first before it can be used by companies and researchers.

For the technique to be reliable it has to be stable and behave as expected. Three errors are used to justify the reliability of this method, namely Type I, Type II and Type III errors.

#### 3.3.1 Type I Error

Given that the expected assumptions do significantly reflect the actual experience, applying the method many times should produce a type I error which is very close to the complement of the company's threshold, i.e.  $\alpha$ . Specifically, the type I error measures the probability of rejecting beta from the expected assumptions when actual experience equals the expected assumptions. Let  $N$  be the number of simulations, then type I error is defined as follows:

$$\text{Type I Error} = \frac{\sum_{j=1}^N (1 - 1_{s_j})}{N} \quad (7)$$

$$1_{S_I} = \begin{cases} 1; & \beta_E \in S_I \\ 0; & \beta_E \notin S_I \end{cases} \quad (8)$$

$\beta_E$  = Beta from the expected assumptions

$S_I$  = Simulated confidence interval for beta

The following is an algorithm to obtain the type I error assuming a closed block of LTC insurance business:

1. Generate 10,000 random LTC healthy policies. Each observation has a specific characteristic represented by the explanatory variables being used in the GLM model.
2. Simulate new claims over the next 5 years based on the incidence rates of the initial expected assumptions.
3. Fit a Poisson GLM to the simulated claim data, with exposure being the off-set so the model will produce expected rates instead of expected new claims. In order to do this, a simplified assumption that all exposure for each observation equals one is used i.e. we are using counts to measure exposure versus benefit amounts.
4. Fix a significance level  $\alpha$  and construct a confidence interval for each estimated beta obtained from step 3.
5. Compare the confidence interval in step 4 to its respective beta from the initial expected assumptions and give a value either one or zero based on equation (8).
6. Repeat steps 1 to 6 for  $N$  times and apply equation (7) to estimate the type I error.

### 3.3.2 Type II Error

If the expected assumptions do not sufficiently reflect actual experience, then there exists at least one beta from the expected assumptions that should be significantly different from its respective beta obtained from actual experience. Applying the method many times should produce a type II error which is very close to the theoretical type II error. Specifically, type II error measures the probability of accepting beta from the expected assumptions when actual experience does not equal the expected assumptions. Let  $N$  be the number of simulations, then type II error is defined as follows:

$$\text{Type II Error} = \frac{\sum_{j=1}^N 1_{S_{II}}}{N} \quad (9)$$

$$1_{S_{II}} = \begin{cases} 1; & \beta_E \in S_{II} \\ 0; & \beta_E \notin S_{II} \end{cases} \quad (10)$$

$\beta_E$  = Beta from the expected assumptions

$S_{II}$  = Simulated confidence interval for beta

The following is an algorithm to obtain the type II error assuming a closed block of LTC policies:

1. Generate 10,000 random observations. Each observation has a specific characteristic represented by explanatory variables.
2. Compute incidence rates for each observation based on the initial expected assumptions.

3. Fix an explanatory variable that will be distorted in order to represent the discrepancy between the actual experience and the initial expected assumptions.
4. Fix a shock factor at a specific level for the explanatory variable selected in step 3.
5. Produce a shocked incidence rate by multiplying the incidence rates obtained in step 2 by this factor for all observations which are affected by the shocked explanatory variable and value produced in steps 3 and 4. Otherwise, the incidence rates remain the same as obtained in step 2.
6. Generate new claims for the set of 10,000 healthy policies over the next 5 years based on the shocked incidence rates obtained in step 5.
7. Fit a Poisson GLM to the new claim with exposure being the off-set so the model will produce expected rates instead of expected new claim. In order to do this, a simplified assumption that all exposure for each observation equals one is used.
8. Fix a significance level  $\alpha$  and construct a confidence interval for each estimated beta obtained from step 7.
9. Compare the confidence interval in step 8 to its respective beta from the initial expected assumptions and give value either one or zero based on equation (10).
10. Repeat steps 1 to 9 for  $N$  times and apply equation (9) to estimate the type II error.

### 3.3.3 Type III Error

Assume there is a need to adjust the initial expected assumptions and this proposed method is employed for changing the respective beta from the initial expected assumptions to the revised form. Then, applying the method many times using the company's own experience should produce the type III error. Specifically, type III error measures the probability of rejecting beta from the

adjusted expected assumptions when actual experience does not equal the expected assumptions. From one aspect it is similar to the type I error by changing the initial expected assumptions to the adjusted version. However the difference between them is that the type III error also captures the adjustment criteria we are employing by this proposed technique. Let  $N$  be the number of simulations, then the type III error is defined as follows:

$$\text{Type III Error} = \frac{\sum_{j=1}^N (1 - 1_{S_{III}})}{N} \quad (11)$$

$$1_{S_{III}} = \begin{cases} 1; & \beta_A \in S_{III} \\ 0; & \beta_A \notin S_{III} \end{cases} \quad (12)$$

$\beta_A$  = Adjusted beta

$S_{III}$  = Simulated confidence interval

The following is an algorithm to obtain type III error assuming a closed block of LTC insurance business:

1. Generate 10,000 LTC healthy policies. Each observation has a specific characteristic represented by the explanatory variables being used in the GLM model.
2. Compute incidence rates for each observation based on the initial expected assumptions.
3. Fix an explanatory variable that will be distorted in order to represent the discrepancy between the actual experience and the initial expected assumptions.
4. Fix a shock factor at a specific level for the explanatory variable selected in step 3.



5. Produce a shocked incidence rate by multiplying the incidence rates obtained in step 2 by this factor for all observations which are affected by the shocked explanatory variable characteristic produced in steps 3 and 4. Otherwise, the incidence rates remain the same as obtained in step 2.
6. Generate new claims for the set of 10,000 healthy policies over the next 7 years based on the shocked incidence rates obtained in step 5.
7. Separate these new claims into two groups; the first 5 years and the last 2 years and call it training data and test data, respectively.
8. Fit a Poisson GLM to the new claims in step 7 for training data with exposure being the off-set so the model will produce expected rates instead of expected new claims. In order to do this, a simplified assumption that all exposure for each observation equals one is used.
9. Fix a significance level  $\alpha_1$  and construct confidence intervals for each estimated beta obtained from step 8.
10. Compare the confidence interval in step 9 to its respective beta from the initial expected assumptions and give a value of either one or zero based on equation (8). If beta from the initial expected assumption falls outside of the confidence interval, adjust it based on the proposed method, otherwise keep the initial value. This produces the adjusted expected assumptions.
11. Fit a Poisson GLM to the new claims in step 7 for the test data with exposure being the off-set so the model will produce expected rates instead of expected new claim. In order to do this, a simplified assumption that all exposure for each observation equals one is used.

12. Fix a significance level  $\alpha_2$  and construct a confidence interval for each estimated beta obtained from step 11.
13. Compare the confidence interval in step 12 to its respective beta from the adjusted expected assumptions and give a value either one or zero based on equation (12).
14. Repeat steps 1 to 13 for  $N$  times and apply equation (11) to estimate the type III error.

### 3.4 Proposed Methodology with Credibility Effect

As we seen before, the level of significance  $\alpha$  can also be viewed as a measure of credibility of a company. The smaller the value of  $\alpha$ , the wider the size of the confidence interval, in other words the smaller  $\alpha$  increases the likelihood to accepting the beta coefficient from the initial expected assumptions. In other words, more trust is put on the expected assumptions if the confidence interval is wide. On the other hand, more trust is put on the actual experience if the confidence interval is narrow. This leads to the question of how to decide a reasonable value of  $\alpha$ .

This is the motivation for bring in credibility into the proposed methodology. The credibility approach provides a method for estimating parameters of a subset of a given population by combining results for that particular subset with results for the population as a whole, which is larger, and more statistically stable<sup>8</sup>. Let  $\hat{\theta}$ ,  $\theta_E$  and  $Z$  be the estimate from the sample (actual experience), estimate from the population (expected assumptions) and credibility factor respectively. Then the credibility or updated estimate,  $\theta_C$ , is given by:

$$\theta_C = Z\hat{\theta} + (1-Z)\theta_E \quad (13)$$

$$0 \leq Z \leq 1 \quad (14)$$

Assume that  $\hat{\theta}$  and  $\theta_E$  are normally distributed with variance  $\sigma_{\hat{\theta}}^2$  and  $\sigma_{\theta_E}^2$  respectively.

Let  $n_s$  be number of sample (actual experience) observations, so the credibility factor  $Z$  based on the greatest accuracy method is given by:

$$Z = \frac{n_s}{n_s + \left(\sigma_{\hat{\theta}}^2 / \sigma_{\theta_E}^2\right)} \quad (15)$$

Let  $\hat{\beta}$  be the estimate of beta by the Poisson GLM from the actual experience and  $\beta_E$  be the estimate from the expected assumptions as seen in section 3.2.2 with variance of  $\hat{\beta}$  and  $\beta_E$  denoted by  $\sigma_{\hat{\beta}}^2$  and  $\sigma_{\beta_E}^2$ , respectively. If  $\Phi(\cdot)$  represents the cumulative distribution function of the Standard Normal Distribution and  $\Phi^{-1}(\cdot)$  is its inverse, then with credibility effect  $\alpha$  in the initial proposed methodology is replaced by  $\alpha_C$  as follows:

$$\alpha_C = \left(1 - \Phi\left(\frac{\sigma_{\hat{\beta}}^2}{\sigma_{\beta_E}^2}\right)\right) \times 2 \quad (16)$$

Equation (16) implies that the size of confidence interval which is equivalent to the company's credibility level is now justified by the variation in  $\hat{\beta}$  and  $\beta_E$ . High variation in  $\hat{\beta}$  leads to a small  $\alpha_C$  (i.e. wider confidence interval) which means less credibility is given to the actual experience and hence it is highly likely to retain  $\beta_E$  as the adjusted beta as given by step 4 of the adjustment procedure in section 3.2.2. On the other hand, high variation in  $\beta_E$  leads to a big

$\alpha_C$  which means more credibility is given to the actual experience and hence it is highly likely to adjust  $\beta_E$  by replacing it with the closest boundary of the confidence interval of  $\hat{\beta}$ .

Generally, adjusted beta  $\beta_A$  can be written as the following:

$$\beta_A = Z \hat{\beta} + (1 - Z) \beta_E \quad (17)$$

The value of  $Z$  in equation (17) is conditioned to the following rules:

- (i) If  $\beta_E < \beta_L$ , then  $Z = \frac{\beta_L - \beta_E}{\hat{\beta} - \beta_E}$ .
- (ii) If  $\beta_E > \beta_U$ , then  $Z = \frac{\beta_E - \beta_U}{\beta_E - \hat{\beta}}$ .
- (iii) If  $\beta_L \leq \beta_E \leq \beta_U$ , then  $Z = 0$ .

This approach sets a better insight to establishing a reasonable credibility threshold prior to implementing the proposed adjustment methodology as a whole. Consequently, more options could be created to set up the credibility threshold either by using a prevailing  $\alpha$  (such as 1%, 5% or 10%), personal judgement (such as a senior actuary who knows the company well) or  $\alpha_C$  as given by equation (16). If we want the confidence interval for a given company to be fixed for all betas, then the selected  $\alpha_C$  could be the maximum, minimum or average of  $\alpha_C$  over all betas.

## **Chapter 4: Analyses and Results**

### **4.1 Establishment of Initial Expected Assumptions of LTC Incidence Rates**

The prior expected assumptions of LTC incidence rates for this work stems from the SOA study<sup>3</sup>. However, there are two issues that arise from this. One, the expected assumptions was obtained through a Poisson GLM but the complete model remains confidential, so for that reason, factor values according to each level of explanatory variable is only available in tabular form. The product of these factors corresponding to specific characteristics of the risk provides the complete expected incidence rates for the respective risk. Two, there are too many explanatory variables in the study, namely issue date, coverage, date of birth, benefit period, gender, elimination period, underwriting class, claim incurred date, underwriting type, claim type, marital status and paid amount. This great number of explanatory variables hinders the efficiency of this research since the fundamental aim of this work is to produce a general methodology which further can be used at any level of desired complexity.

As a result, this work first establishes an initial expected assumption that suits equations (5) and (6) in chapter 3. The number of explanatory variables is reduced to four, namely gender, benefit amount (or paid amount as used by the study), age (derived from date of birth) and elimination period. The choice of these variables is solely to demonstrate our research methodology. The factors were translated into betas by taking their inverse log. However, these

betas did not produce an incidence rate formula which should have been produced by Poisson GLM because some reference groups have a factor value greater than one.

Therefore, a dataset consisting of 10,000 LTC healthy policies is generated and tracked for 5 years based on these betas, and incidence rates is calculated for each of them. Then new claims for this 5 years dataset was generated randomly by generating a random number within interval (0,1) and comparing it to the computed incidence rates. Finally, a Poisson GLM was fitted to the new claims with the exposure being the off-set so the model will produce expected rates instead of expected new claims. A simplified assumption that all exposure for each observation equals one was used.

Explanatory Variable	Level
Gender	Female
	Male
Benefit Amount	Less than \$100
	\$100 or more
Age	[0, 65)
	[65, 75)
	[75, 85)
	85 and older
Elimination Period	Less than 6 months
	6 months or more

Table 1: Explanatory variables with their respective levels

Table 1 shows the chosen explanatory variables at their respective levels. Several levels for each initial chosen explanatory variables were aggregated, again in order to simplify the model to demonstrate our research.

Explanatory Variable	Beta Coefficient	Initial Expected Assumption	Standard Error
Intercept ( $\beta_1$ )	Intercept	-7.027	0.391
Gender ( $\beta_2$ )	Female	0	-
	Male	-0.315	0.397
Benefit Amount ( $\beta_3$ )	Less than \$100	0	-
	\$100 or more	-0.032	0.392
Age ( $\beta_4$ )	[0, 65)	0	-
	[65, 75)	0.711	0.443
	[75, 85)	-0.127	0.636
	85 and older	-1.815	1.035
Waiting/Elimination Period ( $\beta_5$ )	Less than 6 months	0	-
	6 months or more	-0.794	0.425

Table 2: Estimate beta coefficients with their respective standard errors

Table 2 shows the value of each beta coefficient for the established initial expected assumptions of this LTC incidence rates. Beta coefficients with value zero are considered as a reference group under Poisson GLM. The following is the model in Table 2 in terms of equation (6) in chapter 3:

$$IR_i = \exp \left( \sum_{k=1}^5 \beta_k X_k^i \right) = \prod_{k=1}^5 \exp(\beta_k X_k^i) = \prod_{k=1}^5 f_k^i \quad (13)$$

$$\beta_1 = -7.027$$

$$\beta_2 = (0, -0.315)$$

$$\beta_3 = (0, -0.032)$$

$$\beta_4 = (0, 0.711, -0.127, -1.815)$$

$$\beta_5 = (0, -0.794)$$

For example, a full model for observation  $A$  who is a female age 67 with benefit amount \$200 per day and elimination period 3 months is as follows:

$$IR_A = \exp^{-7.027} \times \exp^0 \times \exp^{-0.032} \times \exp^{0.711} \times \exp^0$$

$X_1^i$  is the explanatory variable attached to beta coefficient  $\beta_1$ , and is equals one for all  $i$  which implies  $\beta_1$  represents the intercept of this model.  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$  and  $\beta_5$  are beta coefficients corresponding to explanatory variables gender, benefit amount, age and elimination period, respectively. Beta equals zero implies a reference group and is not estimated by Poisson GLM thus does not have standard error.

## 4.2 Verification of the Reliability of the Proposed Methodology

As mentioned in section 3.3 three statistical errors are used to verify the reliability of the proposed expected assumptions adjustment methodology namely Type I, II and III errors.

### 4.2.1 Type I Error

Type I error measures the probability of rejecting beta from the expected assumptions when actual experience equals expected assumptions. This work uses three common alpha values used in statistics which are 10%, 5% and 1%. No type I error is computed for the betas of the reference group.



Beta	Type I Error		
	$\alpha=10\%$	$\alpha=5\%$	$\alpha=1\%$
$\beta_1$	0.09	0.02	0
$\beta_2$ ( Male)	0.10	0.05	0
$\beta_3$ (\$100 or more)	0.12	0.05	0
$\beta_4$ ([65,75))	0.06	0.03	0.01
$\beta_4$ ([75,85))	0.08	0.03	0
$\beta_4$ (85 and older)	0.12	0.05	0.02
$\beta_5$ (6 months or more)	0.12	0.06	0

Table 3: Type I error from simulations

Table 3 shows that the type I error is close to the chosen alpha and this is indeed to be expected since the generated dataset is assumed to follow the expected assumptions. A consistent type I error sets a solid foundation for the use of a simulation technique in establishing the proposed methodology. Appendixes A1, A2 and A3 show the type I error simulation of confidence intervals produced through an R statistical package for  $\alpha=10\%$  ,  $\alpha=5\%$  and  $\alpha=1\%$  , respectively. Blue and green dots represent the initial expected assumptions (called ‘SOA’ in the graph) and the fitted beta based on simulated data (called ‘Fitted’ in the graph), respectively. Simulated data represents actual experience and a confidence interval is constructed for the estimated beta every time a Poisson GLM is fitted to it.

#### 4.2.2 Type II Error

Type II error measures the probability of accepting beta from the expected assumptions when actual experience does not equal the expected assumptions. In order to compute this error, incidence rates of females is shocked by multiplying the initial incidence rates by a constant value 10. This work uses three common alpha values used in statistics which are 10%, 5% and 1%. No type II error is computed for betas of the reference group.

Beta	Type II Error		
	$\alpha=10\%$	$\alpha=5\%$	$\alpha=1\%$
$\beta_1$	0	0	0
$\beta_2$ ( Male)	0	0	0
$\beta_3$ (\$100 or more)	0.89	0.92	0.97
$\beta_4$ ([65,75))	0.92	0.95	0.98
$\beta_4$ ([75,85))	0.92	0.93	1
$\beta_4$ (85 and older)	0.90	0.96	1
$\beta_5$ (6 months or more)	0.90	0.96	1

Table 4: Type II error from simulations

When a significant shock is applied to females which is the reference group for the gender variable, the beta coefficient for males is expected to change. Male beta is expected to change because the ratio of male incidence rates to female incidence rates will now change due to the significant change in female incidence rates. Overall incidence rate changes follow from change in the male beta since it is a product of several factors as given by equation (2). Therefore the beta coefficient is expected to change in order to keep incidence rates for unaffected risk due to the shock at its original range. Table 4 shows that intercept and male betas have small type II error which is consistent with the presumed expectation. Other betas have big type II errors close to one which is also expected since these betas are not affected by the shock and hence are anticipated to be within their respective confidence intervals. Appendixes A4, A5 and A6 show the type II error simulation confidence intervals produced through an R statistical package for  $\alpha=10\%$ ,  $\alpha=5\%$  and  $\alpha=1\%$ , respectively. Blue and green dots represent the initial expected assumptions (called ‘SOA’ in the graph) and the fitted beta is based on simulated data (called ‘Fitted’ in the graph), respectively. Simulated data represents actual experience and a confidence interval is constructed for the estimate beta every time a Poisson GLM is fitted to it.

### 4.2.3 Type III Error

Type III error measures the probability of rejecting beta from the adjusted expected assumptions when actual experience does not equal the expected assumptions. In order to compute this error, incidence rates of females are shocked by multiplying the initial incidence rates by a constant value 10. In each simulation, the generated dataset is divided into two groups, the testing and the validation dataset. The testing group can be treated as historical experience and the validation group can be treated as actual experience. Therefore two alphas are needed. One alpha applies to the testing group where it is used to determine if the adjustment of beta from the initial assumptions is needed. The other alpha applies to the validation group to see whether the adjusted beta fall in the constructed confidence interval of actual experience. This work uses three common alpha values used in statistics which are 10%, 5% and 1% for both of the alphas. No type III error is computed for betas of the reference group.

Beta	Type III Error		
	$\alpha_1 = 10\%$	$\alpha_1 = 5\%$	$\alpha_1 = 1\%$
$\beta_1$	0.31	0.37	0.46
$\beta_2$ ( Male)	0.27	0.33	0.52
$\beta_3$ (\$100 or more)	0.10	0.10	0.10
$\beta_4$ ([65,75))	0.11	0.11	0.11
$\beta_4$ ([75,85))	0.08	0.08	0.08
$\beta_4$ (85 and older)	0.03	0.03	0.03
$\beta_5$ (6 months or more)	0.10	0.10	0.11

Table 5: Type III error from simulations ( $\alpha_2 = 10\%$ )

Beta	Type III Error		
	$\alpha_1 = 10\%$	$\alpha_1 = 5\%$	$\alpha_1 = 1\%$
$\beta_1$	0.21	0.27	0.38
$\beta_2$ ( Male)	0.10	0.14	0.35
$\beta_3$ (\$100 or more)	0.07	0.06	0.06
$\beta_4$ ([65, 75))	0.04	0.04	0.04
$\beta_4$ ([75, 85))	0.05	0.05	0.05
$\beta_4$ (85 and older)	0.02	0.02	0.02
$\beta_5$ (6 months or more)	0.06	0.06	0.06

Table 6: Type III error from simulations ( $\alpha_2 = 5\%$ )

Beta	Type III Error		
	$\alpha_1 = 10\%$	$\alpha_1 = 5\%$	$\alpha_1 = 1\%$
$\beta_1$	0.06	0.10	0.21
$\beta_2$ ( Male)	0	0	0
$\beta_3$ (\$100 or more)	0.02	0.02	0.02
$\beta_4$ ([65, 75))	0.01	0.01	0
$\beta_4$ ([75, 85))	0.01	0.01	0.01
$\beta_4$ (85 and older)	0	0	0
$\beta_5$ (6 months or more)	0.02	0.02	0.02

Table 7: Type III error from simulations ( $\alpha_2 = 1\%$ )

For betas that are significantly affected by the shock, it is expected that the bigger the first alpha the more appropriate adjustment can be made hence the less type III error is. For betas which are not significantly affected by the shock, type III error is expected to be close to the second alpha. Tables 5, 6 and 7 show the results of the type III error for each selected alpha value in the testing dataset given a fixed alpha in the validation dataset. It can be seen that the type III error increases as the first alpha decreases for the affected betas and is around the second alpha for the unaffected betas. This shows that the proposed methodology does provide a reliable adjustment procedure which gives the flexibility for a company to set up its own threshold limit which is equal

to  $1-\alpha_1$ . Big  $\alpha_1$  implies a small threshold which means a company only allows expected assumptions and actual experience to be different slightly. Big  $\alpha_1$  also implies a narrower confidence interval which increases the likelihood of adjusting the expected assumption, and hence make it closer to the actual experience.

### **4.3 Application of the Proposed Methodology**

The proposed methodology provides a statistically rigorous technique to adjust the expected assumptions and hence reduce the volatility arising from changing expected assumptions to reflect actual experience.

#### **4.3.1 Application of the Proposed Methodology for One Simulation**

In this part, seven years of data is generated based on the initial expected assumption. Then a shock is applied where female incidence rates are multiplied by a constant factor which equals 10. This seven years of data is divided into two group, the first five years represents historical data and the last two years represents actual experience.

The first five years is used for tracking and monitoring where an adjustment will be made to the initial expected beta if it is significantly different from the fitted beta obtained from the last two years. At the end of five years, there are two possible expected assumptions. One is the initial expected assumptions if the proposed methodology is not applied, and the second is the adjusted expected assumptions if the methodology is employed. The charts below show the comparison between initial expected incidence rates (and betas) and actual incidence rates (and betas) obtained at the end of seven years for the two situations, with and without the application of the proposed methodology at three different levels of alpha, 10%, 5% and 1%.

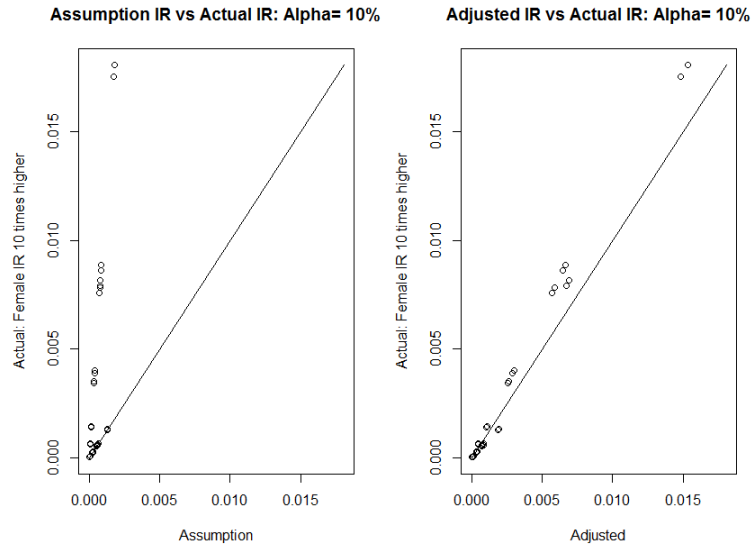


Figure 1: Expected assumptions incidence rates versus actual experience incidence rates with (right) and without (left) adjustment for  $\alpha=10\%$

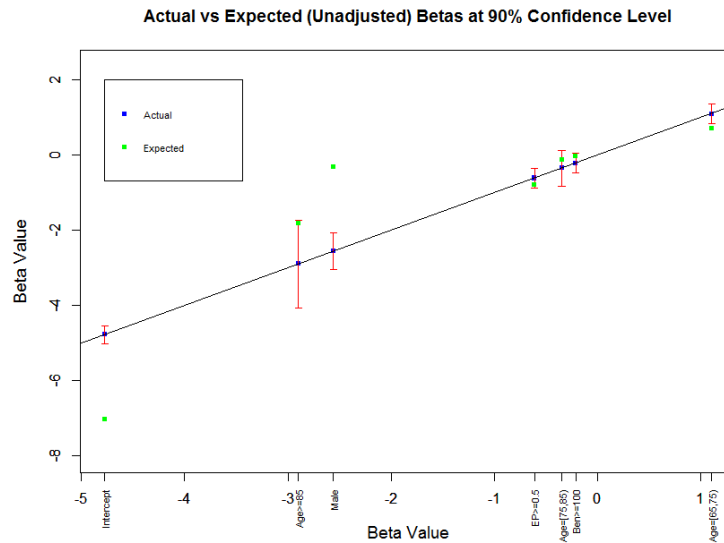


Figure 2: Expected assumptions betas versus actual experience betas (and their respective confidence interval) without adjustment for  $\alpha=10\%$

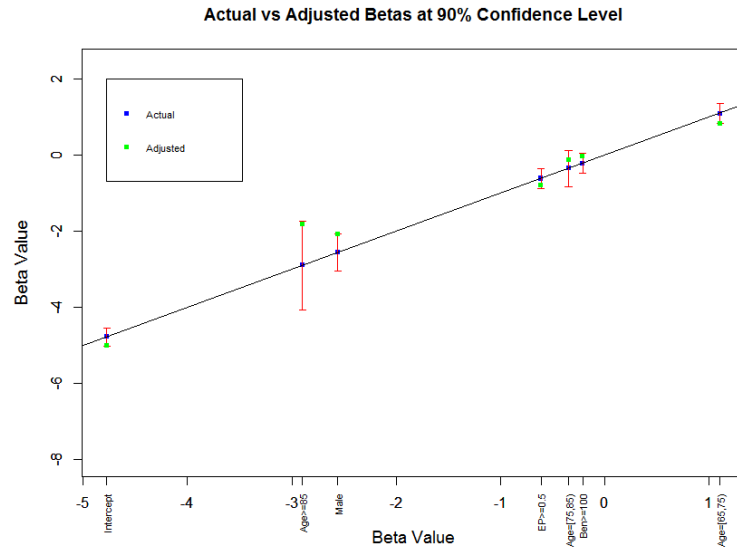


Figure 3: Expected assumptions betas versus actual experience betas (and their respective confidence interval) with adjustment for  $\alpha = 10\%$

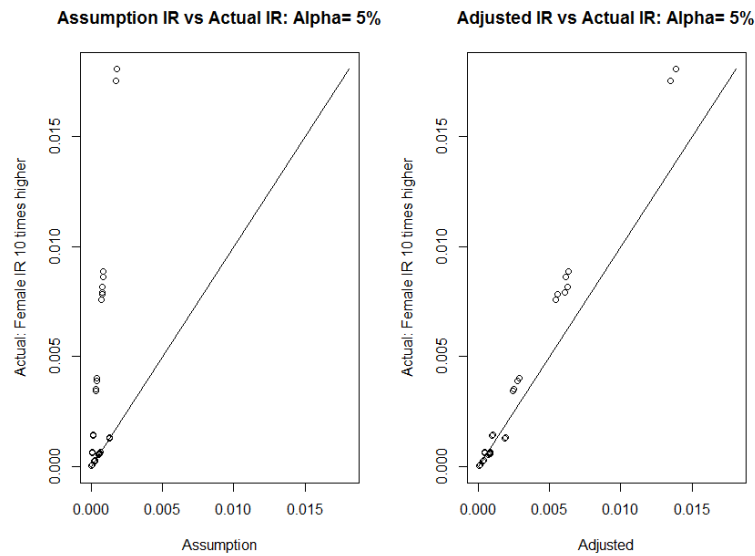


Figure 4: Expected assumptions incidence rates versus actual experience incidence rates with (right) and without (left) adjustment for  $\alpha = 5\%$

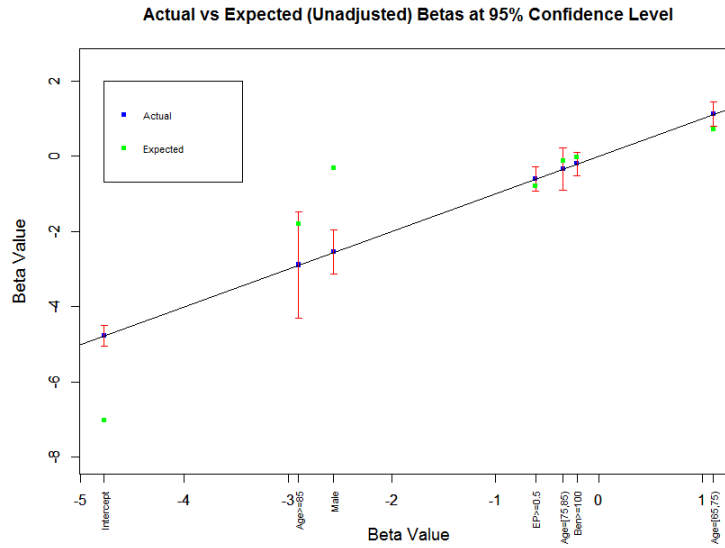


Figure 5: Expected assumptions betas versus actual experience betas (and their respective confidence interval) without adjustment for  $\alpha = 5\%$

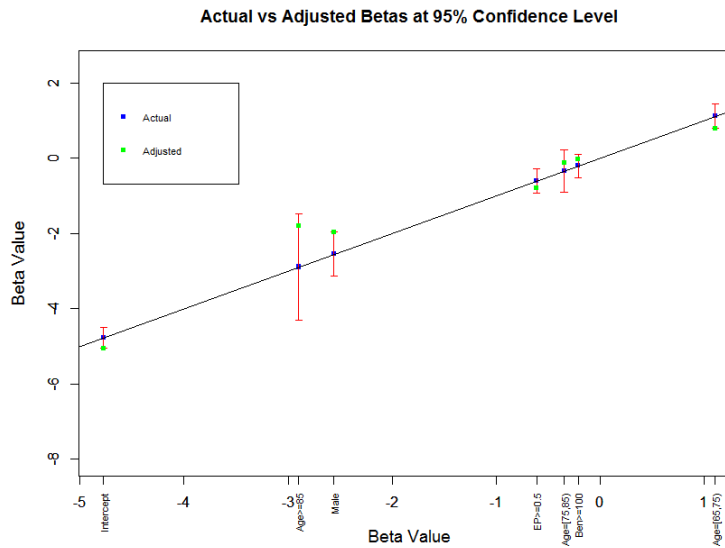


Figure 6: Expected assumptions betas versus actual experience betas (and their respective confidence interval) with adjustment for  $\alpha = 5\%$



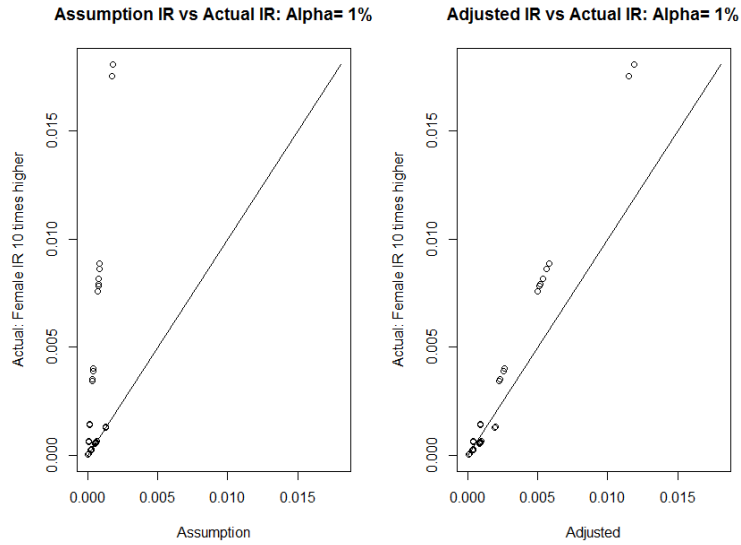


Figure 7: Expected assumptions incidence rates versus actual experience incidence rates with (right) and without (left) adjustment for  $\alpha=1\%$

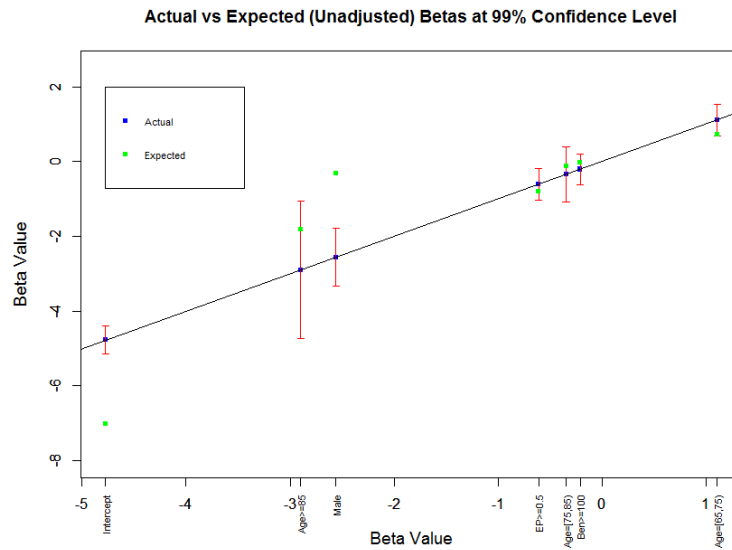


Figure 8: Expected assumptions betas versus actual experience betas (and their respective confidence interval) without adjustment for  $\alpha=1\%$

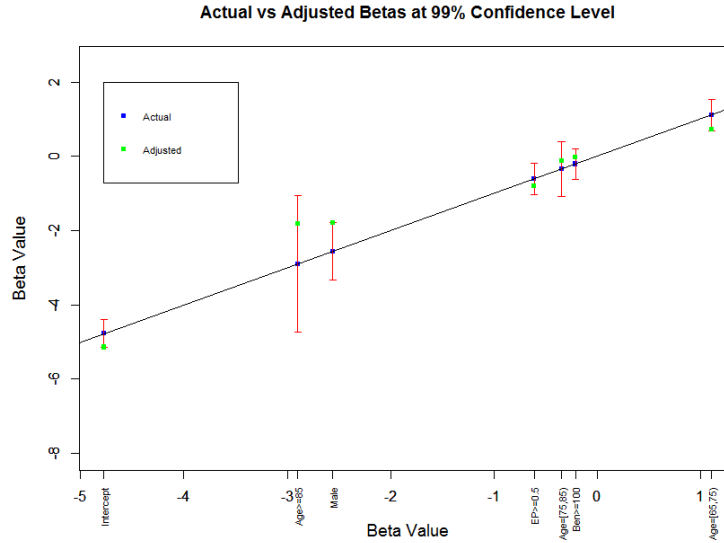


Figure 9: Expected assumptions betas versus actual experience betas (and their respective confidence interval) with adjustment for  $\alpha=1\%$

As mentioned previously, the bigger the alpha, the more likely an adjustment will be made to the expected assumptions and hence the closer it will be to actual experience. Figures 1, 4 and 7 show the applications of the proposed methodology at different alpha value equal to 10%, 5% and 1%, respectively. The left hand side of the three figures shows expected assumptions on the horizontal axis and actual experience on the vertical axis. If expected assumptions reflect actual experience sufficiently, the plots should be around the straight line  $y=x$ . However, the plots in the figures are mostly on the upper part of the line  $y=x$ . This shows that expected assumptions has way lower incidence rates than actual, and this happens due to the shock which is applied to female incidence rates which in this situation is considered as actual experience.

The right hand side of the figure shows adjusted expected assumptions on the horizontal axis and actual experience on the vertical axis. As can be seen, the plots now are closer to the line  $y=x$ . Also, the bigger the alpha the closer the plots are to the line  $y=x$ .

Figures 2, 5 and 8 show the unadjusted betas compared to the confidence interval of betas at confidence levels 90% (or  $\alpha=10\%$ ), 95% (or  $\alpha=5\%$ ) and 99% (or  $\alpha=1\%$ ), respectively. Figures 3, 6 and 9 show the similar adjusted results in terms of confidence intervals similar to the earlier confidence levels. The adjustment procedure brings adjusted betas closer to the actual ones.

#### 4.3.2 Application of the Proposed Methodology for One Hundred Simulations

The proposed method is applied to 7 years data as in section 4.3.1 and the same procedure is repeated one hundred times. The mean of the fitted (actual) and adjusted betas are obtained at  $\alpha=10\%$ ,  $\alpha=5\%$  and  $\alpha=1\%$ . Betas which are not included are the ones in the reference groups.

Beta	Mean		
	Initial Expected Assumptions	Actual	Adjusted
$\beta_1$	-7.027	-4.733	-4.973
$\beta_2$ ( Male)	-0.315	-2.661	-2.126
$\beta_3$ (\$100 or more)	-0.032	-0.002	-0.024
$\beta_4$ ([65,75))	0.711	0.677	0.710
$\beta_4$ ([75,85))	-0.127	-0.218	-0.138
$\beta_4$ (85 and older)	-1.815	-2.000	-1.805
$\beta_5$ (6 months or more)	-0.794	-0.804	-0.792

Table 8: Mean comparison among initial expected assumption, actual and adjusted betas at  $\alpha=10\%$

Beta	Mean		
	Initial Expected Assumptions	Actual	Adjusted
$\beta_1$	-7.027	-4.733	-5.019
$\beta_2$ ( Male)	-0.315	-2.661	-2.024
$\beta_3$ (\$100 or more)	-0.032	-0.002	-0.026
$\beta_4$ ([65,75))	0.711	0.677	0.710
$\beta_4$ ([75,85))	-0.127	-0.218	-0.132
$\beta_4$ (85 and older)	-1.815	-2.000	-1.812
$\beta_5$ (6 months or more)	-0.794	-0.804	-0.793

Table 9: Mean comparison among initial expected assumption, actual and adjusted betas at

$$\alpha = 5\%$$

Beta	Mean		
	Initial Expected Assumptions	Actual	Adjusted
$\beta_1$	-7.027	-4.733	-5.109
$\beta_2$ ( Male)	-0.315	-2.661	-1.823
$\beta_3$ (\$100 or more)	-0.032	-0.002	-0.031
$\beta_4$ ([65,75))	0.711	0.677	0.711
$\beta_4$ ([75,85))	-0.127	-0.218	-0.127
$\beta_4$ (85 and older)	-1.815	-2.000	-1.815
$\beta_5$ (6 months or more)	-0.794	-0.804	-0.794

Table 10: Mean comparison among initial expected assumption, actual and adjusted betas at

$$\alpha = 1\%$$

Tables 8, 9 and 10 show the results when the analysis in section 4.3.1 is repeated 100 times. This is done to examine how the proposed methodology behaves in terms of average results. As expected, higher alpha implies a narrower confidence interval and hence bigger adjustment to the affected expected assumptions. As a result, adjusted betas are closer to the actual ones for bigger alpha values as presented in the tables. Unaffected betas remain close around the original ones.

#### 4.4 Proposed Methodology with Credibility Effect

Alpha values of 10%, 5% and 1% are prevailing significant levels in any statistical-related analyses. In this part, as mentioned earlier, this value is related to a company's threshold. The bigger the alpha, the smaller the company threshold (confidence level of the confidence interval) and hence the greater the chance that the company's initial expected assumptions will be adjusted to reflect actual experience.

This part suggests the appropriate value of alpha obtained by equation (16) in chapter 3 to use for applying the proposed methodology instead of just the universal 10%, 5% and 1% alpha values. Two situations are generated to demonstrate the credibility effect which are stable and unstable actual experience. A five year dataset is generated and then it is shocked in two different ways to produce two different situations.

To produce a stable situation, a constant shock which equals 10 is applied to the five year dataset. Incidence rates of females for all five years are multiplied by this shock. On the other hand, to produce an unstable situation, a constant shock which equals 0.1 is applied to the years two and four dataset only with years one, three and five staying unchanged. Incidence rates of females for years two and four are multiplied by this shock.

As seen in equation (16) in chapter 3, credibility weighted alpha depends on variation in the expected assumptions and actual experience. More variation is implied by a high standard error in the beta estimation. If these variations are the same, the ratio  $\sigma_{\hat{\beta}}^2 / \sigma_{\beta_E}^2$  is equal to one which means alpha equals 32%. Therefore, the application of the proposed methodology using two

credibility alphas produced from the stable and unstable situations is compared to the application using alpha equal to 32%.

#### 4.4.1 Credibility Effect on Stable Dataset

Beta	$\beta_E$	$\beta_A$ (with $\alpha=32\%$ )	$\beta_A$ (with $\alpha_C$ )
$\beta_1$	-7.027	-4.769	-4.651
$\beta_2$ ( Male)	-0.315	-2.456	-2.575
$\beta_3$ (\$100 or more)	-0.032	-0.032	0.028
$\beta_4$ ([65, 75))	0.711	0.602	0.450
$\beta_4$ ([75, 85))	-0.127	-0.127	-0.266
$\beta_4$ (85 and older)	-1.815	-0.716	-0.510
$\beta_5$ (6 months or more)	-0.794	-0.794	-0.760

Table 11: Adjustment results using fixed and credibility alphas for stable dataset

Appendix A7 shows the complete version of Table 11. Standard error for each actual (shocked) beta,  $\sigma_{\hat{\beta}}$ , is smaller than the one for expected assumptions,  $\sigma_{\beta_E}$ . This makes the variance ratio of actual to expected,  $\sigma_{\hat{\beta}}/\sigma_{\beta_E}$ , less than one which produces a credibility alpha,  $\alpha_C$ , greater than the control alpha which is fixed at 32%. Credibility alpha bigger than 32% implies a narrower confidence interval which means more credibility is put on the actual experience.

Appendix A8 shows the results for each beta in terms of confidence intervals. Each figure has four confidence intervals. All confidence intervals are built for the estimated beta from actual experience. While the first and third confidence intervals are constructed based on fixed  $\alpha=32\%$ , the second and fourth are based on the credibility alpha for each beta. Blue and green dots represents the expected assumptions and actual (fitted) betas. The first two confidence intervals

are prior to the adjustment and the last two are the ones obtained applying the proposed adjustment methodology.

Since the credibility alpha is bigger than 32%, as can be seen in the appendix that the credibility confidence intervals (the second and fourth) are narrower than the control confidence intervals (the first and third). Therefore if an adjustment is necessary (compare the third and fourth confidence intervals), the adjusted beta moves closer to the actual beta because small size confidence interval implies a bigger adjustment to the boundary.

#### 4.4.2 Credibility Effect on Unstable Dataset

Beta	$\beta_E$	$\beta_A$ (with $\alpha=32\%$ )	$\beta_A$ (with $\alpha_C$ )
$\beta_1$	-7.027	-7.133	-7.027
$\beta_2$ ( Male)	-0.315	-0.028	-0.265
$\beta_3$ (\$100 or more)	-0.032	-0.032	-0.032
$\beta_4$ ([65,75))	0.711	0.712	0.711
$\beta_4$ ([75,85))	-0.127	-0.127	-0.127
$\beta_4$ (85 and older)	-1.815	-1.815	-1.815
$\beta_5$ (6 months or more)	-0.794	-0.794	-0.794

Table 12: Adjustment results using fixed and credibility alphas for unstable dataset

Appendix A9 shows the complete version of Table 12. Standard error for each actual (shocked) beta,  $\sigma_{\hat{\beta}}$ , is bigger than the standard error for expected assumptions,  $\sigma_{\beta_E}$ . This makes the variance ratio of actual to expected,  $\sigma_{\hat{\beta}}/\sigma_{\beta_E}$ , more than one which produces credibility alpha,  $\alpha_C$ , less than the control alpha which is fixed at 32%. Credibility alpha less than 32% implies a wider confidence interval which means less credibility is put on the actual experience. More weight is put to the expected assumptions hence is more stable and the actual experience.

Appendix A10 shows the results for each beta in terms of confidence intervals. Each figure has four confidence intervals. All confidence intervals are built for the estimated beta from actual experience. While the first and third confidence intervals are constructed based on fixed  $\alpha=32\%$ , the second and fourth are based on the credibility alpha for each beta. Blue and green dots represents the expected assumptions and actual (fitted) betas. The first two confidence intervals are prior to the adjustment and the last two are the ones obtained by applying the proposed adjustment methodology.

Since credibility alpha is smaller than 32%, as can be seen in the appendix, the credibility confidence intervals (the second and fourth) are wider than the control confidence intervals (the first and third). Therefore if an adjustment is necessary (compare the third and fourth confidence intervals), the adjusted beta is further away from the actual beta because a large confidence interval implies a wider boundary from the mean.

#### **4.5 Effect of Historical Data Size on the Application of the Proposed Methodology**

It is common in statistical analyses that a bigger sample size produces better estimates. To see this particular effect in the application of the proposed methodology, 7 years of the dataset is generated. Incidence rates of females are shocked by taking the product of initial incidence rates at a constant value 10. This dataset is divided into two groups, historical data and actual experience. While the observations in the last year are all treated as actual experience, the historical data depends on the historical period. One year historical data is represented by the observations in the sixth year, two years data is represented by the observations of the fifth and sixth years and so on. This implies six years data is represented by the observations of the first, second, third, fourth, fifth and sixth years.



The type III error is used to measure this effect. As seen previously, two alphas are needed for the type III error. Since this section is more interested on the effect of historical data size, values of the two alphas become immaterial for this intent. Therefore both of the alphas are fixed at 10%.

Beta	Type III Error					
	1 year	2 years	3 years	4 years	5 years	6 year
$\beta_1$	0.58	0.40	0.26	0.19	0.19	0.16
$\beta_2$ ( Male)	0.43	0.14	0.09	0.04	0.04	0.04
$\beta_3$ (\$100 or more)	0.10	0.09	0.11	0.10	0.10	0.10
$\beta_4$ ([65,75))	0.07	0.07	0.07	0.07	0.07	0.07
$\beta_4$ ([75,85))	0.03	0.03	0.03	0.03	0.03	0.03
$\beta_4$ (85 and older)	0.09	0.09	0.09	0.09	0.09	0.09
$\beta_5$ (6 months or more)	0.12	0.11	0.10	0.10	0.10	0.11

Table 13: Historical data periods with their respective type III error

As seen in the table above, the affected betas ( $\beta_1$  and  $\beta_2$ ) most of the times have bigger errors compared to the unaffected betas whose errors are around 10%. As experience size increases, error reduces and converges to around 10%. This suggests that a company should wait several years before it can apply the proposed methodology to give a reliable adjustment.

## Chapter 5: Conclusion

Claims tracking and monitoring is divided into three stages. The first is actual to expected ratio plotting, the second is to identify blocks of business whose actual experience is significantly different from expected assumptions and the third is to adjust expected actuarial assumptions if it does not sufficiently reflect actual experience. This work focused on the last stage of the claims tracking and monitoring procedure.

A credibility approach has been used to update actuarial assumptions. This is adequate when actuarial assumptions are developed through a traditional approach by taking the ratio of the total number of events to total number of exposures. However, this is not applicable to actuarial assumptions which are built under a GLM framework.

GLM is widely used for modeling LTC insurance expected assumptions such as incidence, termination and utilization rates. This work proposes a statistically rigorous yet easy to implement and practical methodology to adjust expected assumptions under a GLM framework. A modification of the Society of Actuaries LTC insurance expected incidence rates was used to demonstrate the proposed methodology.

This methodology is parallel to the existing stochastic DAC technique which has been used in industry for unlocking DAC balances. The stochastic DAC technique exploits the straightforward, clear and powerful concept of confidence intervals to keep DAC balances at an

appropriate level and hence reduce the earnings volatility resulting from the DAC unlocking process.

To adopt this methodology, the expected assumptions are assumed to be derived through GLM and its general form as given by equation (2) in chapter 2 is known. Then GLM is fitted to actual experience and a confidence interval is constructed for each of the beta parameters. Similar to the DAC technique, this confidence interval is then compared to the initial expected assumptions beta to decide if an adjustment is needed. No adjustment is necessary if the initial expected assumptions beta falls within the confidence interval, otherwise an adjustment will be performed by shifting it to the confidence interval's closest boundary. The shifting to the boundary is to acknowledge the need to adjust the expected assumptions and also the fact that there exists uncertainty in the actual experience. Like in the credibility approach, this methodology provides consistency, as obtained from the expected assumptions, and reliability, as obtained from the actual experience. This blend stabilizes the difference between expected assumptions and actual experience and therefore reduces volatility.

This methodology is statistically stable since it produces type I, II and III errors as expected. Type I error which measures the probability of rejecting beta from the expected assumptions when actual experience equals the expected assumptions is close to the significance level  $\alpha$ . When there is significant difference, type II error which measures the probability of accepting beta from the expected assumptions when actual experience does not equal the expected assumptions is close to zero. When the methodology is employed, the type III error which measures the probability of rejecting beta from the adjusted expected assumptions when actual experience does not equal the expected assumptions gets closer to  $\alpha$  compared to the error

generated when initial expected assumptions are not adjusted. Since the methodology forces expected assumptions beta which is significantly different from the actual experience to be the actual beta's confidence interval's closest boundary the adjustment brings the type III error gradually closer to  $\alpha$  but definitely smaller compared to not making any adjustment.

It is obvious that the width of the confidence interval depends on  $\alpha$ . This brings the matter of determining the appropriate value of  $\alpha$ . Therefore credibility effect is brought into the proposed methodology and hence a significant level denoted by credibility alpha,  $\alpha_c$ , is defined. Credibility alpha given by equation (16) in chapter 3 is a function of standard errors of the expected assumptions beta and the actual beta. This extension needs this extra piece of information to be known and not just the general form of the model as given by equation (2) in chapter 2. The bigger the variation in actual experience compared to expected assumptions, the larger the variation ratio in equation (16) and hence the smaller the credibility alpha is. This implies a wider confidence interval which means it is more likely to accept beta from the expected assumptions. In other words, actual experience is less credible if it has higher variation than expected assumptions. This extension sets a way to decide an appropriate significant level prior to using the proposed methodology.

Another concern is how much actual experience is needed so that application of the proposed methodology will give a reliable adjustment effect. It is shown in this work that the bigger the size of actual experience, the smaller the error. This implies that one has to have large enough actual experience to rely on the adjustment outcome. However, the optimal size of actual experience before an adjustment is made is beyond the scope of this work.

This work provides a platform for more future research. One possibility is to bring more complexity in investigating type I, II and III errors. Here, the simulation procedure does not take into account termination and recovery effects. More assumptions can be brought into the simulation procedure so that it better reflects what really is happening in the real world. Complexity is not entertained here since this work is at the very fundamental level of developing the proposed methodology.

Another possible future work is to experiment with the credibility alpha. Here each beta has its own standard error. One may be interested to investigate if there is only one credibility alpha for all of the betas. Perhaps choosing the maximum or minimum credibility alpha depending on the level of trust one has on the actual experience would be a possible approach.

As mentioned above, the optimal size of actual experience prior to applying the proposed method is beyond the scope of this work, hence this may also be potential future research.

Also, size of actual experience can also be incorporated into equation (16) to compute credibility alpha. This is of the utmost importance if the actual experience, or sample, is actually part of the expected assumptions, or populations. The inclusion of actual size as a factor for obtaining credibility alpha is very practical since similar to what has been done by SOA<sup>3</sup>, a set of expected assumptions was established through an aggregate dataset which comprised of several participating companies. This established expected assumptions is then to be used by individual participating companies. Hence it is of interest to each of the companies to adjust the expected assumptions to accommodate their own experience. More credibility can be put to individual companies with bigger size of data, and vice versa.

LTC incidence rates were used to demonstrate the proposed methodology. But, our method is not only limited to LTC incidence rates and is widely applicable to any problem of changing a set of initial values where the initial values were constructed under a GLM framework.

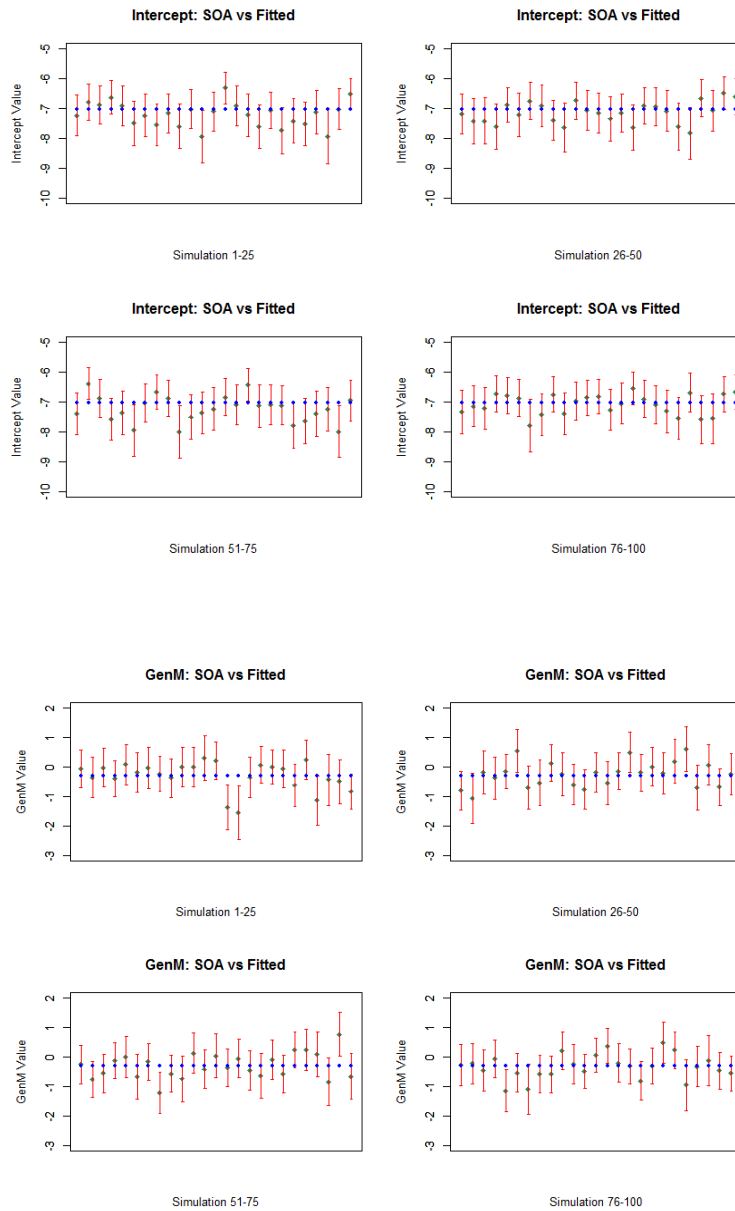
Similar to how the stochastic DAC unlocking technique is an acceptable procedure by regulators and practitioners, this proposed methodology is foreseen to be accepted as a basis of Principle-Based Reserving in the future.

## References

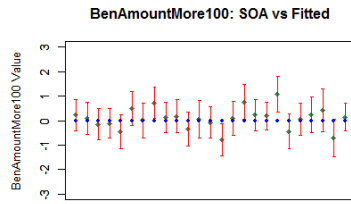
1. Anderson R. & Rosenberg. “Age Standardization of Death Rates: Implementation of the Year 2000 Standard.” *National Vital Statistics Reports* 47.3 (1998): 1-12
2. Benjamin, G. “Selecting Mortality Tables: A Credibility Approach (2008)
3. Donald F. B. “Statistical Credibility Theory.” *Southeastern Actuarial Conference* (2009)
4. Longley-Cook A., Shaw D., Sherril M. & Vadiveloo J. “Stochastic DAC Unlocking for Variable Annuity Products.” *SOA The Financial Reporter* 45 (2001): 1-6
5. Long Term Care Intercompany Experience Study – Aggregate Database 2000-2011 Report, Society of Actuaries
6. Vadiveloo J., Niu G., Xu J., Shen X. & Song T. “Tracking and Monitoring Claims Experience: A Practical Application of Risk Management.” *Risk Management* 31 (2014): 12-15
7. <http://longtermcare.gov/>

# Appendix

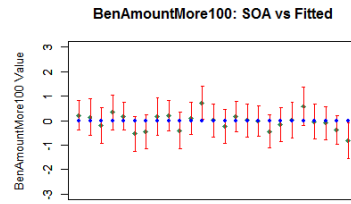
## A1: Simulations of Type I Error ( $\alpha=10\%$ )



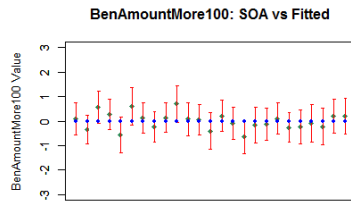




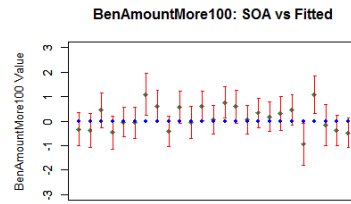
Simulation 1-25



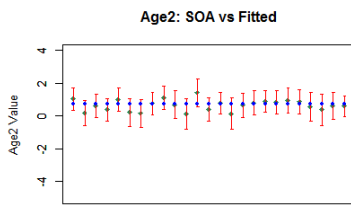
Simulation 26-50



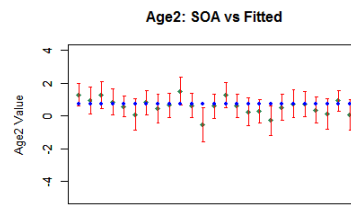
Simulation 51-75



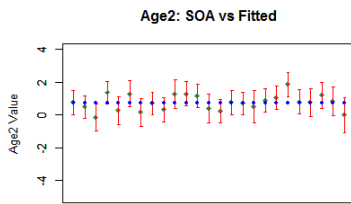
Simulation 76-100



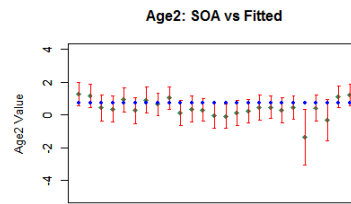
Simulation 1-25



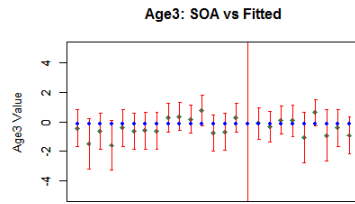
Simulation 26-50



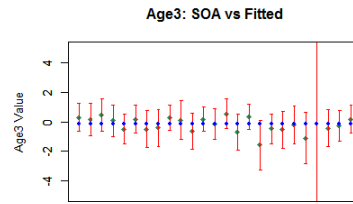
Simulation 51-75



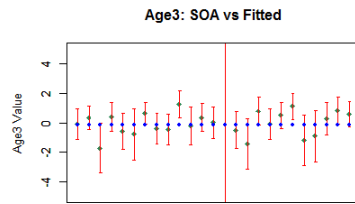
Simulation 76-100



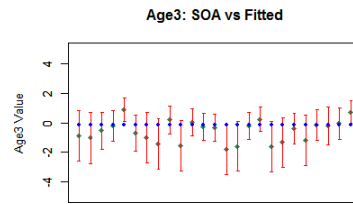
Simulation 1-25



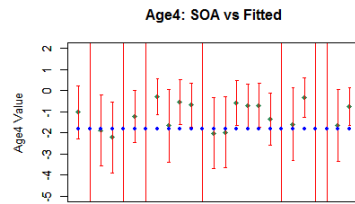
Simulation 26-50



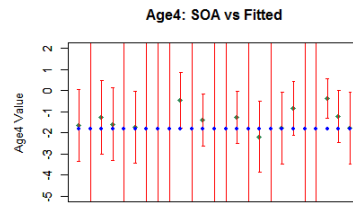
Simulation 51-75



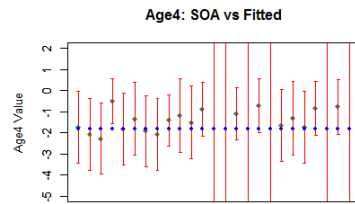
Simulation 76-100



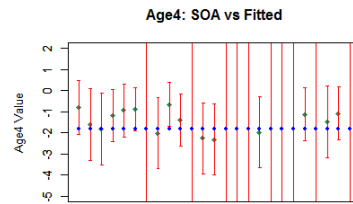
Simulation 1-25



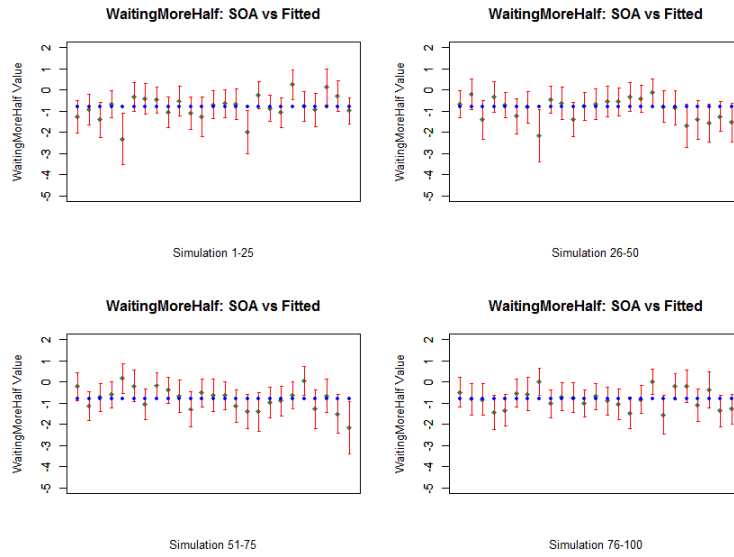
Simulation 26-50



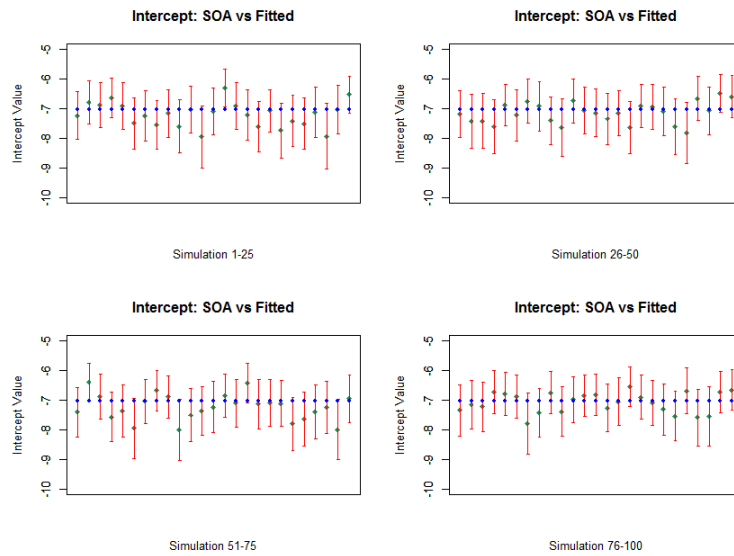
Simulation 51-75

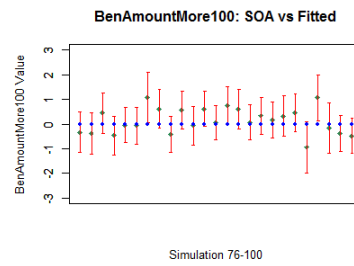
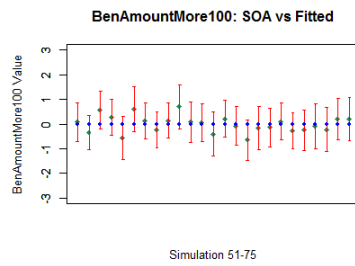
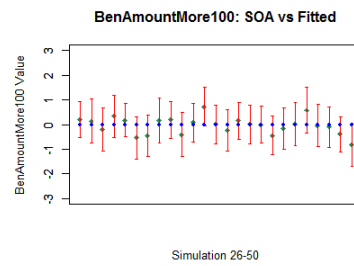
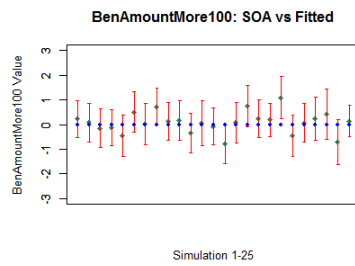
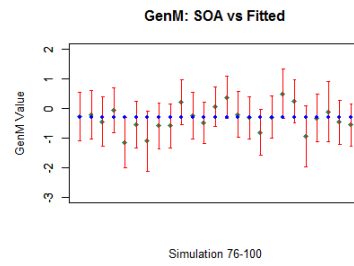
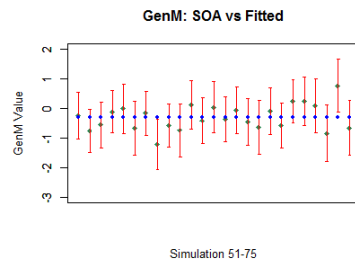
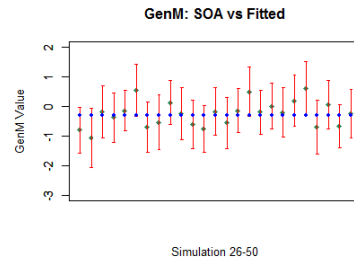
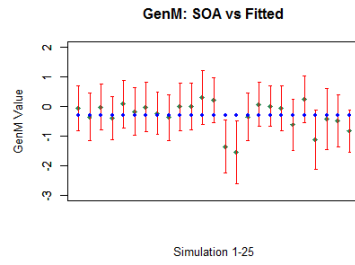


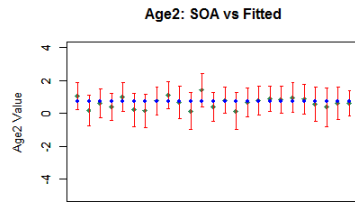
Simulation 76-100



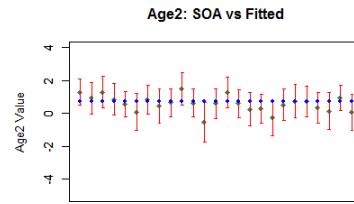
## A2: Simulations of Type I Error ( $\alpha = 5\%$ )



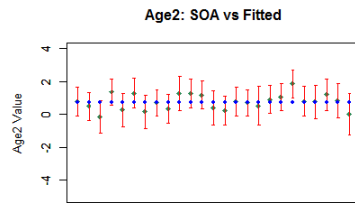




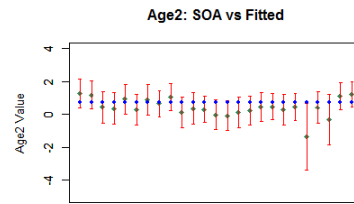
Simulation 1-25



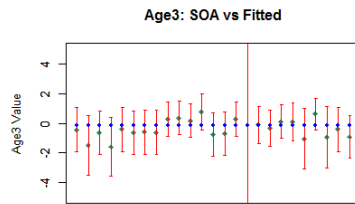
Simulation 26-50



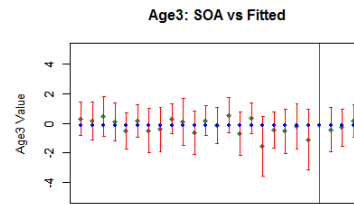
Simulation 51-75



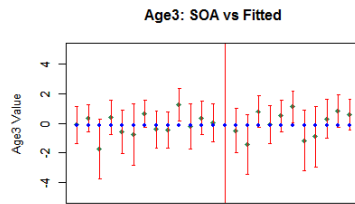
Simulation 76-100



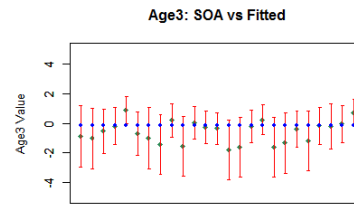
Simulation 1-25



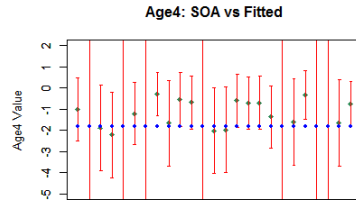
Simulation 26-50



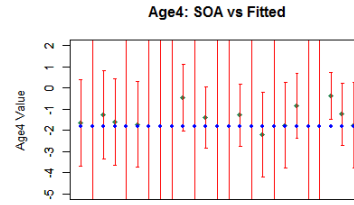
Simulation 51-75



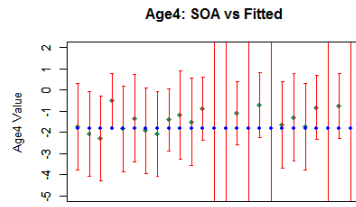
Simulation 76-100



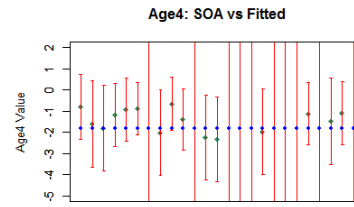
Simulation 1-25



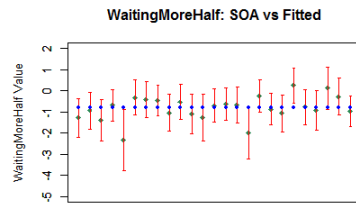
Simulation 26-50



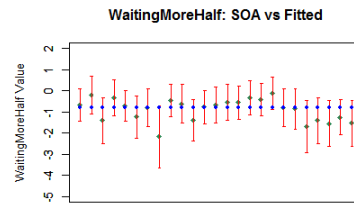
Simulation 51-75



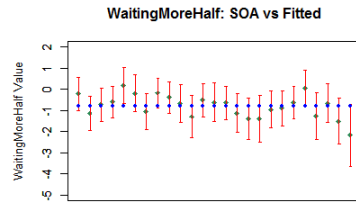
Simulation 76-100



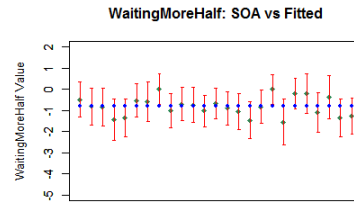
Simulation 1-25



Simulation 26-50

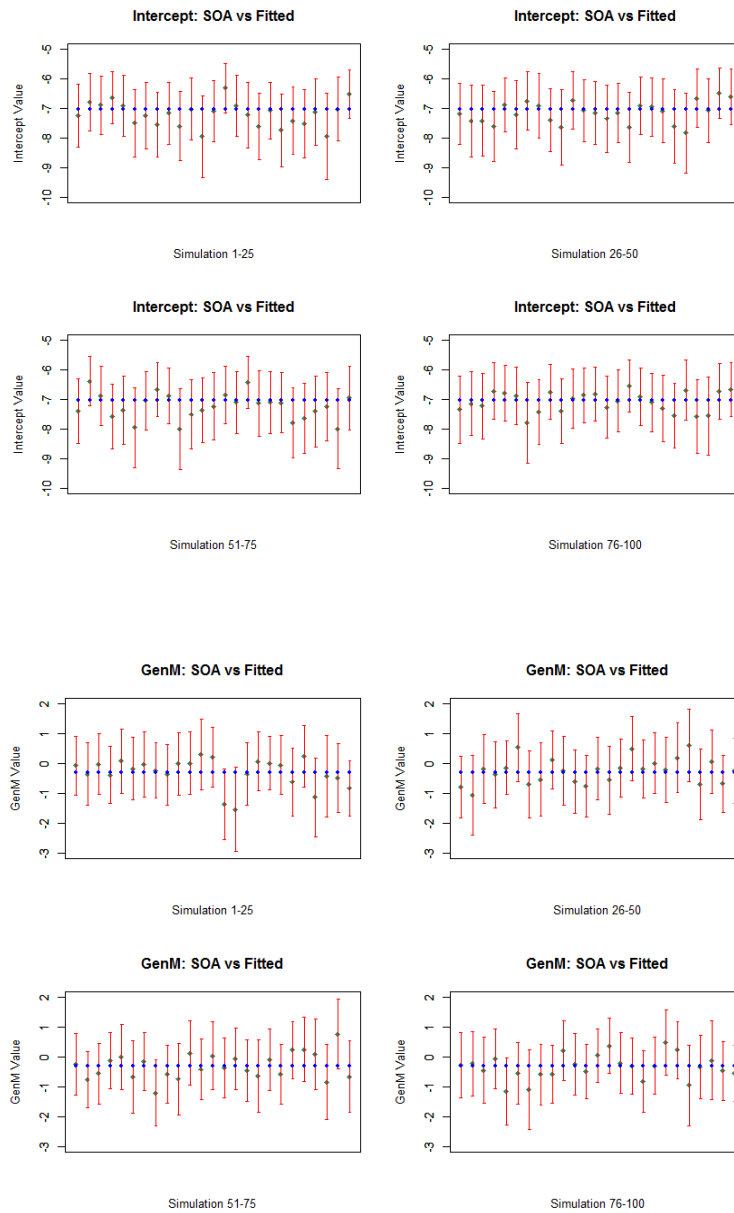


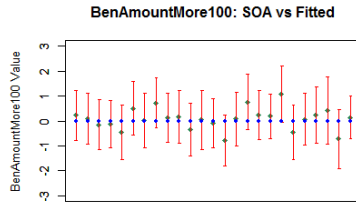
Simulation 51-75



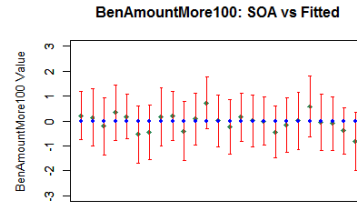
Simulation 76-100

### A3: Simulations of Type I Error ( $\alpha=1\%$ )

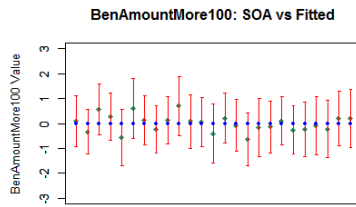




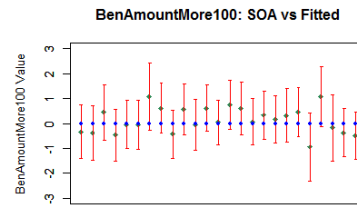
Simulation 1-25



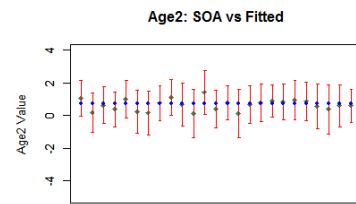
Simulation 26-50



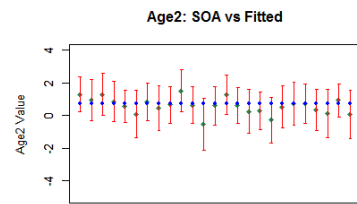
Simulation 51-75



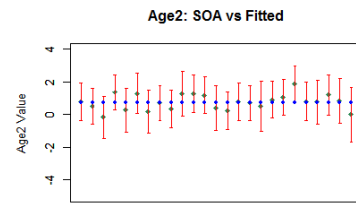
Simulation 76-100



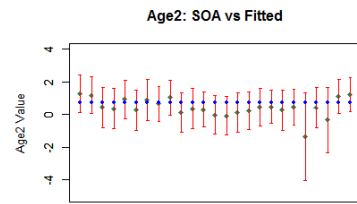
Simulation 1-25



Simulation 26-50

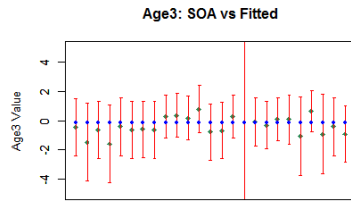


Simulation 51-75

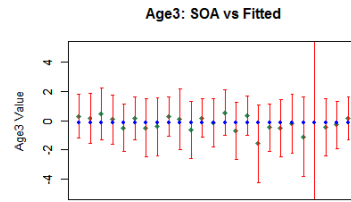


Simulation 76-100

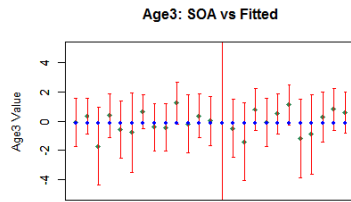




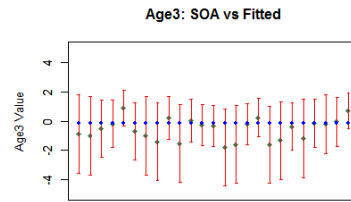
Simulation 1-25



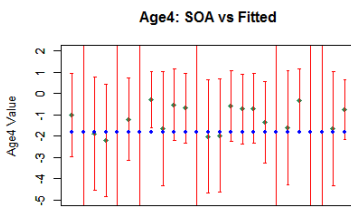
Simulation 26-50



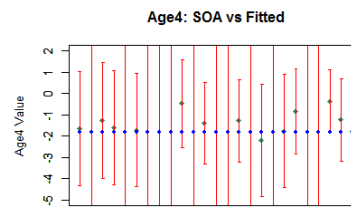
Simulation 51-75



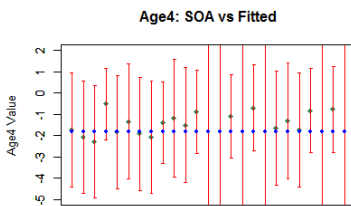
Simulation 76-100



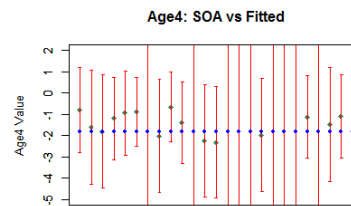
Simulation 1-25



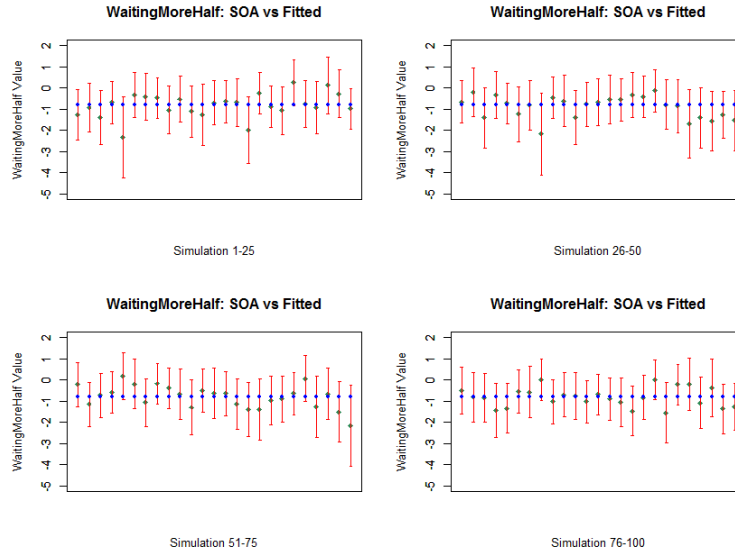
Simulation 26-50



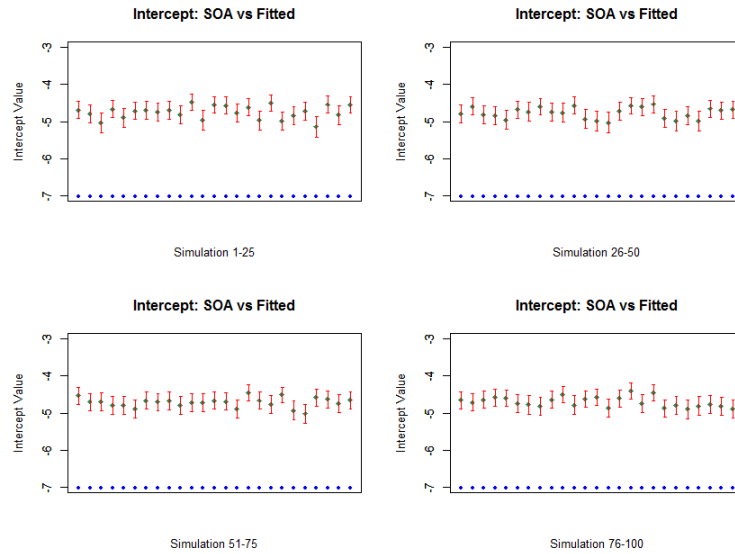
Simulation 51-75

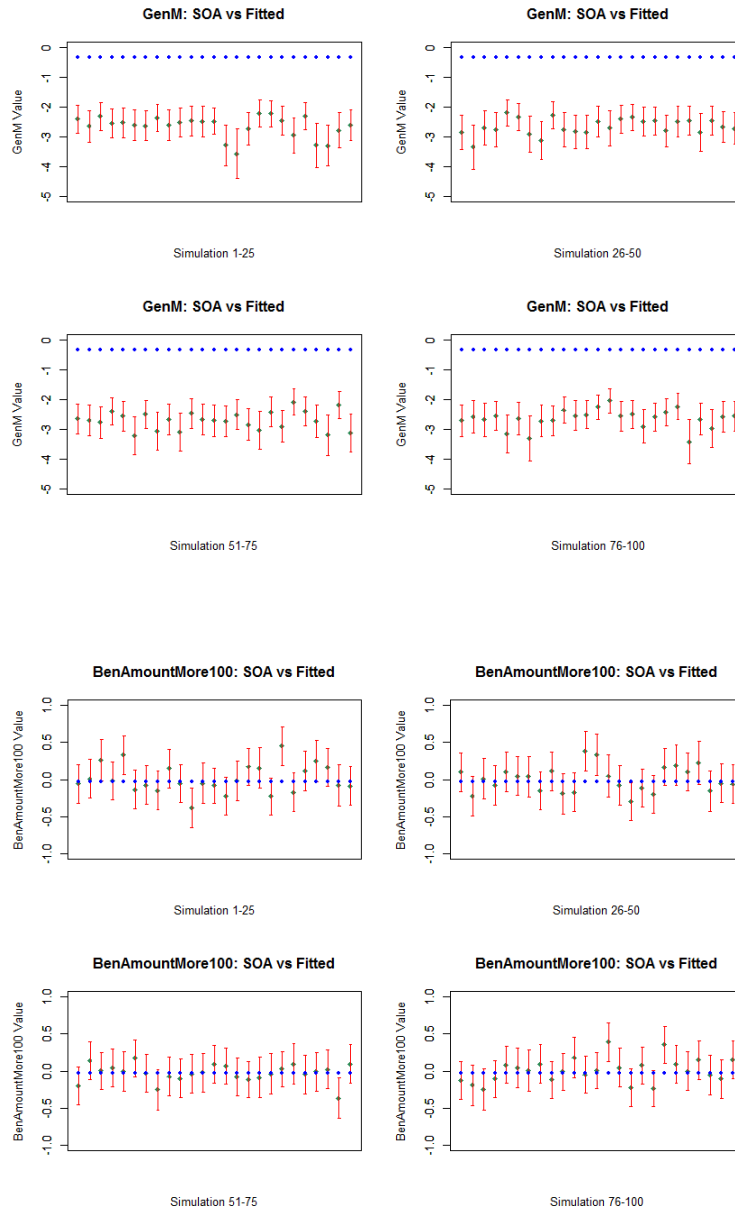


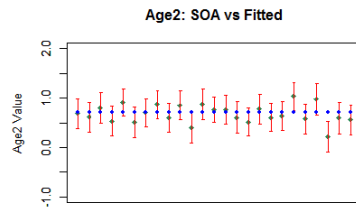
Simulation 76-100



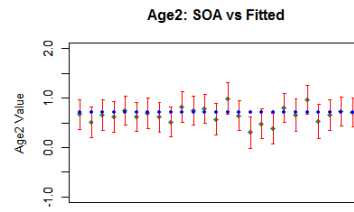
#### A4: Simulations of Type II Error ( $\alpha = 10\%$ )



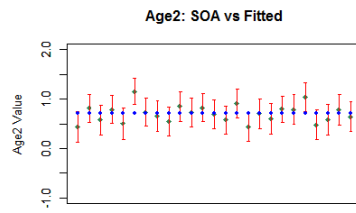




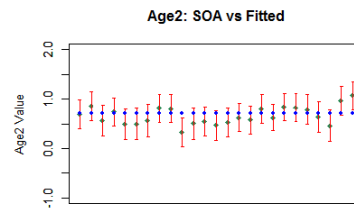
Simulation 1-25



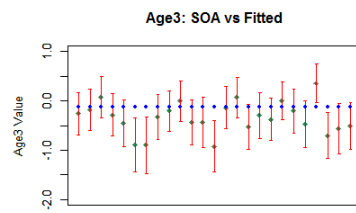
Simulation 26-50



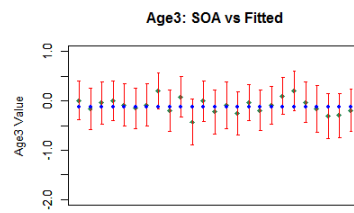
Simulation 51-75



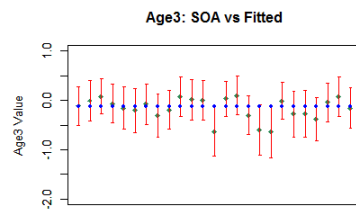
Simulation 76-100



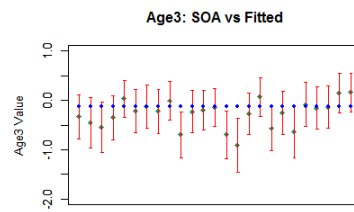
Simulation 1-25



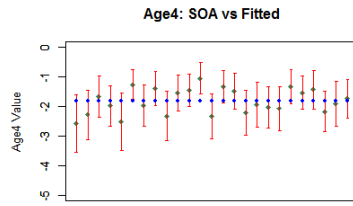
Simulation 26-50



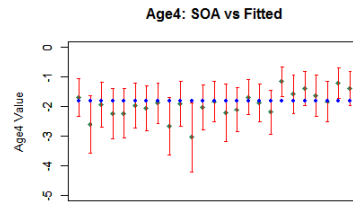
Simulation 51-75



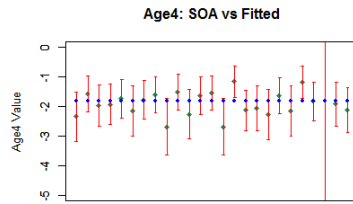
Simulation 76-100



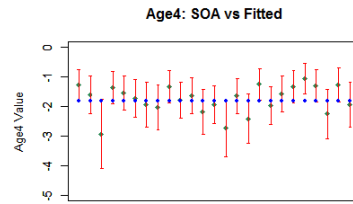
Simulation 1-25



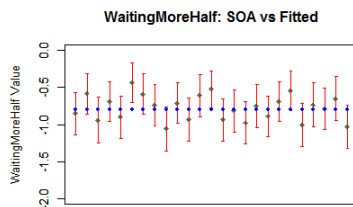
Simulation 26-50



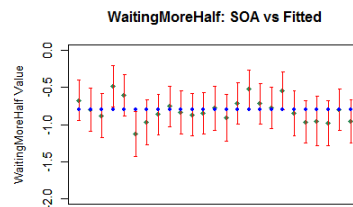
Simulation 51-75



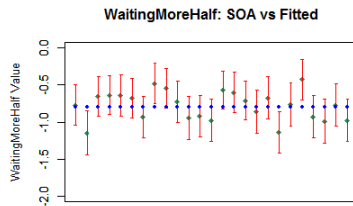
Simulation 76-100



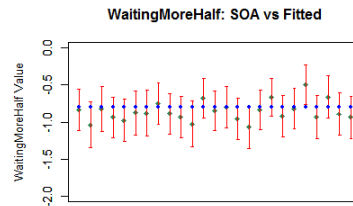
Simulation 1-25



Simulation 26-50

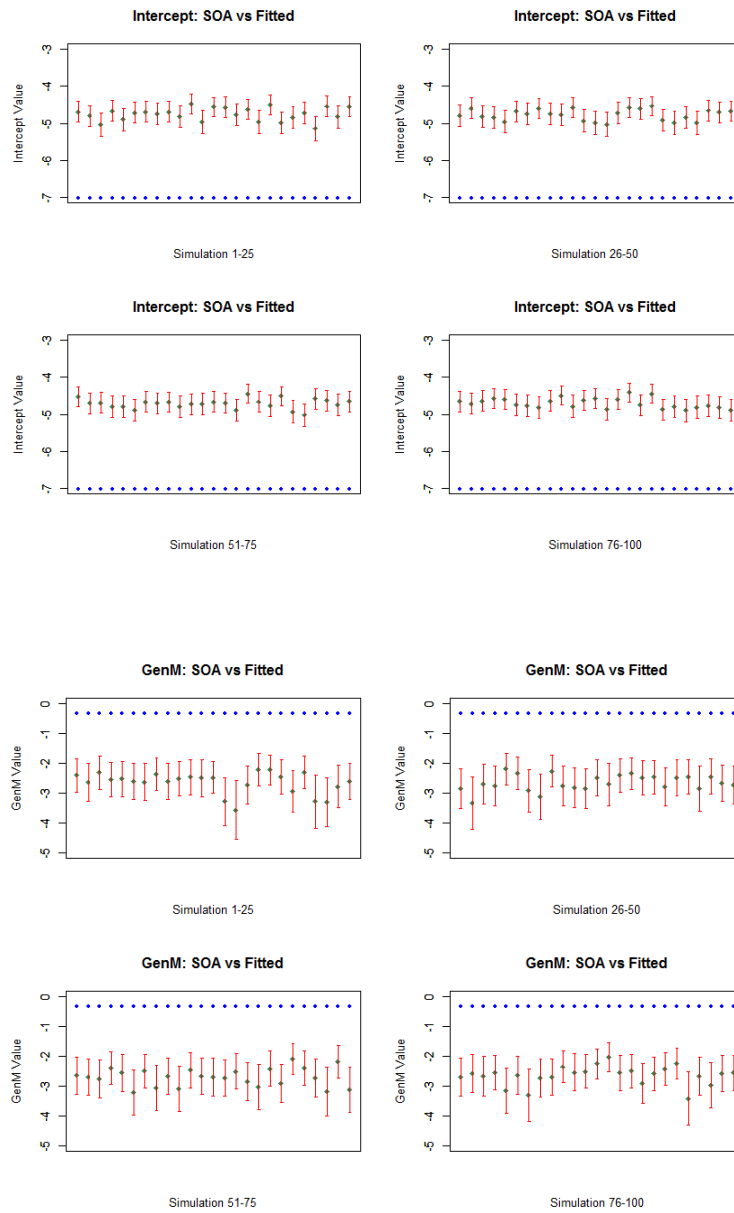


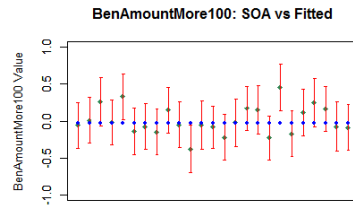
Simulation 51-75



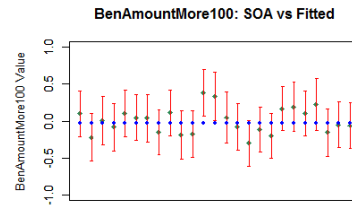
Simulation 76-100

## A5: Simulations of Type II Error ( $\alpha=5\%$ )

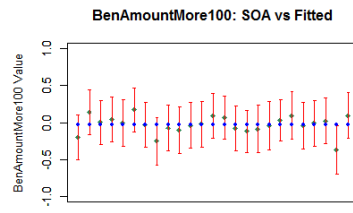




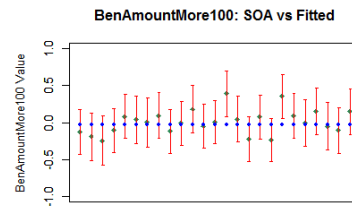
Simulation 1-25



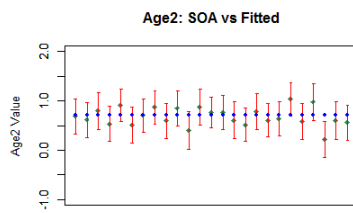
Simulation 26-50



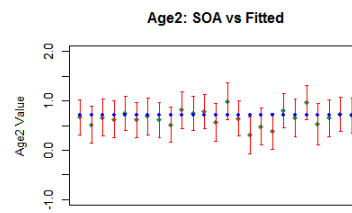
Simulation 51-75



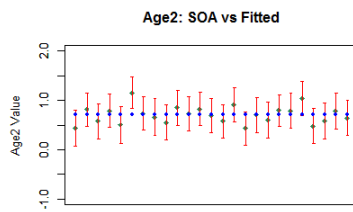
Simulation 76-100



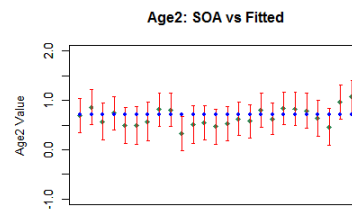
Simulation 1-25



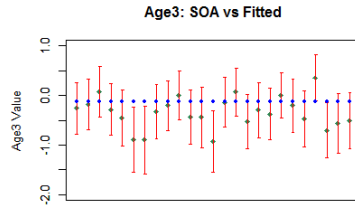
Simulation 26-50



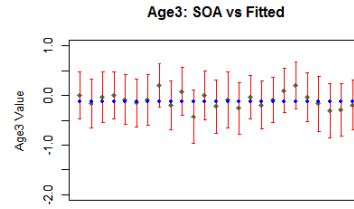
Simulation 51-75



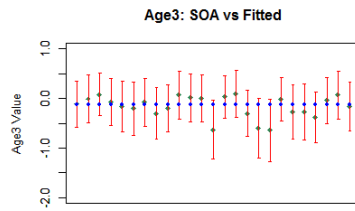
Simulation 76-100



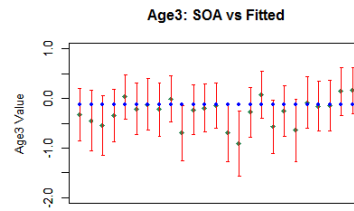
Simulation 1-25



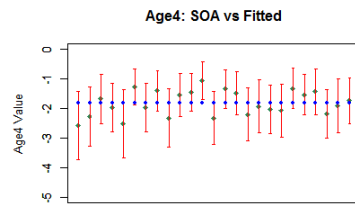
Simulation 26-50



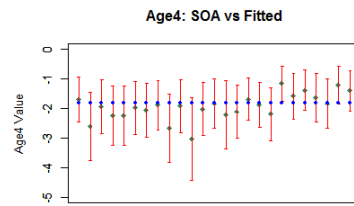
Simulation 51-75



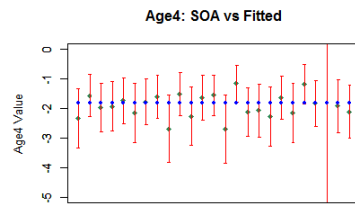
Simulation 76-100



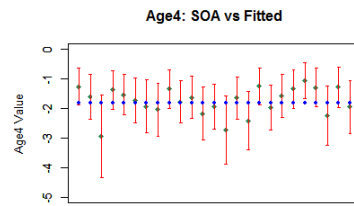
Simulation 1-25



Simulation 26-50

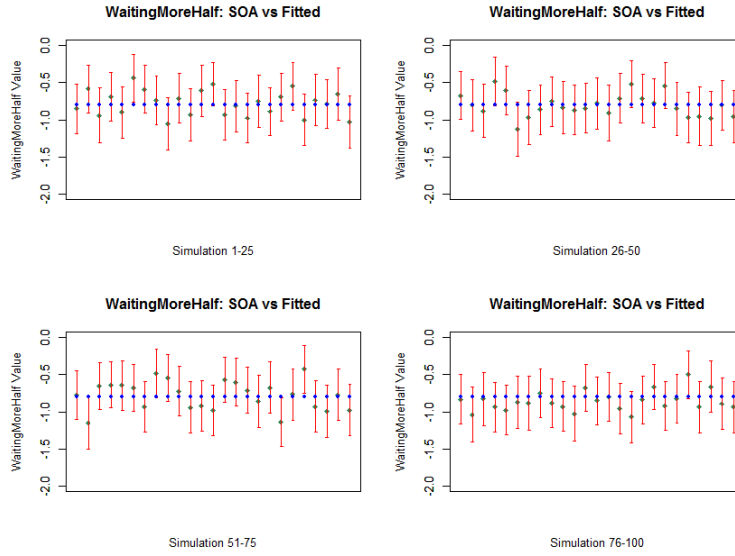


Simulation 51-75

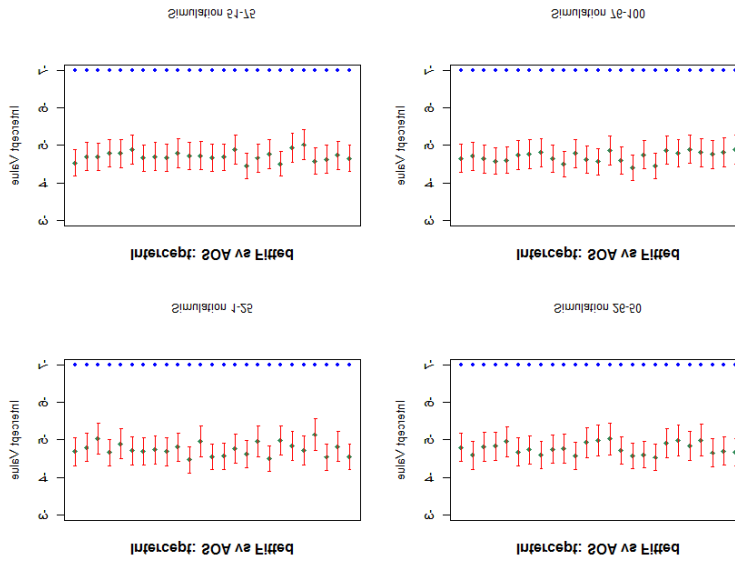


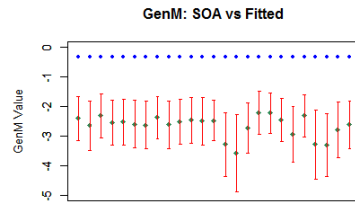
Simulation 76-100



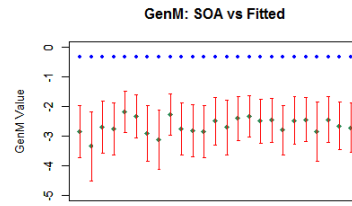


## A6: Simulations of Type II Error ( $\alpha=1\%$ )

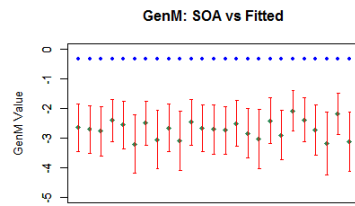




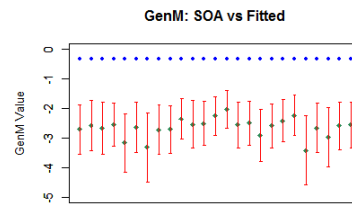
Simulation 1-25



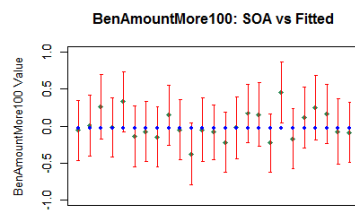
Simulation 26-50



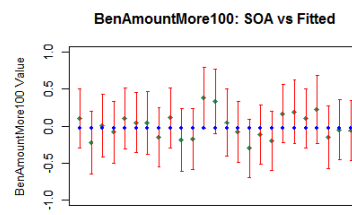
Simulation 51-75



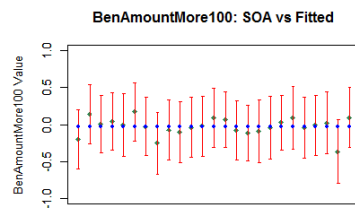
Simulation 76-100



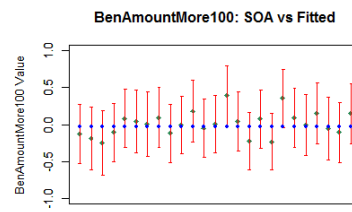
Simulation 1-25



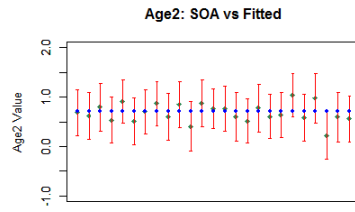
Simulation 26-50



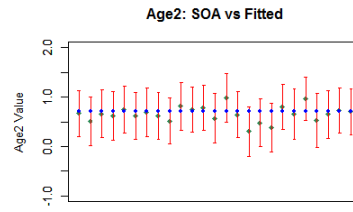
Simulation 51-75



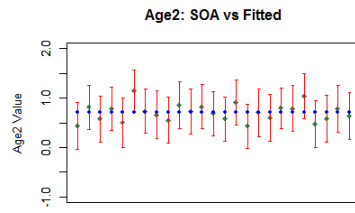
Simulation 76-100



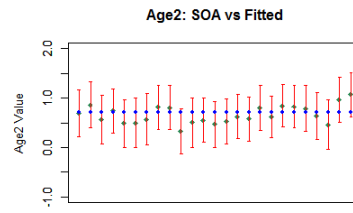
Simulation 1-25



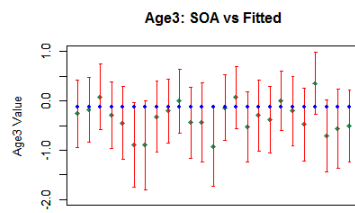
Simulation 26-50



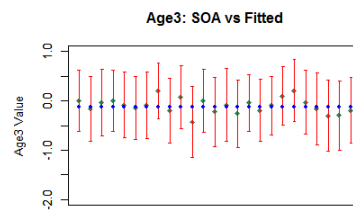
Simulation 51-75



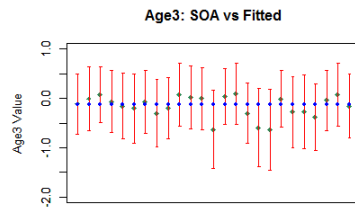
Simulation 76-100



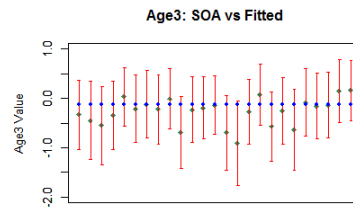
Simulation 1-25



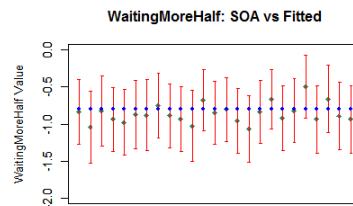
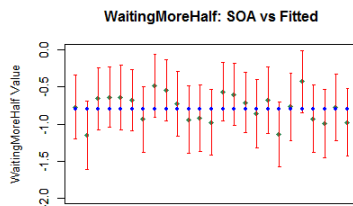
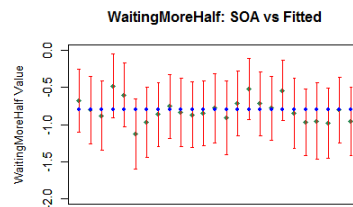
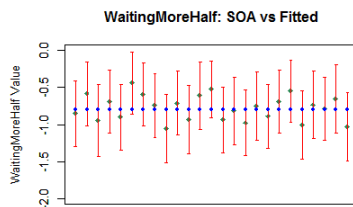
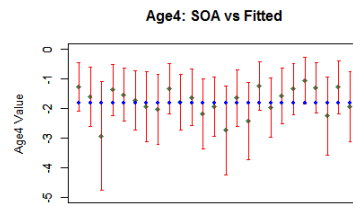
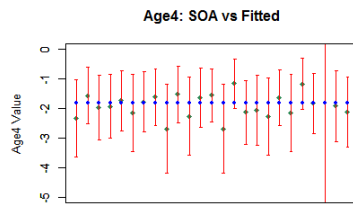
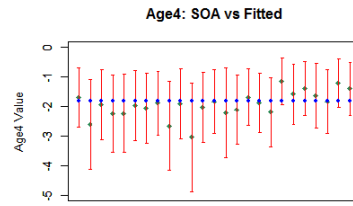
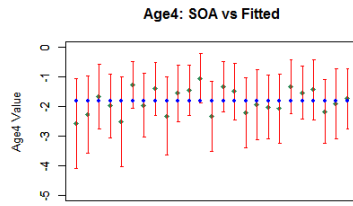
Simulation 26-50



Simulation 51-75



Simulation 76-100

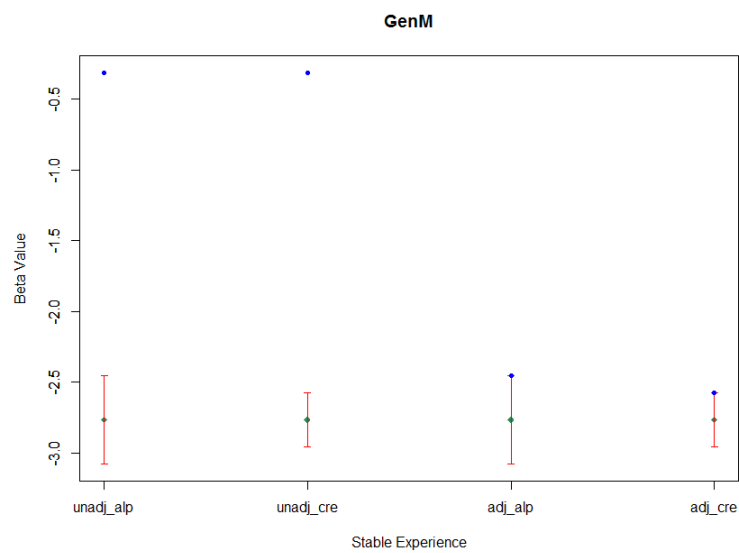
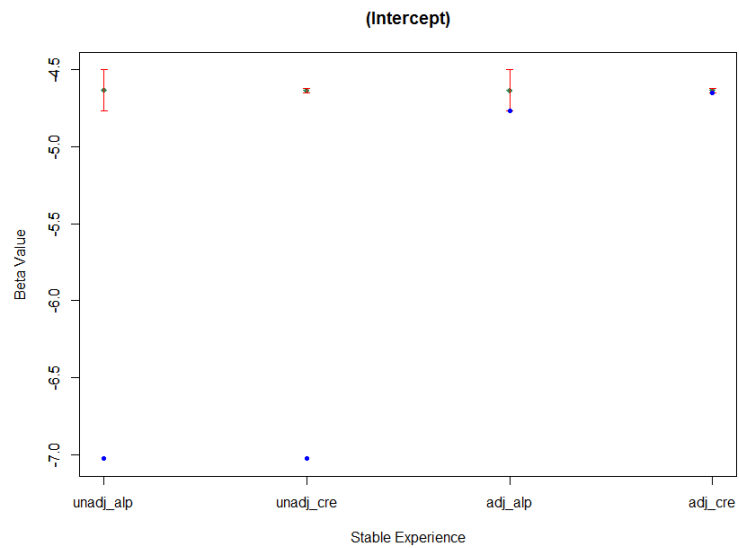


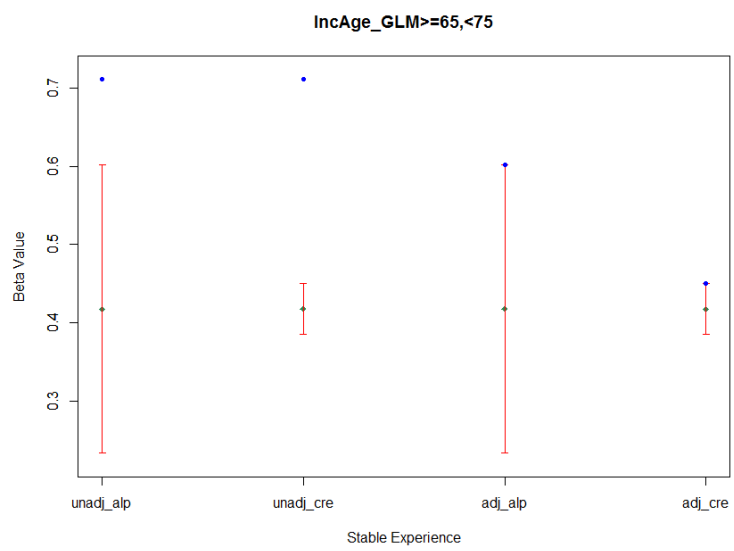
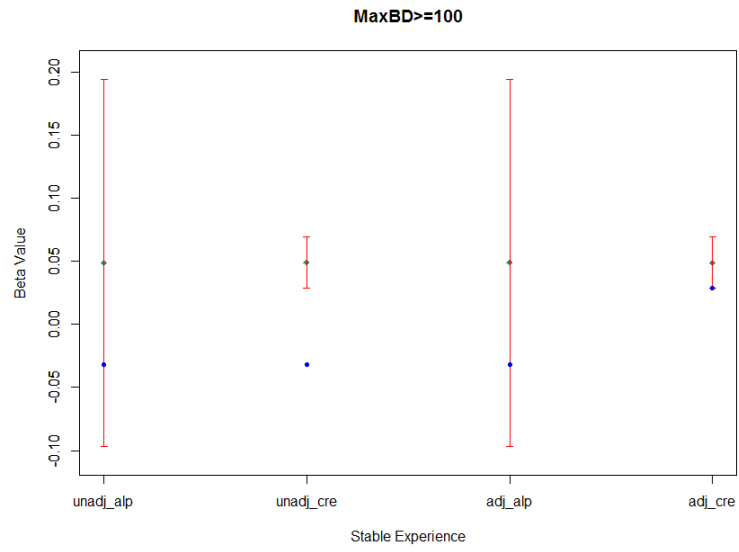
**A7: Complete Adjustment Results Using Fixed and Credibility Alphas for Stable Dataset**

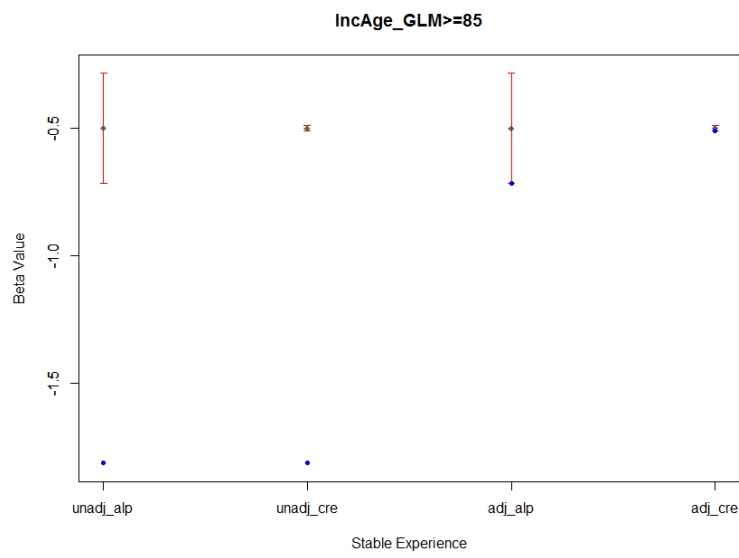
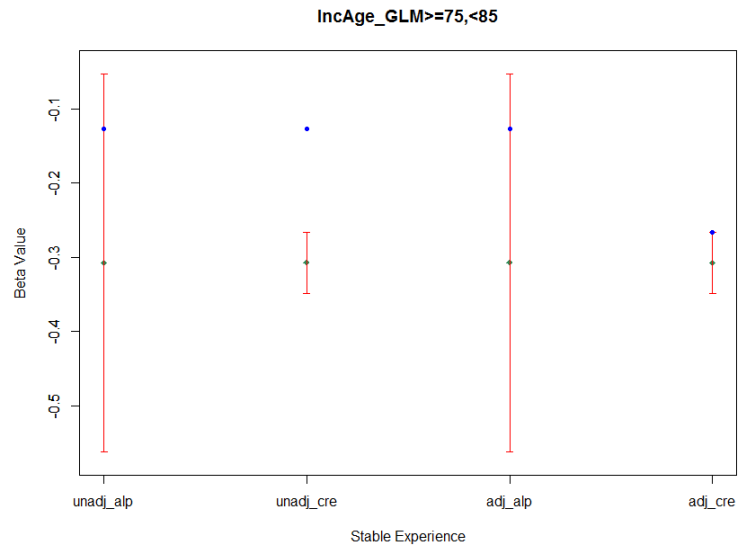
Beta	$\beta_E$	$\sigma_{\beta_E}$	$\alpha$	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	$\sigma_{\hat{\beta}}^2 / \sigma_{\beta_E}^2$
$\beta_1$	-7.027	0.391	0.32	-4.635	0.135	0.119
$\beta_2$ ( Male)	-0.315	0.397	0.32	-2.765	0.311	0.612
$\beta_3$ (\$100 or more)	-0.032	0.392	0.32	0.049	0.146	0.139
$\beta_4$ ([65, 75))	0.711	0.443	0.32	0.418	0.185	0.174
$\beta_4$ ([75, 85))	-0.127	0.636	0.32	-0.307	0.256	0.162
$\beta_4$ (85 and older)	-1.815	1.035	0.32	-0.500	0.217	0.044
$\beta_5$ (6 months or more)	-0.794	0.425	0.32	-0.739	0.157	0.136

Beta	$\alpha_C$	$\beta_A^\alpha$	$\beta_A^{\alpha_C}$	$ \beta_E - \beta_A^\alpha $	$ \beta_E - \beta_A^{\alpha_C} $
$\beta_1$	0.91	-7.133	-7.027	2.257	2.375
$\beta_2$ ( Male)	0.54	-0.028	-0.265	2.141	2.260
$\beta_3$ (\$100 or more)	0.89	-0.032	-0.032	0	0.060
$\beta_4$ ([65, 75))	0.86	0.712	0.711	0.109	0.261
$\beta_4$ ([75, 85))	0.87	-0.127	-0.127	0	0.139
$\beta_4$ (85 and older)	0.97	-1.815	-1.815	1.099	1.305
$\beta_5$ (6 months or more)	0.89	-0.794	-0.794	0	0.033

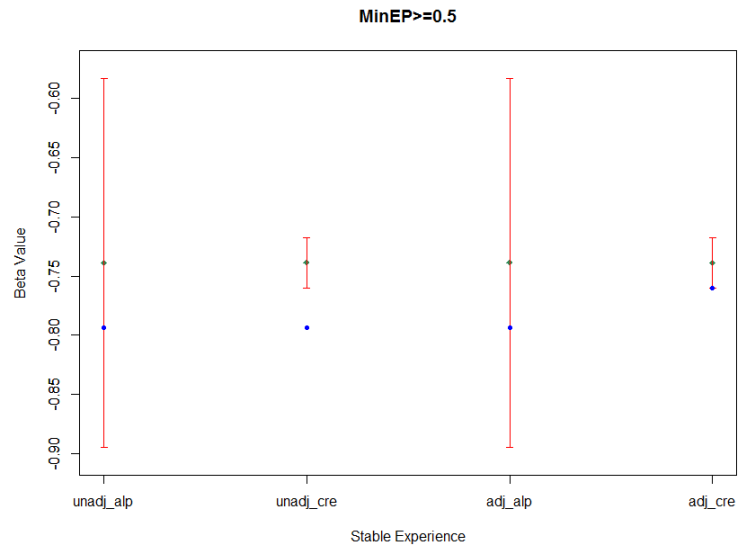
## A8: Adjustment Results Using Fixed and Credibility Alphas for Stable Dataset (Confidence Interval)









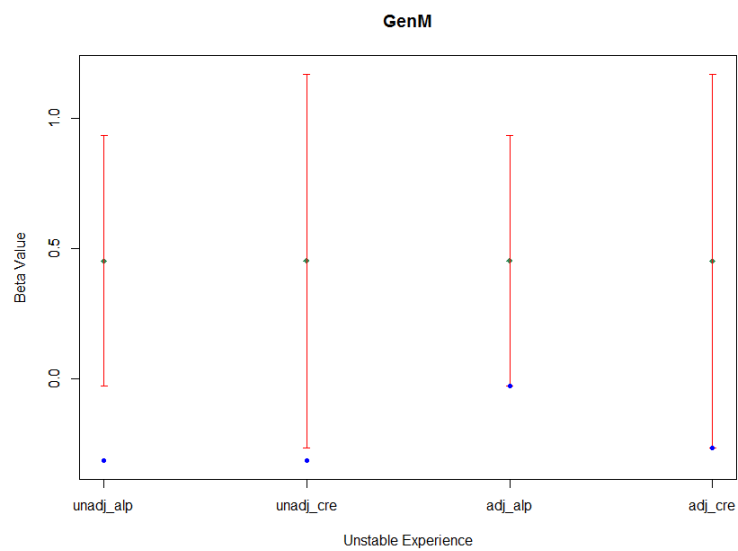
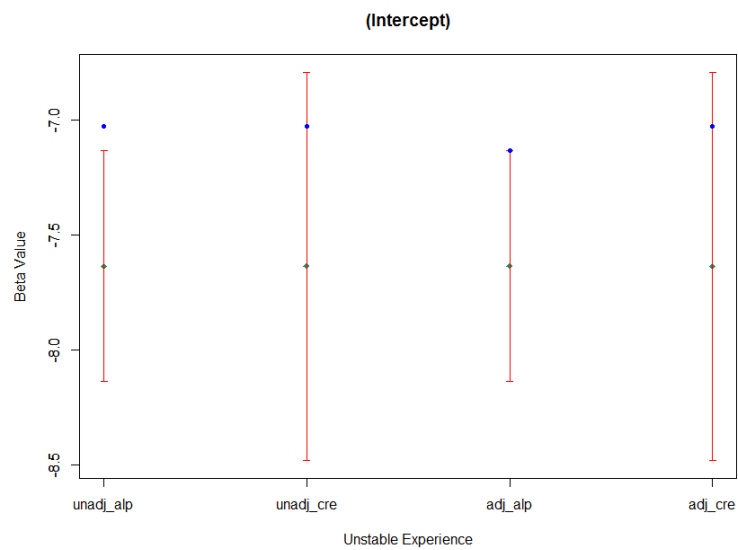


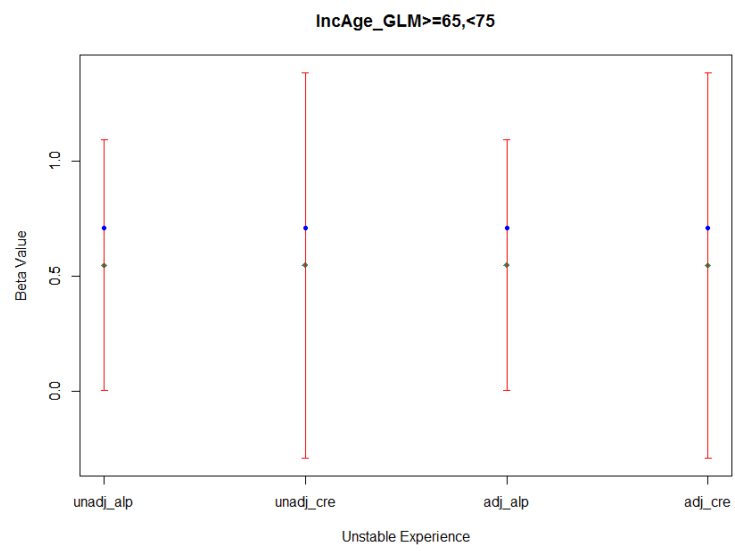
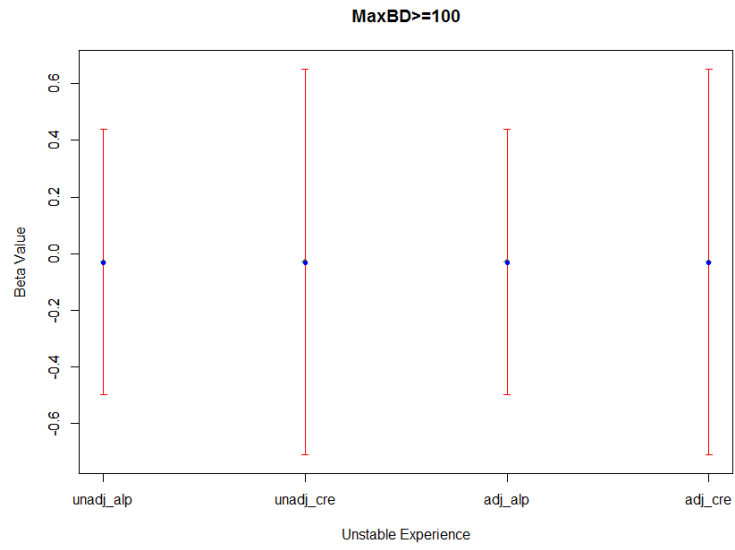
**A9: Complete Adjustment Results Using Fixed and Credibility Alphas for Unstable Dataset**

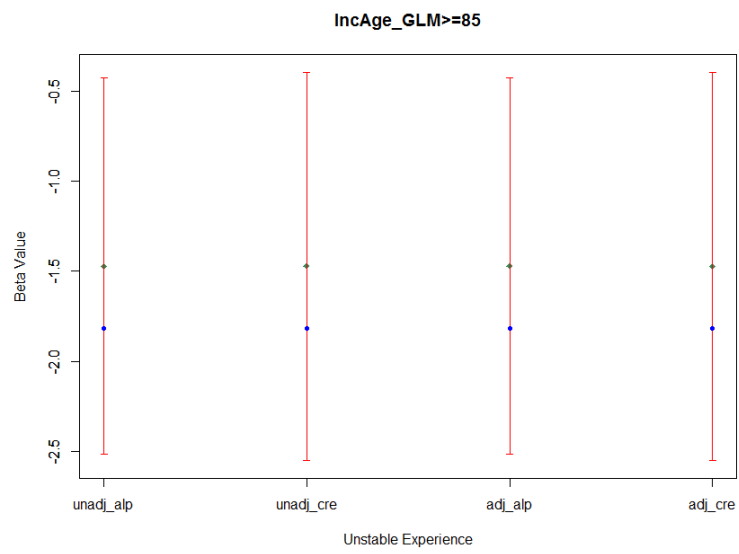
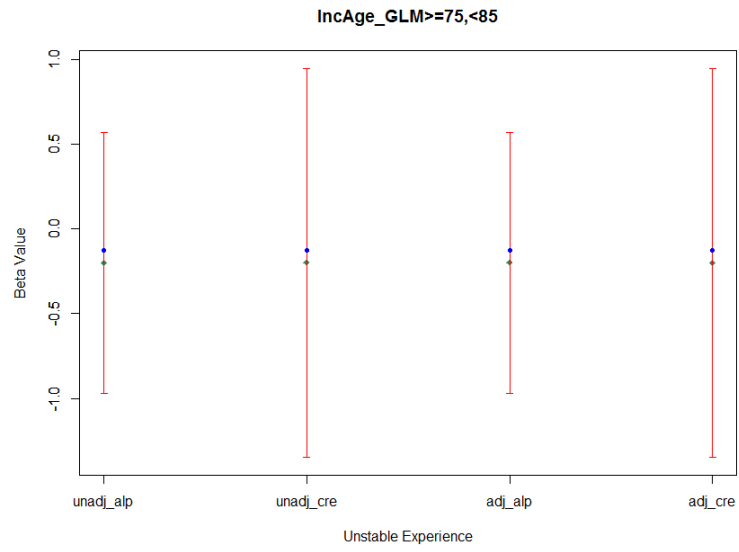
Beta	$\beta_E$	$\sigma_{\beta_E}$	$\alpha$	$\hat{\beta}$	$\sigma_{\hat{\beta}}$	$\sigma_{\hat{\beta}}^2 / \sigma_{\beta_E}^2$
$\beta_1$	-7.027	0.391	0.32	-7.636	0.505	1.669
$\beta_2$ ( Male)	-0.315	0.397	0.32	0.453	0.483	1.484
$\beta_3$ (\$100 or more)	-0.032	0.392	0.32	-0.029	0.471	1.444
$\beta_4$ ([65, 75))	0.711	0.443	0.32	0.548	0.548	1.527
$\beta_4$ ([75, 85))	-0.127	0.636	0.32	-0.200	0.775	1.482
$\beta_4$ (85 and older)	-1.815	1.035	0.32	-1.472	1.049	1.027
$\beta_5$ (6 months or more)	-0.794	0.425	0.32	-1.239	0.567	1.780

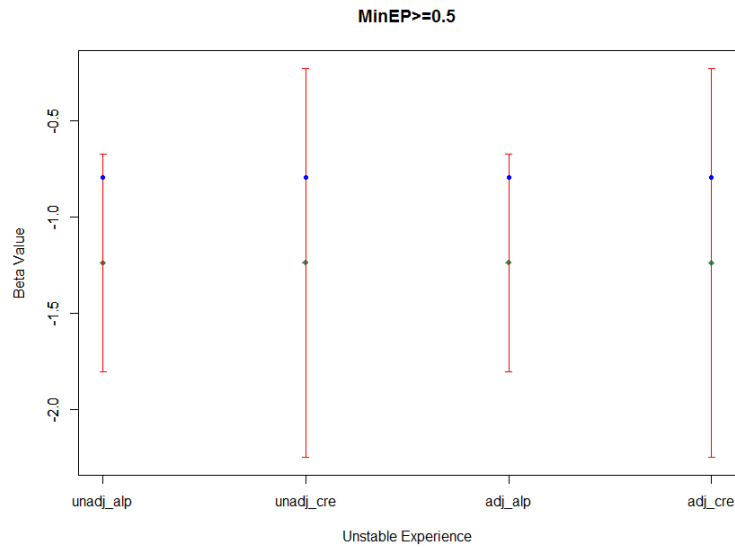
Beta	$\alpha_C$	$\beta_A^\alpha$	$\beta_A^{\alpha_C}$	$ \beta_E - \beta_A^\alpha $	$ \beta_E - \beta_A^{\alpha_C} $
$\beta_1$	0.10	-7.133	-7.027	0.107	0
$\beta_2$ ( Male)	0.14	-0.028	-0.265	0.287	0.050
$\beta_3$ (\$100 or more)	0.15	-0.032	-0.032	0	0
$\beta_4$ ([65, 75))	0.13	0.711	0.711	0	0
$\beta_4$ ([75, 85))	0.14	-0.127	-0.127	0	0
$\beta_4$ (85 and older)	0.30	-1.815	-1.815	0	0
$\beta_5$ (6 months or more)	0.08	-0.794	-0.0794	0	0

# A10: Adjustment Results Using Fixed and Credibility Alphas for Unstable Dataset (Confidence Interval)









## A11: R Code for Type I Error

```
factor1 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor1_Base.csv",
header=TRUE)

factor2 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor2_Gen.csv",
header=TRUE)

factor3 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor3_MaxBD.csv",
header=TRUE)

factor4 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor4_IncAge.csv",
header=TRUE)

factor5 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor5_MinEP.csv",
header=TRUE)

initial_parameter <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Parameter.csv",
header=TRUE)

n <- 100

intercept <- matrix(1, n, 6)

colnames(intercept) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")

intercept <- data.frame(intercept)

GenF <- matrix(1, n, 6)
```

```
colnames(GenF) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
GenF <- data.frame(GenF)
```

```
GenM <- matrix(1, n, 6)
```

```
colnames(GenM) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
GenM <- data.frame(GenM)
```

```
BenAmountLess100 <- matrix(1, n, 6)
```

```
colnames(BenAmountLess100) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
BenAmountLess100 <- data.frame(BenAmountLess100)
```

```
BenAmountMore100 <- matrix(1, n, 6)
```

```
colnames(BenAmountMore100) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
BenAmountMore100 <- data.frame(BenAmountMore100)
```

```
Age1 <- matrix(1, n, 6)
```

```
colnames(Age1) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
Age1 <- data.frame(Age1)
```

```
Age2 <- matrix(1, n, 6)
```

```
colnames(Age2) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
Age2 <- data.frame(Age2)
```

```
Age3 <- matrix(1, n, 6)
```

```
colnames(Age3) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
Age3 <- data.frame(Age3)
```

```
Age4 <- matrix(1, n, 6)
```

```
colnames(Age4) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```

Age4 <- data.frame(Age4)

WaitingHalf <- matrix(1, n, 6)

colnames(WaitingHalf) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")

WaitingHalf <- data.frame(WaitingHalf)

WaitingMoreHalf <- matrix(1, n, 6)

colnames(WaitingMoreHalf) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")

WaitingMoreHalf <- data.frame(WaitingMoreHalf)

for(i in 1:n){

  ## Generate 10000 data points for 5 years with 4 variables

  data1 <- matrix(1, 10000, 4, byrow=TRUE)

  #rownames(data1) <- 1:10000

  colnames(data1) <- c("Gen", "MaxBD", "IncAge", "MinEP")

  #str(data1)

  data1 <- data.frame(data1)

  #str(data1)

  set.seed(200+i)

  possible_Gen <- c("F", "M")

  Gen_vector <- sample(possible_Gen, 10000, replace=TRUE)

  data1$Gen <- Gen_vector

  set.seed(400+i)

  possible_MaxBD <- c("<100", ">=100")

  MaxBD_vector <- sample(possible_MaxBD, 10000, replace=TRUE)

```



```

data1$MaxBD <- MaxBD_vector

set.seed(600+i)

possible_IncAge <- 22:100

IncAge_vector <- sample(possible_IncAge, 10000, replace=TRUE)

data1$IncAge <- IncAge_vector


set.seed(800+i)

possible_MinEP <- c("<0.5", ">=0.5")

MinEP_vector <- sample(possible_MinEP, 10000, replace=TRUE)

data1$MinEP <- MinEP_vector

data2 <- data1

data3 <- data1

data4 <- data1

data5 <- data1


data2$IncAge <- data1$IncAge + 1

data3$IncAge <- data1$IncAge + 2

data4$IncAge <- data1$IncAge + 3

data5$IncAge <- data1$IncAge + 4


full_data <- rbind(data1, data2, data3, data4, data5)

full_data <- transform(full_data, fac1=1, fac2=1, fac3=1, fac4=1, fac5=1, exp_inc_rate=1, random_num=1, new_claim=1, exposure=1)

#full_data$IncAge <- as.integer(full_data$IncAge)

full_data$IncAge <- ifelse(full_data$IncAge<36, 0, ifelse(full_data$IncAge<100, full_data$IncAge, 100))


ID_factor <- matrix(1, nrow(full_data), 5, byrow=TRUE)

colnames(ID_factor) <- c("id1", "id2", "id3", "id4", "id5")

```

```

ID_factor <- data.frame(ID_factor)

#ID_factor$Id1

ID_factor$Id2 <- full_data$Gen

ID_factor$Id3 <- full_data$MaxBD

ID_factor$Id4 <- ifelse(full_data$IncAge<64, "<65",
                        ifelse(full_data$IncAge<75, ">=65,<75",
                                ifelse(full_data$IncAge<85, ">=75,<85",
                                        ">=85"))))

ID_factor$Id5 <- full_data$MinEP

## Update factor values

full_data$fac1 <- as.numeric(factor1$Base[1])

full_data$fac2 <- as.numeric(factor2[match(ID_factor$Id2, factor2$Gen), which(colnames(factor2)=="Gen_Val")])

full_data$fac3 <- as.numeric(factor3[match(ID_factor$Id3, factor3$MaxBD), which(colnames(factor3)=="MaxBD_Val")])

full_data$fac4 <- as.numeric(factor4[match(ID_factor$Id4, factor4$IncAge), which(colnames(factor4)=="IncAge_Val")])

full_data$fac5 <- as.numeric(factor5[match(ID_factor$Id5, factor5$MinEP), which(colnames(factor5)=="MinEP_Val")])

## Update xpected incidence rates

full_data$exp_inc_rate <- full_data$fac1*full_data$fac2*full_data$fac3*full_data$fac4*full_data$fac5

## Generate new claims and fit Poisson GLM

set.seed(1000+i)

ran_num <- runif(50000, 0, 1)

full_data$random_num <- ran_num

full_data$new_claim <- ifelse(full_data$random_num<full_data$exp_inc_rate, 1, 0)

## Modify numerical variables for Poisson GLM

full_data$IncAge_GLM <- ifelse(full_data$IncAge<64, "<65",

```

```

        ifelse(full_data$IncAge<75, ">=65,<75",

        ifelse(full_data$IncAge<85, ">=75,<85",

        ">=85"))))

## Fit Poisson GLM

lr1 <- glm(formula = new_claim/exposure ~

        Gen + MaxBD + IncAge_GLM + MinEP,

        data=full_data, family=poisson)

#lr1

#summary(lr1)

para_stdDev <- summary(lr1)$coefficients[, 1:2]

#parameters <- confint(lr1)

#name_para <- names(lr1$coefficients)

#length(lr1$coefficients)

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=rownames(para_stdDev), para_stdDev)

## Match SOA parameter and fitted parameter

initial_parameter_i <- initial_parameter

initial_parameter_i$estimate <- para_stdDev[match(initial_parameter_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Estimate")]

initial_parameter_i$estimate <- ifelse(is.na(initial_parameter_i$estimate)==TRUE, 0, initial_parameter_i$estimate)

initial_parameter_i$stdError <- para_stdDev[match(initial_parameter_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Std.Error")]

initial_parameter_i$stdError <- ifelse(is.na(initial_parameter_i$stdError)==TRUE, 0, initial_parameter_i$stdError)

## Construct confidence interval for estimated beta

confidence <- 0.90

c1 <- (1-confidence)/2

```

```

c2 <- c1 + confidence

initial_parameter_i$lower_estimate <- qnorm(c1, mean=initial_parameter_i$estimate, sd=initial_parameter_i$stdError)

initial_parameter_i$upper_estimate <- qnorm(c2, mean=initial_parameter_i$estimate, sd=initial_parameter_i$stdError)

initial_parameter_i$fail <- ifelse(initial_parameter_i$Beta_Val<initial_parameter_i$lower_estimate, 1,
                                ifelse(initial_parameter_i$Beta_Val>initial_parameter_i$upper_estimate, 1, 0))

#sum(initial_parameter_i$fail)

intercept[i, 1] <- initial_parameter_i$Beta_Val[1]

intercept[i, 2] <- initial_parameter_i$estimate[1]

intercept[i, 3] <- initial_parameter_i$stdError[1]

intercept[i, 4] <- initial_parameter_i$lower_estimate[1]

intercept[i, 5] <- initial_parameter_i$upper_estimate[1]

intercept[i, 6] <- initial_parameter_i$fail[1]


GenF[i, 1] <- initial_parameter_i$Beta_Val[2]

GenF[i, 2] <- initial_parameter_i$estimate[2]

GenF[i, 3] <- initial_parameter_i$stdError[2]

GenF[i, 4] <- initial_parameter_i$lower_estimate[2]

GenF[i, 5] <- initial_parameter_i$upper_estimate[2]

GenF[i, 6] <- initial_parameter_i$fail[2]


GenM[i, 1] <- initial_parameter_i$Beta_Val[3]

GenM[i, 2] <- initial_parameter_i$estimate[3]

GenM[i, 3] <- initial_parameter_i$stdError[3]

GenM[i, 4] <- initial_parameter_i$lower_estimate[3]

GenM[i, 5] <- initial_parameter_i$upper_estimate[3]

GenM[i, 6] <- initial_parameter_i$fail[3]


BenAmountLess100[i, 1] <- initial_parameter_i$Beta_Val[4]

```

```

BenAmountLess100[i, 2] <- initial_parameter_i$estimate[4]

BenAmountLess100[i, 3] <- initial_parameter_i$stdError[4]

BenAmountLess100[i, 4] <- initial_parameter_i$lower_estimate[4]

BenAmountLess100[i, 5] <- initial_parameter_i$upper_estimate[4]

BenAmountLess100[i, 6] <- initial_parameter_i$fail[4]

```

```

BenAmountMore100[i, 1] <- initial_parameter_i$Beta_Val[5]

BenAmountMore100[i, 2] <- initial_parameter_i$estimate[5]

BenAmountMore100[i, 3] <- initial_parameter_i$stdError[5]

BenAmountMore100[i, 4] <- initial_parameter_i$lower_estimate[5]

BenAmountMore100[i, 5] <- initial_parameter_i$upper_estimate[5]

BenAmountMore100[i, 6] <- initial_parameter_i$fail[5]

```

```

Age1[i, 1] <- initial_parameter_i$Beta_Val[6]

Age1[i, 2] <- initial_parameter_i$estimate[6]

Age1[i, 3] <- initial_parameter_i$stdError[6]

Age1[i, 4] <- initial_parameter_i$lower_estimate[6]

Age1[i, 5] <- initial_parameter_i$upper_estimate[6]

Age1[i, 6] <- initial_parameter_i$fail[6]

```

```

Age2[i, 1] <- initial_parameter_i$Beta_Val[7]

Age2[i, 2] <- initial_parameter_i$estimate[7]

Age2[i, 3] <- initial_parameter_i$stdError[7]

Age2[i, 4] <- initial_parameter_i$lower_estimate[7]

Age2[i, 5] <- initial_parameter_i$upper_estimate[7]

Age2[i, 6] <- initial_parameter_i$fail[7]

```

```

Age3[i, 1] <- initial_parameter_i$Beta_Val[8]

```

```

Age3[i, 2] <- initial_parameter_i$estimate[8]

Age3[i, 3] <- initial_parameter_i$stdError[8]

Age3[i, 4] <- initial_parameter_i$lower_estimate[8]

Age3[i, 5] <- initial_parameter_i$supper_estimate[8]

Age3[i, 6] <- initial_parameter_i$fail[8]


Age4[i, 1] <- initial_parameter_i$Beta_Val[9]

Age4[i, 2] <- initial_parameter_i$estimate[9]

Age4[i, 3] <- initial_parameter_i$stdError[9]

Age4[i, 4] <- initial_parameter_i$lower_estimate[9]

Age4[i, 5] <- initial_parameter_i$supper_estimate[9]

Age4[i, 6] <- initial_parameter_i$fail[9]


WaitingHalf[i, 1] <- initial_parameter_i$Beta_Val[10]

WaitingHalf[i, 2] <- initial_parameter_i$estimate[10]

WaitingHalf[i, 3] <- initial_parameter_i$stdError[10]

WaitingHalf[i, 4] <- initial_parameter_i$lower_estimate[10]

WaitingHalf[i, 5] <- initial_parameter_i$supper_estimate[10]

WaitingHalf[i, 6] <- initial_parameter_i$fail[10]

WaitingMoreHalf[i, 1] <- initial_parameter_i$Beta_Val[11]

WaitingMoreHalf[i, 2] <- initial_parameter_i$estimate[11]

WaitingMoreHalf[i, 3] <- initial_parameter_i$stdError[11]

WaitingMoreHalf[i, 4] <- initial_parameter_i$lower_estimate[11]

WaitingMoreHalf[i, 5] <- initial_parameter_i$supper_estimate[11]

WaitingMoreHalf[i, 6] <- initial_parameter_i$fail[11]

}

ErrorTypeI <- rep(1, 11)

```

```

ErrorTypeI <- data.frame(ErrorTypeI)

rownames(ErrorTypeI) <- c("intercept", "GenF", "GenM", "BenAmountLess100", "BenAmountMore100",
                          "Age1", "Age2", "Age3", "Age4", "WaitingHalf", "WaitingMoreHalf")

ErrorTypeI[1,] <- sum(intercept$Fail/100)

ErrorTypeI[2,] <- sum(GenF$Fail/100)

ErrorTypeI[3,] <- sum(GenM$Fail/100)

ErrorTypeI[4,] <- sum(BenAmountLess100$Fail/100)

ErrorTypeI[5,] <- sum(BenAmountMore100$Fail/100)

ErrorTypeI[6,] <- sum(Age1$Fail/100)

ErrorTypeI[7,] <- sum(Age2$Fail/100)

ErrorTypeI[8,] <- sum(Age3$Fail/100)

ErrorTypeI[9,] <- sum(Age4$Fail/100)

ErrorTypeI[10,] <- sum(WaitingHalf$Fail/100)

ErrorTypeI[11,] <- sum(WaitingMoreHalf$Fail/100)

ErrorTypeI

```

## A12: R Code for Type II Error

```

factor1 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor1_Base.csv",
header=TRUE)

factor2 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor2_Gen.csv",
header=TRUE)

factor3 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor3_MaxBD.csv",
header=TRUE)

factor4 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor4_IncAge.csv",
header=TRUE)

factor5 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor5_MinEP.csv",
header=TRUE)

initial_parameter <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Parameter.csv",
header=TRUE)

```

```
n <- 100
```

```
intercept <- matrix(1, n, 6)
```

```
colnames(intercept) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
intercept <- data.frame(intercept)
```

```
GenF <- matrix(1, n, 6)
```

```
colnames(GenF) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
GenF <- data.frame(GenF)
```

```
GenM <- matrix(1, n, 6)
```

```
colnames(GenM) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
GenM <- data.frame(GenM)
```

```
BenAmountLess100 <- matrix(1, n, 6)
```

```
colnames(BenAmountLess100) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
BenAmountLess100 <- data.frame(BenAmountLess100)
```

```
BenAmountMore100 <- matrix(1, n, 6)
```

```
colnames(BenAmountMore100) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
BenAmountMore100 <- data.frame(BenAmountMore100)
```

```
Age1 <- matrix(1, n, 6)
```

```
colnames(Age1) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
Age1 <- data.frame(Age1)
```

```
Age2 <- matrix(1, n, 6)
```

```
colnames(Age2) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```



```
Age2 <- data.frame(Age2)
```

```
Age3 <- matrix(1, n, 6)
```

```
colnames(Age3) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
Age3 <- data.frame(Age3)
```

```
Age4 <- matrix(1, n, 6)
```

```
colnames(Age4) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
Age4 <- data.frame(Age4)
```

```
WaitingHalf <- matrix(1, n, 6)
```

```
colnames(WaitingHalf) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
WaitingHalf <- data.frame(WaitingHalf)
```

```
WaitingMoreHalf <- matrix(1, n, 6)
```

```
colnames(WaitingMoreHalf) <- c("SOA", "Fitted", "Error", "Lower", "Upper", "Fail")
```

```
WaitingMoreHalf <- data.frame(WaitingMoreHalf)
```

```
for(i in 1:n){
```

```
  ## Generate 10000 data points for 5 years with 4 variables
```

```
  data1 <- matrix(1, 10000, 4, byrow=TRUE)
```

```
  #rownames(data1) <- 1:10000
```

```
  colnames(data1) <- c("Gen", "MaxBD", "IncAge", "MinEP")
```

```
  #str(data1)
```

```
  data1 <- data.frame(data1)
```

```
  #str(data1)
```

```
  set.seed(200+i)
```

```

possible_Gen <- c("F", "M")

Gen_vector <- sample(possible_Gen, 10000, replace=TRUE)

data1$Gen <- Gen_vector


set.seed(400+i)

possible_MaxBD <- c("<100", ">=100")

MaxBD_vector <- sample(possible_MaxBD, 10000, replace=TRUE)

data1$MaxBD <- MaxBD_vector


set.seed(600+i)

possible_IncAge <- 22:100

IncAge_vector <- sample(possible_IncAge, 10000, replace=TRUE)

data1$IncAge <- IncAge_vector


set.seed(800+i)

possible_MinEP <- c("<0.5", ">=0.5")

MinEP_vector <- sample(possible_MinEP, 10000, replace=TRUE)

data1$MinEP <- MinEP_vector


data2 <- data1

data3 <- data1

data4 <- data1

data5 <- data1


data2$IncAge <- data1$IncAge + 1

data3$IncAge <- data1$IncAge + 2

data4$IncAge <- data1$IncAge + 3

data5$IncAge <- data1$IncAge + 4

```

```

full_data <- rbind(data1, data2, data3, data4, data5)

full_data <- transform(full_data, fac1=1, fac2=1, fac3=1, fac4=1, fac5=1, exp_inc_rate=1, random_num=1, new_claim=1, exposure=1)

#full_data$IncAge <- as.integer(full_data$IncAge)

full_data$IncAge <- ifelse(full_data$IncAge<36, 0, ifelse(full_data$IncAge<100, full_data$IncAge, 100))


ID_factor <- matrix(1, nrow(full_data), 5, byrow=TRUE)

colnames(ID_factor) <- c("id1", "id2", "id3", "id4", "id5")

ID_factor <- data.frame(ID_factor)


#ID_factor$Id1

ID_factor$Id2 <- full_data$Gen

ID_factor$Id3 <- full_data$MaxBD

ID_factor$Id4 <- ifelse(full_data$IncAge<64, "<65",
  ifelse(full_data$IncAge<75, ">=65,<75",
    ifelse(full_data$IncAge<85, ">=75,<85",
      ">=85"))))

ID_factor$Id5 <- full_data$MinEP


## Update factor values

full_data$fac1 <- as.numeric(factor1$Base[1])

full_data$fac2 <- as.numeric(factor2[match(ID_factor$Id2, factor2$Gen), which(colnames(factor2)=="Gen_Val")])

full_data$fac3 <- as.numeric(factor3[match(ID_factor$Id3, factor3$MaxBD), which(colnames(factor3)=="MaxBD_Val")])

full_data$fac4 <- as.numeric(factor4[match(ID_factor$Id4, factor4$IncAge), which(colnames(factor4)=="IncAge_Val")])

full_data$fac5 <- as.numeric(factor5[match(ID_factor$Id5, factor5$MinEP), which(colnames(factor5)=="MinEP_Val")])


## Update expected incidence rates

full_data$exp_inc_rate <- full_data$fac1*full_data$fac2*full_data$fac3*full_data$fac4*full_data$fac5

```

```

## Shock female expected incidence rate by factor shock_f

shock_f <- 10

full_data$exp_inc_rate <- ifelse(full_data$Gen=="F", shock_f*full_data$exp_inc_rate, full_data$exp_inc_rate)


## Generate new claims and fit Poisson GLM

set.seed(1000+i)

ran_num <- runif(50000, 0, 1)

full_data$random_num <- ran_num

full_data$new_claim <- ifelse(full_data$random_num<full_data$exp_inc_rate, 1, 0)


## Modify numerical variables for Poisson GLM

full_data$IncAge_GLM <- ifelse(full_data$IncAge<64, "<65",
                               ifelse(full_data$IncAge<75, ">=65,<75",
                                       ifelse(full_data$IncAge<85, ">=75,<85",
                                             ">=85"))))


## Fit Poisson GLM

lr1 <- glm(formula = new_claim/exposure ~
           Gen + MaxBD + IncAge_GLM + MinEP,
           data=full_data, family=poisson)

#lr1

#summary(lr1)

para_stdDev <- summary(lr1)$coefficients[, 1:2]

#parameters <- confint(lr1)

#name_para <- names(lr1$coefficients)

#length(lr1$coefficients)

```

```

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=rownames(para_stdDev), para_stdDev)

## Match SOA parameter and fitted parameter

initial_parameter_i <- initial_parameter

initial_parameter_i$estimate <- para_stdDev[match(initial_parameter_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Estimate")]

initial_parameter_i$estimate <- ifelse(is.na(initial_parameter_i$estimate)==TRUE), 0, initial_parameter_i$estimate)

initial_parameter_i$stdError <- para_stdDev[match(initial_parameter_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Std..Error")]

initial_parameter_i$stdError <- ifelse(is.na(initial_parameter_i$stdError)==TRUE), 0, initial_parameter_i$stdError)

## Construct confidence interval for estimated beta

confidence <- 0.99

c1 <- (1-confidence)/2

c2 <- c1 + confidence

initial_parameter_i$lower_estimate <- qnorm(c1, mean=initial_parameter_i$estimate, sd=initial_parameter_i$stdError)

initial_parameter_i$supper_estimate <- qnorm(c2, mean=initial_parameter_i$estimate, sd=initial_parameter_i$stdError)

initial_parameter_i$fail <- ifelse(initial_parameter_i$Beta_Val<initial_parameter_i$lower_estimate, 0,
                                ifelse(initial_parameter_i$Beta_Val>initial_parameter_i$supper_estimate, 0, 1))

#sum(initial_parameter$fail)

intercept[i, 1] <- initial_parameter_i$Beta_Val[1]

intercept[i, 2] <- initial_parameter_i$estimate[1]

intercept[i, 3] <- initial_parameter_i$stdError[1]

intercept[i, 4] <- initial_parameter_i$lower_estimate[1]

intercept[i, 5] <- initial_parameter_i$supper_estimate[1]

intercept[i, 6] <- initial_parameter_i$fail[1]

GenF[i, 1] <- initial_parameter_i$Beta_Val[2]

```

```

GenF[i, 2] <- initial_parameter_i$estimate[2]

GenF[i, 3] <- initial_parameter_i$stdError[2]

GenF[i, 4] <- initial_parameter_i$lower_estimate[2]

GenF[i, 5] <- initial_parameter_i$supper_estimate[2]

GenF[i, 6] <- initial_parameter_i$fail[2]


GenM[i, 1] <- initial_parameter_i$Beta_Val[3]

GenM[i, 2] <- initial_parameter_i$estimate[3]

GenM[i, 3] <- initial_parameter_i$stdError[3]

GenM[i, 4] <- initial_parameter_i$lower_estimate[3]

GenM[i, 5] <- initial_parameter_i$supper_estimate[3]

GenM[i, 6] <- initial_parameter_i$fail[3]


BenAmountLess100[i, 1] <- initial_parameter_i$Beta_Val[4]

BenAmountLess100[i, 2] <- initial_parameter_i$estimate[4]

BenAmountLess100[i, 3] <- initial_parameter_i$stdError[4]

BenAmountLess100[i, 4] <- initial_parameter_i$lower_estimate[4]

BenAmountLess100[i, 5] <- initial_parameter_i$supper_estimate[4]

BenAmountLess100[i, 6] <- initial_parameter_i$fail[4]


BenAmountMore100[i, 1] <- initial_parameter_i$Beta_Val[5]

BenAmountMore100[i, 2] <- initial_parameter_i$estimate[5]

BenAmountMore100[i, 3] <- initial_parameter_i$stdError[5]

BenAmountMore100[i, 4] <- initial_parameter_i$lower_estimate[5]

BenAmountMore100[i, 5] <- initial_parameter_i$supper_estimate[5]

BenAmountMore100[i, 6] <- initial_parameter_i$fail[5]


Age1[i, 1] <- initial_parameter_i$Beta_Val[6]

```

```

Age1[i, 2] <- initial_parameter_i$estimate[6]

Age1[i, 3] <- initial_parameter_i$stdError[6]

Age1[i, 4] <- initial_parameter_i$lower_estimate[6]

Age1[i, 5] <- initial_parameter_i$upper_estimate[6]

Age1[i, 6] <- initial_parameter_i$fail[6]


Age2[i, 1] <- initial_parameter_i$Beta_Val[7]

Age2[i, 2] <- initial_parameter_i$estimate[7]

Age2[i, 3] <- initial_parameter_i$stdError[7]

Age2[i, 4] <- initial_parameter_i$lower_estimate[7]

Age2[i, 5] <- initial_parameter_i$upper_estimate[7]

Age2[i, 6] <- initial_parameter_i$fail[7]


Age3[i, 1] <- initial_parameter_i$Beta_Val[8]

Age3[i, 2] <- initial_parameter_i$estimate[8]

Age3[i, 3] <- initial_parameter_i$stdError[8]

Age3[i, 4] <- initial_parameter_i$lower_estimate[8]

Age3[i, 5] <- initial_parameter_i$upper_estimate[8]

Age3[i, 6] <- initial_parameter_i$fail[8]


Age4[i, 1] <- initial_parameter_i$Beta_Val[9]

Age4[i, 2] <- initial_parameter_i$estimate[9]

Age4[i, 3] <- initial_parameter_i$stdError[9]

Age4[i, 4] <- initial_parameter_i$lower_estimate[9]

Age4[i, 5] <- initial_parameter_i$upper_estimate[9]

Age4[i, 6] <- initial_parameter_i$fail[9]


WaitingHalf[i, 1] <- initial_parameter_i$Beta_Val[10]

```

```

WaitingHalf[i, 2] <- initial_parameter_i$estimate[10]

WaitingHalf[i, 3] <- initial_parameter_i$stdError[10]

WaitingHalf[i, 4] <- initial_parameter_i$lower_estimate[10]

WaitingHalf[i, 5] <- initial_parameter_i$upper_estimate[10]

WaitingHalf[i, 6] <- initial_parameter_i$fail[10]


WaitingMoreHalf[i, 1] <- initial_parameter_i$Beta_Val[11]

WaitingMoreHalf[i, 2] <- initial_parameter_i$estimate[11]

WaitingMoreHalf[i, 3] <- initial_parameter_i$stdError[11]

WaitingMoreHalf[i, 4] <- initial_parameter_i$lower_estimate[11]

WaitingMoreHalf[i, 5] <- initial_parameter_i$upper_estimate[11]

WaitingMoreHalf[i, 6] <- initial_parameter_i$fail[11]

}


ErrorTypeII <- rep(1, 11)

ErrorTypeII <- data.frame(ErrorTypeII)

rownames(ErrorTypeII) <- c("intercept", "GenF", "GenM", "BenAmountLess100", "BenAmountMore100",

      "Age1", "Age2", "Age3", "Age4", "WaitingHalf", "WaitingMoreHalf")

ErrorTypeII[1,] <- sum(intercept$Fail/100)

ErrorTypeII[2,] <- sum(GenF$Fail/100)

ErrorTypeII[3,] <- sum(GenM$Fail/100)

ErrorTypeII[4,] <- sum(BenAmountLess100$Fail/100)

ErrorTypeII[5,] <- sum(BenAmountMore100$Fail/100)

ErrorTypeII[6,] <- sum(Age1$Fail/100)

ErrorTypeII[7,] <- sum(Age2$Fail/100)

ErrorTypeII[8,] <- sum(Age3$Fail/100)

ErrorTypeII[9,] <- sum(Age4$Fail/100)

ErrorTypeII[10,] <- sum(WaitingHalf$Fail/100)

```



```
ErrorTypeII[11,] <- sum(WaitingMoreHalf$Fail/100)
```

```
ErrorTypeII
```

### **A13: R Code for Type III Error**

```
factor1 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor1_Base.csv",  
header=TRUE)
```

```
factor2 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor2_Gen.csv",  
header=TRUE)
```

```
factor3 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor3_MaxBD.csv",  
header=TRUE)
```

```
factor4 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor4_IncAge.csv",  
header=TRUE)
```

```
factor5 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Factor5_MinEP.csv",  
header=TRUE)
```

```
initial_parameter <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Parameter.csv",  
header=TRUE)
```

```
n <- 100
```

```
intercept_1 <- matrix(1, n, 6)
```

```
colnames(intercept_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")
```

```
intercept_1 <- data.frame(intercept_1)
```

```
GenF_1 <- matrix(1, n, 6)
```

```
colnames(GenF_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")
```

```
GenF_1 <- data.frame(GenF_1)
```

```
GenM_1 <- matrix(1, n, 6)
```

```
colnames(GenM_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")
```

```
GenM_1 <- data.frame(GenM_1)
```

```

BenAmountLess100_1 <- matrix(1, n, 6)

colnames(BenAmountLess100_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

BenAmountLess100_1 <- data.frame(BenAmountLess100_1)


BenAmountMore100_1 <- matrix(1, n, 6)

colnames(BenAmountMore100_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

BenAmountMore100_1 <- data.frame(BenAmountMore100_1)


Age1_1 <- matrix(1, n, 6)

colnames(Age1_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

Age1_1 <- data.frame(Age1_1)


Age2_1 <- matrix(1, n, 6)

colnames(Age2_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

Age2_1 <- data.frame(Age2_1)


Age3_1 <- matrix(1, n, 6)

colnames(Age3_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

Age3_1 <- data.frame(Age3_1)


Age4_1 <- matrix(1, n, 6)

colnames(Age4_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

Age4_1 <- data.frame(Age4_1)


WaitingHalf_1 <- matrix(1, n, 6)

colnames(WaitingHalf_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

WaitingHalf_1 <- data.frame(WaitingHalf_1)

```

```

WaitingMoreHalf_1 <- matrix(1, n, 6)

colnames(WaitingMoreHalf_1) <- c("Expected", "Fitted", "Error", "Lower", "Upper", "Adjusted")

WaitingMoreHalf_1 <- data.frame(WaitingMoreHalf_1)


intercept_2 <- matrix(1, n, 6)

colnames(intercept_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

intercept_2 <- data.frame(intercept_2)


GenF_2 <- matrix(1, n, 6)

colnames(GenF_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

GenF_2 <- data.frame(GenF_2)


GenM_2 <- matrix(1, n, 6)

colnames(GenM_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

GenM_2 <- data.frame(GenM_2)


BenAmountLess100_2 <- matrix(1, n, 6)

colnames(BenAmountLess100_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

BenAmountLess100_2 <- data.frame(BenAmountLess100_2)


BenAmountMore100_2 <- matrix(1, n, 6)

colnames(BenAmountMore100_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

BenAmountMore100_2 <- data.frame(BenAmountMore100_2)


Age1_2 <- matrix(1, n, 6)

colnames(Age1_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

Age1_2 <- data.frame(Age1_2)

```

```

Age2_2 <- matrix(1, n, 6)

colnames(Age2_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

Age2_2 <- data.frame(Age2_2)


Age3_2 <- matrix(1, n, 6)

colnames(Age3_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

Age3_2 <- data.frame(Age3_2)


Age4_2 <- matrix(1, n, 6)

colnames(Age4_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

Age4_2 <- data.frame(Age4_2)


WaitingHalf_2 <- matrix(1, n, 6)

colnames(WaitingHalf_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

WaitingHalf_2 <- data.frame(WaitingHalf_2)


WaitingMoreHalf_2 <- matrix(1, n, 6)

colnames(WaitingMoreHalf_2) <- c("Adjusted", "Fitted", "Error", "Lower", "Upper", "fail")

WaitingMoreHalf_2 <- data.frame(WaitingMoreHalf_2)


for(i in 1:n){

  ## Generate 10000 data points for 7 years with 4 variables

  data1 <- matrix(1, 10000, 4, byrow=TRUE)

  #rownames(data1) <- 1:10000

  colnames(data1) <- c("Gen", "MaxBD", "IncAge", "MinEP")

  #str(data1)

  data1 <- data.frame(data1)

  #str(data1)

```

```

set.seed(200+i)

possible_Gen <- c("F", "M")

Gen_vector <- sample(possible_Gen, 10000, replace=TRUE)

data1$Gen <- Gen_vector


set.seed(400+i)

possible_MaxBD <- c("<100", ">=100")

MaxBD_vector <- sample(possible_MaxBD, 10000, replace=TRUE)

data1$MaxBD <- MaxBD_vector


set.seed(600+i)

possible_IncAge <- 22:100

IncAge_vector <- sample(possible_IncAge, 10000, replace=TRUE)

data1$IncAge <- IncAge_vector


set.seed(800+i)

possible_MinEP <- c("<0.5", ">=0.5")

MinEP_vector <- sample(possible_MinEP, 10000, replace=TRUE)

data1$MinEP <- MinEP_vector


data2 <- data1

data3 <- data1

data4 <- data1

data5 <- data1

data6 <- data1

data7 <- data1


data2$IncAge <- data1$IncAge + 1

```

```

data3$IncAge <- data1$IncAge + 2

data4$IncAge <- data1$IncAge + 3

data5$IncAge <- data1$IncAge + 4

data6$IncAge <- data1$IncAge + 5

data7$IncAge <- data1$IncAge + 6


full_data <- rbind(data1, data2, data3, data4, data5, data6, data7)

full_data <- transform(full_data, fac1=1, fac2=1, fac3=1, fac4=1, fac5=1, exp_inc_rate=1, random_num=1, new_claim=1, exposure=1)

#full_data$IncAge <- as.integer(full_data$IncAge)

full_data$IncAge <- ifelse(full_data$IncAge<36, 0, ifelse(full_data$IncAge<100, full_data$IncAge, 100))


ID_factor <- matrix(1, nrow(full_data), 5, byrow=TRUE)

colnames(ID_factor) <- c("id1", "id2", "id3", "id4", "id5")

ID_factor <- data.frame(ID_factor)


#ID_factor$id1

ID_factor$id2 <- full_data$Gen

ID_factor$id3 <- full_data$MaxBD

ID_factor$id4 <- ifelse(full_data$IncAge<64, "<65",
                        ifelse(full_data$IncAge<75, ">=65,<75",
                                ifelse(full_data$IncAge<85, ">=75,<85",
                                        ">=85"))))

ID_factor$id5 <- full_data$MinEP


## Update factor values

full_data$fac1 <- as.numeric(factor1$Base[1])

full_data$fac2 <- as.numeric(factor2[match(ID_factor$id2, factor2$Gen), which(colnames(factor2)=="Gen_Val")])

```

```

full_data$fac3 <- as.numeric(factor3[match(ID_factor$Id3, factor3$MaxBD), which(colnames(factor3)=="MaxBD_Val")])

full_data$fac4 <- as.numeric(factor4[match(ID_factor$Id4, factor4$IncAge), which(colnames(factor4)=="IncAge_Val")])

full_data$fac5 <- as.numeric(factor5[match(ID_factor$Id5, factor5$MinEP), which(colnames(factor5)=="MinEP_Val")])


## Update xpected incidence rates

full_data$exp_inc_rate <- full_data$fac1*full_data$fac2*full_data$fac3*full_data$fac4*full_data$fac5


## Shock female expected incidence rate by factor shock_f

shock_f <- 10

full_data$exp_inc_rate <- ifelse(full_data$Gen=="F", shock_f*full_data$exp_inc_rate, full_data$exp_inc_rate)


## Generate new claims and fit Poisson GLM

set.seed(1000+i)

ran_num <- runif(70000, 0, 1)

full_data$random_num <- ran_num

full_data$new_claim <- ifelse(full_data$random_num<full_data$exp_inc_rate, 1, 0)


## Modify numerical variables for Poisson GLM

full_data$IncAge_GLM <- ifelse(full_data$IncAge<64, "<65",
                             ifelse(full_data$IncAge<75, ">=65,<75",
                                     ifelse(full_data$IncAge<85, ">=75,<85",
                                             ">=85"))))


## Split data into 2 parts: Testing & Validation

full_data_complete <- full_data

full_data <- full_data_complete[1:50000, ]

full_data_validation <- full_data_complete[50001:70000, ]

```

```

## Fit Poisson GLM to testing data

lr1 <- glm(formula = new_claim/exposure ~

          Gen + MaxBD + IncAge_GLM + MinEP,

          data=full_data, family=poisson)

#lr1

#summary(lr1)

para_stdDev <- summary(lr1)$coefficients[, 1:2]

#parameters <- confint(lr1)

#name_para <- names(lr1$coefficients)

#length(lr1$coefficients)

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=rownames(para_stdDev), para_stdDev)


## Match expected parameter and fitted parameter

initial_parameter_i <- initial_parameter

initial_parameter_i$estimate <- para_stdDev[match(initial_parameter_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Estimate")]

initial_parameter_i$estimate <- ifelse(is.na(initial_parameter_i$estimate)==TRUE), 0, initial_parameter_i$estimate)

initial_parameter_i$stdError <- para_stdDev[match(initial_parameter_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Std..Error")]

initial_parameter_i$stdError <- ifelse(is.na(initial_parameter_i$stdError)==TRUE), 0, initial_parameter_i$stdError)


## Construct confidence interval for estimated beta

confidence <- 0.99

c1 <- (1-confidence)/2

c2 <- c1 + confidence

initial_parameter_i$lower_estimate <- qnorm(c1, mean=initial_parameter_i$estimate, sd=initial_parameter_i$stdError)

initial_parameter_i$upper_estimate <- qnorm(c2, mean=initial_parameter_i$estimate, sd=initial_parameter_i$stdError)

```



```

initial_parameter_i$adjusted <- ifelse(initial_parameter_i$Beta_Val<initial_parameter_i$lower_estimate, initial_parameter_i$lower_estimate,

                                     ifelse(initial_parameter_i$Beta_Val>initial_parameter_i$upper_estimate, initial_parameter_i$upper_estimate,

                                     initial_parameter_i$Beta_Val))

## Update parameter table from Poisson GLM fit 1

intercept_1[i, 1] <- initial_parameter_i$Beta_Val[1]

intercept_1[i, 2] <- initial_parameter_i$Sestimate[1]

intercept_1[i, 3] <- initial_parameter_i$stdError[1]

intercept_1[i, 4] <- initial_parameter_i$lower_estimate[1]

intercept_1[i, 5] <- initial_parameter_i$upper_estimate[1]

intercept_1[i, 6] <- initial_parameter_i$adjusted[1]


GenF_1[i, 1] <- initial_parameter_i$Beta_Val[2]

GenF_1[i, 2] <- initial_parameter_i$Sestimate[2]

GenF_1[i, 3] <- initial_parameter_i$stdError[2]

GenF_1[i, 4] <- initial_parameter_i$lower_estimate[2]

GenF_1[i, 5] <- initial_parameter_i$upper_estimate[2]

GenF_1[i, 6] <- initial_parameter_i$adjusted[2]


GenM_1[i, 1] <- initial_parameter_i$Beta_Val[3]

GenM_1[i, 2] <- initial_parameter_i$Sestimate[3]

GenM_1[i, 3] <- initial_parameter_i$stdError[3]

GenM_1[i, 4] <- initial_parameter_i$lower_estimate[3]

GenM_1[i, 5] <- initial_parameter_i$upper_estimate[3]

GenM_1[i, 6] <- initial_parameter_i$adjusted[3]


BenAmountLess100_1[i, 1] <- initial_parameter_i$Beta_Val[4]

BenAmountLess100_1[i, 2] <- initial_parameter_i$Sestimate[4]

```

```

BenAmountLess100_1[i, 3] <- initial_parameter_i$stdError[4]

BenAmountLess100_1[i, 4] <- initial_parameter_i$lower_estimate[4]

BenAmountLess100_1[i, 5] <- initial_parameter_i$supper_estimate[4]

BenAmountLess100_1[i, 6] <- initial_parameter_i$adjusted[4]


BenAmountMore100_1[i, 1] <- initial_parameter_i$Beta_Val[5]

BenAmountMore100_1[i, 2] <- initial_parameter_i$estimate[5]

BenAmountMore100_1[i, 3] <- initial_parameter_i$stdError[5]

BenAmountMore100_1[i, 4] <- initial_parameter_i$lower_estimate[5]

BenAmountMore100_1[i, 5] <- initial_parameter_i$supper_estimate[5]

BenAmountMore100_1[i, 6] <- initial_parameter_i$adjusted[5]


Age1_1[i, 1] <- initial_parameter_i$Beta_Val[6]

Age1_1[i, 2] <- initial_parameter_i$estimate[6]

Age1_1[i, 3] <- initial_parameter_i$stdError[6]

Age1_1[i, 4] <- initial_parameter_i$lower_estimate[6]

Age1_1[i, 5] <- initial_parameter_i$supper_estimate[6]

Age1_1[i, 6] <- initial_parameter_i$adjusted[6]


Age2_1[i, 1] <- initial_parameter_i$Beta_Val[7]

Age2_1[i, 2] <- initial_parameter_i$estimate[7]

Age2_1[i, 3] <- initial_parameter_i$stdError[7]

Age2_1[i, 4] <- initial_parameter_i$lower_estimate[7]

Age2_1[i, 5] <- initial_parameter_i$supper_estimate[7]

Age2_1[i, 6] <- initial_parameter_i$adjusted[7]


Age3_1[i, 1] <- initial_parameter_i$Beta_Val[8]

Age3_1[i, 2] <- initial_parameter_i$estimate[8]

```

```

Age3_1[i, 3] <- initial_parameter_i$stdError[8]

Age3_1[i, 4] <- initial_parameter_i$lower_estimate[8]

Age3_1[i, 5] <- initial_parameter_i$upper_estimate[8]

Age3_1[i, 6] <- initial_parameter_i$adjusted[8]


Age4_1[i, 1] <- initial_parameter_i$Beta_Val[9]

Age4_1[i, 2] <- initial_parameter_i$estimate[9]

Age4_1[i, 3] <- initial_parameter_i$stdError[9]

Age4_1[i, 4] <- initial_parameter_i$lower_estimate[9]

Age4_1[i, 5] <- initial_parameter_i$upper_estimate[9]

Age4_1[i, 6] <- initial_parameter_i$adjusted[9]


WaitingHalf_1[i, 1] <- initial_parameter_i$Beta_Val[10]

WaitingHalf_1[i, 2] <- initial_parameter_i$estimate[10]

WaitingHalf_1[i, 3] <- initial_parameter_i$stdError[10]

WaitingHalf_1[i, 4] <- initial_parameter_i$lower_estimate[10]

WaitingHalf_1[i, 5] <- initial_parameter_i$upper_estimate[10]

WaitingHalf_1[i, 6] <- initial_parameter_i$adjusted[10]


WaitingMoreHalf_1[i, 1] <- initial_parameter_i$Beta_Val[11]

WaitingMoreHalf_1[i, 2] <- initial_parameter_i$estimate[11]

WaitingMoreHalf_1[i, 3] <- initial_parameter_i$stdError[11]

WaitingMoreHalf_1[i, 4] <- initial_parameter_i$lower_estimate[11]

WaitingMoreHalf_1[i, 5] <- initial_parameter_i$upper_estimate[11]

WaitingMoreHalf_1[i, 6] <- initial_parameter_i$adjusted[11]


## Fit Poisson GLM to validation data

lr2 <- glm(formula = new_claim/exposure ~

```

```

    Gen + MaxBD + IncAge_GLM + MinEP,

    data=full_data_validation, family=poisson)

#lr2

#summary(lr2)

para_stdDev <- summary(lr2)$coefficients[, 1:2]

#parameters <- confint(lr2)

#name_para <- names(lr2$coefficients)

#length(lr2$coefficients)

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=rownames(para_stdDev), para_stdDev)


## Match expected parameter (from GLM1) and fitted parameter (from GLM2)

initial_parameter2_i <- cbind(as.data.frame(initial_parameter_i$Beta), initial_parameter_i$adjusted)

names(initial_parameter2_i) <- c("Beta", "Beta_Val")

initial_parameter2_i$estimate <- para_stdDev[match(initial_parameter2_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Estimate")]

initial_parameter2_i$estimate <- ifelse(is.na(initial_parameter2_i$estimate)==TRUE), 0, initial_parameter2_i$estimate)

initial_parameter2_i$stdError <- para_stdDev[match(initial_parameter2_i$Beta, para_stdDev$betas),
which(colnames(para_stdDev)=="Std..Error")]

initial_parameter2_i$stdError <- ifelse(is.na(initial_parameter2_i$stdError)==TRUE), 0, initial_parameter2_i$stdError)


## Construct confidence interval for estimated beta

confidence <- 0.99

c1 <- (1-confidence)/2

c2 <- c1 + confidence

initial_parameter2_i$lower_estimate <- qnorm(c1, mean=initial_parameter2_i$estimate, sd=initial_parameter2_i$stdError)

initial_parameter2_i$upper_estimate <- qnorm(c2, mean=initial_parameter2_i$estimate, sd=initial_parameter2_i$stdError)

initial_parameter2_i$fail <- ifelse(initial_parameter2_i$Beta_Val<initial_parameter2_i$lower_estimate, 1,

                                ifelse(initial_parameter2_i$Beta_Val>initial_parameter2_i$upper_estimate, 1,

```

0))

```
## Update parameter table from Poisson GLM fit 2
```

```
intercept_2[i, 1] <- initial_parameter2_i$Beta_Val[1]
```

```
intercept_2[i, 2] <- initial_parameter2_i$estimate[1]
```

```
intercept_2[i, 3] <- initial_parameter2_i$stdError[1]
```

```
intercept_2[i, 4] <- initial_parameter2_i$lower_estimate[1]
```

```
intercept_2[i, 5] <- initial_parameter2_i$upper_estimate[1]
```

```
intercept_2[i, 6] <- initial_parameter2_i$fail[1]
```

```
GenF_2[i, 1] <- initial_parameter2_i$Beta_Val[2]
```

```
GenF_2[i, 2] <- initial_parameter2_i$estimate[2]
```

```
GenF_2[i, 3] <- initial_parameter2_i$stdError[2]
```

```
GenF_2[i, 4] <- initial_parameter2_i$lower_estimate[2]
```

```
GenF_2[i, 5] <- initial_parameter2_i$upper_estimate[2]
```

```
GenF_2[i, 6] <- initial_parameter2_i$fail[2]
```

```
GenM_2[i, 1] <- initial_parameter2_i$Beta_Val[3]
```

```
GenM_2[i, 2] <- initial_parameter2_i$estimate[3]
```

```
GenM_2[i, 3] <- initial_parameter2_i$stdError[3]
```

```
GenM_2[i, 4] <- initial_parameter2_i$lower_estimate[3]
```

```
GenM_2[i, 5] <- initial_parameter2_i$upper_estimate[3]
```

```
GenM_2[i, 6] <- initial_parameter2_i$fail[3]
```

```
BenAmountLess100_2[i, 1] <- initial_parameter2_i$Beta_Val[4]
```

```
BenAmountLess100_2[i, 2] <- initial_parameter2_i$estimate[4]
```

```
BenAmountLess100_2[i, 3] <- initial_parameter2_i$stdError[4]
```

```
BenAmountLess100_2[i, 4] <- initial_parameter2_i$lower_estimate[4]
```

```

BenAmountLess100_2[i, 5] <- initial_parameter2_i$upper_estimate[4]

BenAmountLess100_2[i, 6] <- initial_parameter2_i$fail[4]


BenAmountMore100_2[i, 1] <- initial_parameter2_i$Beta_Val[5]

BenAmountMore100_2[i, 2] <- initial_parameter2_i$estimate[5]

BenAmountMore100_2[i, 3] <- initial_parameter2_i$stdError[5]

BenAmountMore100_2[i, 4] <- initial_parameter2_i$lower_estimate[5]

BenAmountMore100_2[i, 5] <- initial_parameter2_i$upper_estimate[5]

BenAmountMore100_2[i, 6] <- initial_parameter2_i$fail[5]


Age1_2[i, 1] <- initial_parameter2_i$Beta_Val[6]

Age1_2[i, 2] <- initial_parameter2_i$estimate[6]

Age1_2[i, 3] <- initial_parameter2_i$stdError[6]

Age1_2[i, 4] <- initial_parameter2_i$lower_estimate[6]

Age1_2[i, 5] <- initial_parameter2_i$upper_estimate[6]

Age1_2[i, 6] <- initial_parameter2_i$fail[6]


Age2_2[i, 1] <- initial_parameter2_i$Beta_Val[7]

Age2_2[i, 2] <- initial_parameter2_i$estimate[7]

Age2_2[i, 3] <- initial_parameter2_i$stdError[7]

Age2_2[i, 4] <- initial_parameter2_i$lower_estimate[7]

Age2_2[i, 5] <- initial_parameter2_i$upper_estimate[7]

Age2_2[i, 6] <- initial_parameter2_i$fail[7]


Age3_2[i, 1] <- initial_parameter2_i$Beta_Val[8]

Age3_2[i, 2] <- initial_parameter2_i$estimate[8]

Age3_2[i, 3] <- initial_parameter2_i$stdError[8]

Age3_2[i, 4] <- initial_parameter2_i$lower_estimate[8]

Age3_2[i, 5] <- initial_parameter2_i$upper_estimate[8]

```

```
Age3_2[i, 6] <- initial_parameter2_i$fail[8]
```

```
Age4_2[i, 1] <- initial_parameter2_i$Beta_Val[9]
```

```
Age4_2[i, 2] <- initial_parameter2_i$estimate[9]
```

```
Age4_2[i, 3] <- initial_parameter2_i$stdError[9]
```

```
Age4_2[i, 4] <- initial_parameter2_i$lower_estimate[9]
```

```
Age4_2[i, 5] <- initial_parameter2_i$upper_estimate[9]
```

```
Age4_2[i, 6] <- initial_parameter2_i$fail[9]
```

```
WaitingHalf_2[i, 1] <- initial_parameter2_i$Beta_Val[10]
```

```
WaitingHalf_2[i, 2] <- initial_parameter2_i$estimate[10]
```

```
WaitingHalf_2[i, 3] <- initial_parameter2_i$stdError[10]
```

```
WaitingHalf_2[i, 4] <- initial_parameter2_i$lower_estimate[10]
```

```
WaitingHalf_2[i, 5] <- initial_parameter2_i$upper_estimate[10]
```

```
WaitingHalf_2[i, 6] <- initial_parameter2_i$fail[10]
```

```
WaitingMoreHalf_2[i, 1] <- initial_parameter2_i$Beta_Val[11]
```

```
WaitingMoreHalf_2[i, 2] <- initial_parameter2_i$estimate[11]
```

```
WaitingMoreHalf_2[i, 3] <- initial_parameter2_i$stdError[11]
```

```
WaitingMoreHalf_2[i, 4] <- initial_parameter2_i$lower_estimate[11]
```

```
WaitingMoreHalf_2[i, 5] <- initial_parameter2_i$upper_estimate[11]
```

```
WaitingMoreHalf_2[i, 6] <- initial_parameter2_i$fail[11]
```

```
}
```

```
ErrorTypeIII <- rep(1, 11)
```

```
ErrorTypeIII <- data.frame(ErrorTypeIII)
```

```
rownames(ErrorTypeIII) <- c("intercept", "GenF", "GenM", "BenAmountLess100", "BenAmountMore100",
```

```
    "Age1", "Age2", "Age3", "Age4", "WaitingHalf", "WaitingMoreHalf")
```

```
ErrorTypeIII[1,] <- sum(intercept_2$fail/100)
```

```

ErrorTypeIII[2,] <- sum(GenF_2$fail/100)

ErrorTypeIII[3,] <- sum(GenM_2$fail/100)

ErrorTypeIII[4,] <- sum(BenAmountLess100_2$fail/100)

ErrorTypeIII[5,] <- sum(BenAmountMore100_2$fail/100)

ErrorTypeIII[6,] <- sum(Age1_2$fail/100)

ErrorTypeIII[7,] <- sum(Age2_2$fail/100)

ErrorTypeIII[8,] <- sum(Age3_2$fail/100)

ErrorTypeIII[9,] <- sum(Age4_2$fail/100)

ErrorTypeIII[10,] <- sum(WaitingHalf_2$fail/100)

ErrorTypeIII[11,] <- sum(WaitingMoreHalf_2$fail/100)

ErrorTypeIII

```

## A14: R Code for Credibility Effect

```

factor1 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\Standard Format Assumption Beta Modified1\\Factor1_Base.csv",
header=TRUE)

factor2 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\Standard Format Assumption Beta Modified1\\Factor2_Gen.csv",
header=TRUE)

factor3 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\Standard Format Assumption Beta Modified1\\Factor3_MaxBD.csv",
header=TRUE)

factor4 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\Standard Format Assumption Beta Modified1\\Factor4_IncAge.csv",
header=TRUE)

factor5 <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\Standard Format Assumption Beta Modified1\\Factor5_MinEP.csv",
header=TRUE)

initial_parameter <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\Standard Format Assumption Beta
Modified1\\Parameter.csv", header=TRUE)

## Step 1: Generate 5 years data from expected assumptions ##

## Generate 10000 data points for 5 years with 4 variables

data1 <- matrix(1, 10000, 4, byrow=TRUE)

```



```

#rownames(data1) <- 1:10000

colnames(data1) <- c("Gen", "MaxBD", "IncAge", "MinEP")

#str(data1)

data1 <- data.frame(data1)

#str(data1)


set.seed(1)

possible_Gen <- c("F", "M")

Gen_vector <- sample(possible_Gen, 10000, replace=TRUE)

data1$Gen <- Gen_vector


set.seed(2)

possible_MaxBD <- c("<100", ">=100")

MaxBD_vector <- sample(possible_MaxBD, 10000, replace=TRUE)

data1$MaxBD <- MaxBD_vector


set.seed(3)

possible_IncAge <- 22:100

IncAge_vector <- sample(possible_IncAge, 10000, replace=TRUE)

data1$IncAge <- IncAge_vector


set.seed(4)

possible_MinEP <- c("<0.5", ">=0.5")

MinEP_vector <- sample(possible_MinEP, 10000, replace=TRUE)

data1$MinEP <- MinEP_vector


data2 <- data1

data3 <- data1

```

```

data4 <- data1

data5 <- data1


data2$IncAge <- data1$IncAge + 1

data3$IncAge <- data1$IncAge + 2

data4$IncAge <- data1$IncAge + 3

data5$IncAge <- data1$IncAge + 4


full_data <- rbind(data1, data2, data3, data4, data5)

full_data <- transform(full_data, fac1=1, fac2=1, fac3=1, fac4=1, fac5=1, exp_inc_rate=1, random_num=1, new_claim=1, exposure=1)

#full_data$IncAge <- as.integer(full_data$IncAge)

full_data$IncAge <- ifelse(full_data$IncAge<36, 0, ifelse(full_data$IncAge<100, full_data$IncAge, 100))


ID_factor <- matrix(1, nrow(full_data), 5, byrow=TRUE)

colnames(ID_factor) <- c("id1", "id2", "id3", "id4", "id5")

ID_factor <- data.frame(ID_factor)


#ID_factor$id1

ID_factor$id2 <- full_data$Gen

ID_factor$id3 <- full_data$MaxBD

ID_factor$id4 <- ifelse(full_data$IncAge<64, "<65",
                        ifelse(full_data$IncAge<75, ">=65,<75",
                                ifelse(full_data$IncAge<85, ">=75,<85",
                                        ">=85"))))

ID_factor$id5 <- full_data$MinEP


## Update factor values

full_data$fac1 <- as.numeric(factor1$Base[1])

```

```

full_data$fac2 <- as.numeric(factor2[match(ID_factor$id2, factor2$Gen), which(colnames(factor2)=="Gen_Val")])

full_data$fac3 <- as.numeric(factor3[match(ID_factor$id3, factor3$MaxBD), which(colnames(factor3)=="MaxBD_Val")])

full_data$fac4 <- as.numeric(factor4[match(ID_factor$id4, factor4$IncAge), which(colnames(factor4)=="IncAge_Val")])

full_data$fac5 <- as.numeric(factor5[match(ID_factor$id5, factor5$MinEP), which(colnames(factor5)=="MinEP_Val")])


## Generate new claims and fit Poisson GLM

set.seed(128)

ran_num <- runif(50000, 0, 1)

full_data$random_num <- ran_num

full_data$new_claim <- ifelse(full_data$random_num<full_data$exp_inc_rate, 1, 0)


## Modify numerical variables for Poisson GLM

full_data$IncAge_GLM <- ifelse(full_data$IncAge<64, "<65",
                               ifelse(full_data$IncAge<75, ">=65,<75",
                                       ifelse(full_data$IncAge<85, ">=75,<85",
                                             ">=85")))

## Fit Poisson GLM

lr1 <- glm(formula = new_claim/exposure ~
           Gen + MaxBD + IncAge_GLM + MinEP,
           data=full_data, family=poisson)

#lr1

#summary(lr1)

summary.glm(lr1)$dispersion

para_stdDev <- summary(lr1)$coefficients[, 1:2]

#parameters <- confint(lr1)

#name_para <- names(lr1$coefficients)

#length(lr1$coefficients)

```

```

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=row.names(para_stdDev), para_stdDev)

write.table(para_stdDev, row.names=FALSE, "C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\New Assumption.csv", sep=",")


expected_assumptions <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\New Assumption.csv", header=TRUE)

names(expected_assumptions)[names(expected_assumptions)=="Estimate"] <- "Beta_Initial"

names(expected_assumptions)[names(expected_assumptions)=="Std..Error"] <- "Std_Error_Beta_Initial"


## End: Step 1 ##


## Step 2: Generate 5 years shock & stable (shock is applied for all five years) data ##


data1 <- matrix(1, 10000, 4, byrow=TRUE)

#rownames(data1) <- 1:10000

colnames(data1) <- c("Gen", "MaxBD", "IncAge", "MinEP")

#str(data1)

data1 <- data.frame(data1)

#str(data1)


set.seed(1)

possible_Gen <- c("F", "M")

Gen_vector <- sample(possible_Gen, 10000, replace=TRUE)

data1$Gen <- Gen_vector


set.seed(2)

possible_MaxBD <- c("<100", ">=100")

MaxBD_vector <- sample(possible_MaxBD, 10000, replace=TRUE)

```

```

data1$MaxBD <- MaxBD_vector

set.seed(3)

possible_IncAge <- 22:100

IncAge_vector <- sample(possible_IncAge, 10000, replace=TRUE)

data1$IncAge <- IncAge_vector


set.seed(4)

possible_MinEP <- c("<0.5", ">=0.5")

MinEP_vector <- sample(possible_MinEP, 10000, replace=TRUE)

data1$MinEP <- MinEP_vector


data2 <- data1

data3 <- data1

data4 <- data1

data5 <- data1


data2$IncAge <- data1$IncAge + 1

data3$IncAge <- data1$IncAge + 2

data4$IncAge <- data1$IncAge + 3

data5$IncAge <- data1$IncAge + 4


full_data <- rbind(data1, data2, data3, data4, data5)

full_data <- transform(full_data, fac1=1, fac2=1, fac3=1, fac4=1, fac5=1, exp_inc_rate=1, random_num=1, new_claim=1, exposure=1)

#full_data$IncAge <- as.integer(full_data$IncAge)

full_data$IncAge <- ifelse(full_data$IncAge<36, 0, ifelse(full_data$IncAge<100, full_data$IncAge, 100))


ID_factor <- matrix(1, nrow(full_data), 5, byrow=TRUE)

```

```

colnames(ID_factor) <- c("id1", "id2", "id3", "id4", "id5")

ID_factor <- data.frame(ID_factor)

#ID_factor$id1

ID_factor$id2 <- full_data$Gen

ID_factor$id3 <- full_data$MaxBD

ID_factor$id4 <- ifelse(full_data$IncAge<64, "<65",
                        ifelse(full_data$IncAge<75, ">=65,<75",
                                ifelse(full_data$IncAge<85, ">=75,<85",
                                        ">=85"))))

ID_factor$id5 <- full_data$MinEP

## Update factor values

full_data$fac1 <- as.numeric(factor1$Base[1])

full_data$fac2 <- as.numeric(factor2[match(ID_factor$id2, factor2$Gen), which(colnames(factor2)=="Gen_Val")])

full_data$fac3 <- as.numeric(factor3[match(ID_factor$id3, factor3$MaxBD), which(colnames(factor3)=="MaxBD_Val")])

full_data$fac4 <- as.numeric(factor4[match(ID_factor$id4, factor4$IncAge), which(colnames(factor4)=="IncAge_Val")])

full_data$fac5 <- as.numeric(factor5[match(ID_factor$id5, factor5$MinEP), which(colnames(factor5)=="MinEP_Val")])

## Update xpected incidence rates

full_data$exp_inc_rate <- full_data$fac1*full_data$fac2*full_data$fac3*full_data$fac4*full_data$fac5

## Shock female expected incidence rate by factor shock_f

shock_f <- 10

full_data$exp_inc_rate <- ifelse(full_data$Gen=="F", shock_f*full_data$exp_inc_rate, full_data$exp_inc_rate)

## Generate new claims and fit Poisson GLM

```

```

set.seed(128)

ran_num <- runif(50000, 0, 1)

full_data$random_num <- ran_num

full_data$new_claim <- ifelse(full_data$random_num<full_data$exp_inc_rate, 1, 0)


## Modify numerical variables for Poisson GLM

full_data$IncAge_GLM <- ifelse(full_data$IncAge<64, "<65",

                               ifelse(full_data$IncAge<75, ">=65,<75",

                                       ifelse(full_data$IncAge<85, ">=75,<85",

                                              ">=85"))))

## Fit Poisson GLM

lr1 <- glm(formula = new_claim/exposure ~

           Gen + MaxBD + IncAge_GLM + MinEP,

           data=full_data, family=poisson)

#lr1

#summary(lr1)

summary.glm(lr1)$dispersion

para_stdDev <- summary(lr1)$coefficients[, 1:2]

#parameters <- confint(lr1)

#name_para <- names(lr1$coefficients)

#length(lr1$coefficients)

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=rownames(para_stdDev), para_stdDev)

write.table(para_stdDev, row.names=FALSE, "C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson
GLM\\Shocked Stable Assumption.csv", sep=",")

```

```

shocked_stable_assumptions <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson GLM\\Shocked
Stable Assumption.csv", header=TRUE)

names(shocked_stable_assumptions)[names(shocked_stable_assumptions)=="Estimate"] <- "Beta_Shocked_Stable"

names(shocked_stable_assumptions)[names(shocked_stable_assumptions)=="Std..Error"] <- "Std_Error_Beta_Shocked_Stable"


## End: Step 2 ##

## Step 3: Generate 5 years shock & unstable (shock is applied only for years 2 & 4) data ##


data1 <- matrix(1, 10000, 4, byrow=TRUE)

#rownames(data1) <- 1:10000

colnames(data1) <- c("Gen", "MaxBD", "IncAge", "MinEP")

#str(data1)

data1 <- data.frame(data1)

#str(data1)


set.seed(1)

possible_Gen <- c("F", "M")

Gen_vector <- sample(possible_Gen, 10000, replace=TRUE)

data1$Gen <- Gen_vector


set.seed(2)

possible_MaxBD <- c("<100", ">=100")

MaxBD_vector <- sample(possible_MaxBD, 10000, replace=TRUE)

data1$MaxBD <- MaxBD_vector


set.seed(3)

possible_IncAge <- 22:100

IncAge_vector <- sample(possible_IncAge, 10000, replace=TRUE)

data1$IncAge <- IncAge_vector

```



```

set.seed(4)

possible_MinEP <- c("<0.5", ">=0.5")

MinEP_vector <- sample(possible_MinEP, 10000, replace=TRUE)

data1$MinEP <- MinEP_vector


data2 <- data1

data3 <- data1

data4 <- data1

data5 <- data1


data2$IncAge <- data1$IncAge + 1

data3$IncAge <- data1$IncAge + 2

data4$IncAge <- data1$IncAge + 3

data5$IncAge <- data1$IncAge + 4

full_data <- rbind(data1, data2, data3, data4, data5)

full_data <- transform(full_data, fac1=1, fac2=1, fac3=1, fac4=1, fac5=1, exp_inc_rate=1, random_num=1, new_claim=1, exposure=1)

#full_data$IncAge <- as.integer(full_data$IncAge)

full_data$IncAge <- ifelse(full_data$IncAge<36, 0, ifelse(full_data$IncAge<100, full_data$IncAge, 100))


ID_factor <- matrix(1, nrow(full_data), 5, byrow=TRUE)

colnames(ID_factor) <- c("id1", "id2", "id3", "id4", "id5")

ID_factor <- data.frame(ID_factor)


#ID_factor$id1

ID_factor$id2 <- full_data$Gen

ID_factor$id3 <- full_data$MaxBD

ID_factor$id4 <- ifelse(full_data$IncAge<64, "<65",

```

```

    ifelse(full_data$IncAge<75, ">=65,<75",

    ifelse(full_data$IncAge<85, ">=75,<85",

    ">=85"))))

ID_factor$id5 <- full_data$MinEP

## Update factor values

full_data$fac1 <- as.numeric(factor1$Base[1])

full_data$fac2 <- as.numeric(factor2[match(ID_factor$id2, factor2$Gen), which(colnames(factor2)=="Gen_Val")])

full_data$fac3 <- as.numeric(factor3[match(ID_factor$id3, factor3$MaxBD), which(colnames(factor3)=="MaxBD_Val")])

full_data$fac4 <- as.numeric(factor4[match(ID_factor$id4, factor4$IncAge), which(colnames(factor4)=="IncAge_Val")])

full_data$fac5 <- as.numeric(factor5[match(ID_factor$id5, factor5$MinEP), which(colnames(factor5)=="MinEP_Val")])

## Update expected incidence rates

full_data$exp_inc_rate <- full_data$fac1*full_data$fac2*full_data$fac3*full_data$fac4*full_data$fac5

## Shock female expected incidence rate for years 2 & 4 by factor 1/shock_f

shock_f <- 10

full_data1 <- full_data[1:10000, ]

full_data2 <- full_data[10001:20000, ]

full_data3 <- full_data[20001:30000, ]

full_data4 <- full_data[30001:40000, ]

full_data5 <- full_data[40001:50000, ]

#full_data1$exp_inc_rate <- ifelse(full_data1$Gen=="F", (1/shock_f)*full_data1$exp_inc_rate, full_data1$exp_inc_rate)

full_data2$exp_inc_rate <- ifelse(full_data2$Gen=="F", (1/shock_f)*full_data2$exp_inc_rate, full_data2$exp_inc_rate)

#full_data3$exp_inc_rate <- ifelse(full_data3$Gen=="F", (1/shock_f)*full_data3$exp_inc_rate, full_data3$exp_inc_rate)

full_data4$exp_inc_rate <- ifelse(full_data4$Gen=="F", (1/shock_f)*full_data4$exp_inc_rate, full_data4$exp_inc_rate)

#full_data5$exp_inc_rate <- ifelse(full_data5$Gen=="F", (1/shock_f)*full_data5$exp_inc_rate, full_data5$exp_inc_rate)

full_data <- rbind(full_data1, full_data2, full_data3, full_data4, full_data5)

```

```

full_data <- data.frame(full_data)

## Generate new claims and fit Poisson GLM

set.seed(128)

ran_num <- runif(50000, 0, 1)

full_data$random_num <- ran_num

full_data$new_claim <- ifelse(full_data$random_num<full_data$exp_inc_rate, 1, 0)

## Modify numerical variables for Poisson GLM

full_data$IncAge_GLM <- ifelse(full_data$IncAge<64, "<65",
                               ifelse(full_data$IncAge<75, ">=65,<75",
                                       ifelse(full_data$IncAge<85, ">=75,<85",
                                             ">=85"))))

## Fit Poisson GLM

lr1 <- glm(formula = new_claim/exposure ~
           Gen + MaxBD + IncAge_GLM + MinEP,
           data=full_data, family=poisson)

#lr1

#summary(lr1)

summary.glm(lr1)$dispersion

para_stdDev <- summary(lr1)$coefficients[, 1:2]

#parameters <- confint(lr1)

#name_para <- names(lr1$coefficients)

#length(lr1$coefficients)

para_stdDev <- data.frame(para_stdDev)

para_stdDev <- cbind(betas=rownames(para_stdDev), para_stdDev)

write.table(para_stdDev, row.names=FALSE, "C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson
GLM\\Shocked Unstable Assumption.csv", sep=",")

```

```

shocked_unstable_assumptions <- read.csv("C:\\Users\\Rozita\\Desktop\\PhD Dissertation\\Data\\New Assumption_From Poisson
GLM\\Shocked Unstable Assumption.csv", header=TRUE)

names(shocked_unstable_assumptions)[names(shocked_unstable_assumptions)=="Estimate"] <- "Beta_Shocked_Unstable"

names(shocked_unstable_assumptions)[names(shocked_unstable_assumptions)=="Std..Error"] <- "Std_Error_Beta_Shocked_Unstable"

## End: Step 3 ##

## Step 4: Adjustment for shocked stable beta

adjustment_stable <- merge(expected_assumptions, shocked_stable_assumptions, by="betas")

adjustment_stable$alpha <- 0.32

adjustment_stable$confidence_alpha <- 1- adjustment_stable$alpha

adjustment_stable$variance_ratio <- (adjustment_stable$Std_Error_Beta_Shocked_Stable^2)/(adjustment_stable$Std_Error_Beta_Initial^2)

adjustment_stable$alpha_credibility <- (1 - pnorm(adjustment_stable$variance_ratio))*2

adjustment_stable$confidence_credibility <- 1- adjustment_stable$alpha_credibility

adjustment_stable$c1_alpha <- adjustment_stable$alpha/2

adjustment_stable$c2_alpha <- adjustment_stable$c1_alpha + adjustment_stable$confidence_alpha

adjustment_stable$c1_credibility <- adjustment_stable$alpha_credibility/2

adjustment_stable$c2_credibility <- adjustment_stable$c1_credibility + adjustment_stable$confidence_credibility

adjustment_stable$lower_alpha <- qnorm(adjustment_stable$c1_alpha, mean=adjustment_stable$Beta_Shocked_Stable,
sd=adjustment_stable$Std_Error_Beta_Shocked_Stable)

adjustment_stable$upper_alpha <- qnorm(adjustment_stable$c2_alpha, mean=adjustment_stable$Beta_Shocked_Stable,
sd=adjustment_stable$Std_Error_Beta_Shocked_Stable)

adjustment_stable$lower_credibility <- qnorm(adjustment_stable$c1_credibility, mean=adjustment_stable$Beta_Shocked_Stable,
sd=adjustment_stable$Std_Error_Beta_Shocked_Stable)

adjustment_stable$upper_credibility <- qnorm(adjustment_stable$c2_credibility, mean=adjustment_stable$Beta_Shocked_Stable,
sd=adjustment_stable$Std_Error_Beta_Shocked_Stable)

adjustment_stable$adjusted_beta_alpha <- ifelse(adjustment_stable$Beta_Initial < adjustment_stable$lower_alpha,
adjustment_stable$lower_alpha, ifelse(adjustment_stable$Beta_Initial > adjustment_stable$upper_alpha, adjustment_stable$upper_alpha,
adjustment_stable$Beta_Initial))

adjustment_stable$adjusted_beta_credibility <- ifelse(adjustment_stable$Beta_Initial < adjustment_stable$lower_credibility,
adjustment_stable$lower_credibility, ifelse(adjustment_stable$Beta_Initial > adjustment_stable$upper_credibility,
adjustment_stable$upper_credibility, adjustment_stable$Beta_Initial))

```

```

adjustment_stable$diff_adjusted_alpha <- abs(adjustment_stable$Beta_Initial - adjustment_stable$adjusted_beta_alpha)

adjustment_stable$diff_adjusted_credibility <- abs(adjustment_stable$Beta_Initial - adjustment_stable$adjusted_beta_credibility)

## End: Step 4 ##

## Step 5: Adjustment for shocked stable beta

adjustment_unstable <- merge(expected_assumptions, shocked_unstable_assumptions, by="betas")

adjustment_unstable$alpha <- 0.32

adjustment_unstable$confidence_alpha <- 1 - adjustment_unstable$alpha

adjustment_unstable$variance_ratio <-
  (adjustment_unstable$Std_Error_Beta_Shocked_Unstable^2)/(adjustment_unstable$Std_Error_Beta_Initial^2)

adjustment_unstable$alpha_credibility <- (1 - pnorm(adjustment_unstable$variance_ratio))*2

adjustment_unstable$confidence_credibility <- 1 - adjustment_unstable$alpha_credibility

adjustment_unstable$c1_alpha <- adjustment_unstable$alpha/2

adjustment_unstable$c2_alpha <- adjustment_unstable$c1_alpha + adjustment_unstable$confidence_alpha

adjustment_unstable$c1_credibility <- adjustment_unstable$alpha_credibility/2

adjustment_unstable$c2_credibility <- adjustment_unstable$c1_credibility + adjustment_unstable$confidence_credibility

adjustment_unstable$lower_alpha <- qnorm(adjustment_unstable$c1_alpha, mean=adjustment_unstable$Beta_Shocked_Unstable,
sd=adjustment_unstable$Std_Error_Beta_Shocked_Unstable)

adjustment_unstable$upper_alpha <- qnorm(adjustment_unstable$c2_alpha, mean=adjustment_unstable$Beta_Shocked_Unstable,
sd=adjustment_unstable$Std_Error_Beta_Shocked_Unstable)

adjustment_unstable$lower_credibility <- qnorm(adjustment_unstable$c1_credibility, mean=adjustment_unstable$Beta_Shocked_Unstable,
sd=adjustment_unstable$Std_Error_Beta_Shocked_Unstable)

adjustment_unstable$upper_credibility <- qnorm(adjustment_unstable$c2_credibility, mean=adjustment_unstable$Beta_Shocked_Unstable,
sd=adjustment_unstable$Std_Error_Beta_Shocked_Unstable)

adjustment_unstable$adjusted_beta_alpha <- ifelse(adjustment_unstable$Beta_Initial < adjustment_unstable$lower_alpha,
adjustment_unstable$lower_alpha, ifelse(adjustment_unstable$Beta_Initial > adjustment_unstable$upper_alpha,
adjustment_unstable$upper_alpha, adjustment_unstable$Beta_Initial))

adjustment_unstable$adjusted_beta_credibility <- ifelse(adjustment_unstable$Beta_Initial < adjustment_unstable$lower_credibility,
adjustment_unstable$lower_credibility, ifelse(adjustment_unstable$Beta_Initial > adjustment_unstable$upper_credibility,
adjustment_unstable$upper_credibility, adjustment_unstable$Beta_Initial))

adjustment_unstable$diff_adjusted_alpha <- abs(adjustment_unstable$Beta_Initial - adjustment_unstable$adjusted_beta_alpha)

adjustment_unstable$diff_adjusted_credibility <- abs(adjustment_unstable$Beta_Initial - adjustment_unstable$adjusted_beta_credibility)

## End: Step 5 ##

```