

Issues in Applying Credibility to Group Long-Term Disability Insurance

Sponsored by
Society of Actuaries Health Section

Prepared by
Paul L. Correia
Milliman, Inc.
September 2013



© 2013 Society of Actuaries, All Right Reserved

The opinions expressed and conclusions reached by the author are his own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information.

Table of Contents

| | |
|--|----|
| Table of Contents | 2 |
| Introduction | 3 |
| Practical Applications of Credibility Concepts in Group LTD | 5 |
| Challenges in Applying Credibility in LTD..... | 11 |
| Credibility Survey for Actuaries Practicing in LTD..... | 16 |
| Credibility Survey for State Regulators | 36 |
| Credibility Survey for Actuaries Practicing in Medical Insurance | 40 |
| Credibility Adjustment Factors for Medical Loss Ratio Refund Calculation | 45 |
| Actuarial Guidelines Pertaining to the Use of Credibility | 48 |
| Annotated Bibliography | 53 |

Introduction

The Society of Actuaries Health Section has engaged Milliman in a research project on issues in applying credibility in group long-term disability (LTD) insurance. Among the project's objectives are the following three key components:

1. Collecting information on challenges and current methods for applying credibility in LTD,
2. Researching published material and compiling an annotated bibliography of sources that discuss credibility in a way that is applicable to LTD, and
3. Researching and summarizing actuarial guidelines that pertain to the use of credibility in LTD.

The most well-known approaches for applying credibility have been derived from either classical or Bayesian paradigms. In the classical approach to credibility, observations are taken directly from experience data to determine the credibility of the experience, and all prior information is ignored. The resulting credibility estimates depend on the confidence factor and allowable error that define the underlying confidence interval. On the other hand, prior information does play a large role in determining credibility when Bayesian methods are used. Also, Bayesian credibility is dependent on conditional probabilities and subjective beliefs of posterior outcomes, unlike classical credibility.

Most of the applications of credibility in LTD are based on classical paradigms, and in particular the limited fluctuation model. Examples of some applications of limited fluctuation credibility are provided in the first section of this report, to give the reader background information on the use of credibility in LTD insurance.

The second section addresses some of the challenges in applying credibility in LTD. As part of our research, we performed a survey of credibility applications at 11 LTD insurers, all of whom rank among the top 15 writers of LTD insurance according to the 2012 Gen Re Group Disability and Group Term Life Market Survey. The survey asked carriers detailed questions about credibility applications in pricing and valuation, aimed at gathering information on challenges and current methods for applying credibility in LTD. A list of insurers that participated in the survey is given below:

| | |
|-------------------|-------------------|
| Assurant | Mutual of Omaha |
| Cigna | Principal |
| Guardian Life | Reliance Standard |
| The Hartford | Standard |
| Liberty Mutual | Unum |
| Lincoln Financial | |

Some of the survey questions were directed at credibility applications in group life and short-term disability insurance, to identify potential opportunities where those practices might be extended to LTD. We also surveyed state regulators for their perspective on credibility applications in LTD. Their responses provide unique and important perspectives on issues in applying credibility in LTD.

In our research of published literature on credibility, we identified 22 sources that discuss the topic in a way that is applicable to LTD. Although many more sources were reviewed in the process, we have excluded those that are not applicable to LTD. However, some of the papers that discuss credibility applications in Workers' Compensation insurance are included in the annotated bibliography because of the similarities between LTD and Workers Compensation claims.

Actuarial guidelines governing the use of credibility in LTD are summarized in the final section of this report. These include Actuarial Standard of Practice No. 25, the actuarial guidelines pertaining to the 2012 GLTD and 2005 GTLW valuation standards, and the state of Florida's criteria for applying credibility in LTD manual rate development and experience rating.

We would like to thank the following members of the Project Oversight Group for their support and contributions on this project:

| | |
|-----------------|-----------------|
| Warren Cohen | Jinn Lin |
| Thomas Corcoran | Barbara Scott |
| Brian Dunham | Sandee Schuster |
| Jennifer Fleck | Steven Siegel |
| Rick Leavitt | Bram Spector |

This report is intended for the benefit of the Society of Actuaries. Although we understand that this report will be made widely available to third parties, Milliman does not assume any duty or liability to such third parties with its work. Furthermore, this report should be distributed and reviewed only in its entirety.

In preparing this report, we have relied on information provided by actuaries who participated in our surveys, and on information from published research papers. To the extent that any of this information is incomplete or inaccurate, the results of our work may be materially affected.

Qualifications

I, Paul L. Correia, am a consulting actuary for Milliman, Inc. I am a member of the American Academy of Actuaries and meet its qualification standards for writing this report.

Practical Applications of Credibility Concepts in Group LTD

Credibility concepts are applied in the valuation of LTD claims and in the pricing of LTD products. The 2012 GLTD valuation standard uses a limited fluctuation credibility model for blending company specific experience with the valuation table. The pricing of LTD insurance products relies on applications of credibility in both experience rating and in manual rate development. These applications are discussed below.

Valuation

The 2012 GLTD valuation standard is a principle-based approach to reserving. The standard requires LTD insurers to either fully or partially reflect company-specific claim termination experience in their valuation assumptions, provided the experience is credible. Large LTD carriers whose claim termination experience is considered fully credible are required to calculate claim reserves using termination assumptions based entirely on company experience. Mid-sized carriers, whose termination experience does not meet full credibility standards, can partially reflect their own experience when the experience is at least partially credible. Insurers whose experience is not credible are expected to use 100% of the 2012 GLTD valuation table for calculating reserves.

The weight of company-specific experience is determined from a limited fluctuation credibility model. One of the underlying assumptions in limited fluctuation credibility theory is that the random variables (i.e. claim terminations) are independent. In reality, however, LTD claim terminations are not independent, and can be influenced by many factors like changes in claims management practices or changes in the economy. The non-independence of LTD claim terminations is thought to increase volatility and therefore reduce the credibility of claims experience. To address this issue, “selected variance factors” have been included in the limited fluctuation credibility model, aimed at adjusting credibility estimates for the non-independence of LTD claim terminations. These “selected variance factors” are shown below:

| LTD Claim Duration | Selected Variance Factor |
|-------------------------|--------------------------|
| 4 to 24 Months | 4.0 |
| 25 to 60 Months | 3.0 |
| 61 to 120 Months | 2.5 |
| Greater Than 120 Months | 2.0 |

The factors decrease with increasing claim duration because LTD terminations tend to be more volatile in early durations of claim. One reason for this is because claim terminations in early durations are dominated by recoveries which are believed to be highly correlated with cause of disability. Also, benefits from other sources and the change in the definition of disability usually occur in early durations

of claim, and these are believed to impact claim terminations as well. On the other hand, LTD claim terminations in the tail tend to be more stable, dominated by deaths as opposed to recoveries. The structure of the “selected variance factors” ensures that fewer terminations are required to be considered fully credible in later durations of claim, consistent with how volatility is believed to decrease with increasing duration.

Full credibility in the 2012 GLTD standard (i.e. the requirement to use 100% of company-specific experience in the valuation of LTD claims) is determined from the number of expected terminations that yields an 85% probability that observed terminations are within 5% of expected terminations, taking differences in volatility by duration into account. This confidence interval can be expressed symbolically as follows:

$$0.05 = 1.44 \times \sqrt{\frac{\text{Selected Variance Factor}}{\text{Number of Expected Terminations}}}$$

For example, the number of expected terminations that is required for considering LTD claims within durations 4 to 24 months fully credible is equal to $4 \times (1.44 / 0.05)^2 = 3,318$. However, only half as many terminations (i.e. 1,659) are required for considering claim experience in durations greater than 120 months fully credible, due to the “selected variance factors”.

According to the 2012 GLTD valuation standard, insurers whose experience is not considered fully credible are required to calculate a partial credibility factor for blending company-specific experience with the termination rates from the 2012 GLTD Valuation Table. The limited fluctuation credibility model used for calculating partial credibility is given below:

$$\text{Credibility Factor} = \text{Minimum of (100\% and } \left(\sqrt{\frac{\text{Number of Expected Terminations}}{\text{Number of Terminations Required for Full Credibility}}} \right))$$

For example, suppose an insurer expects 500 terminations from a block of claims in durations 4 to 24 months. Based on the 2012 GLTD valuation standard, this particular cohort (i.e. claims in durations 4 to 24 months) requires 3,318 expected terminations to be considered fully credible. From the equation above, the insurer can calculate a partial credibility factor equal to $\sqrt{500 / 3,318} = 0.39$, and therefore use a blended termination rate for calculating reserves which consists of 39% of company-specific experience and 61% of terminations from the 2012 GLTD Valuation Table.

Experience Rating

In experience rating, the weight given to historical experience is based on the assumed credibility of the experience. The most common formula for experience rating LTD products is given below:

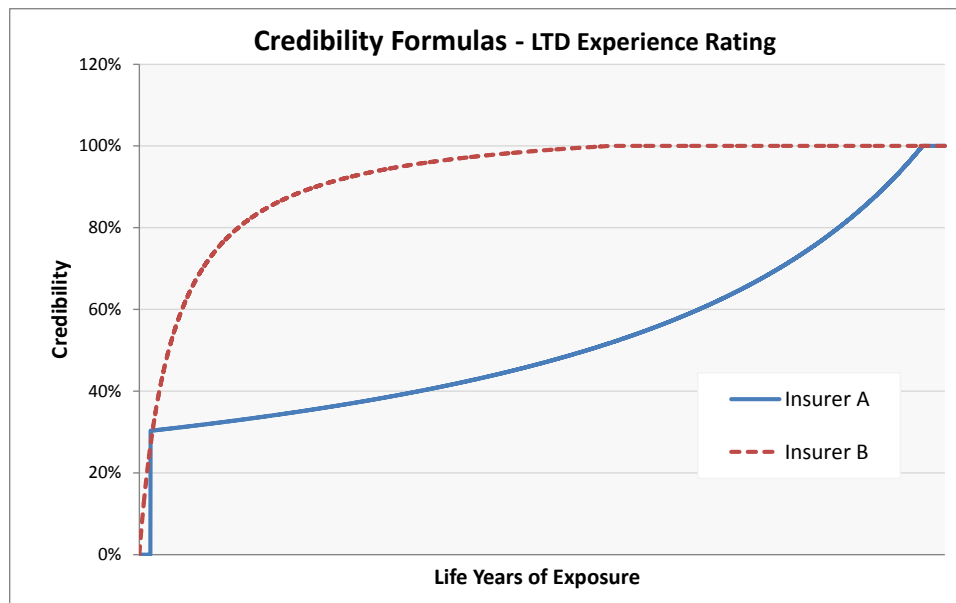
$$\text{Premium Rate} = (\text{Credibility Factor}) \times (\text{Experience Rate}) + (1 - \text{Credibility Factor}) \times (\text{Manual Rate})$$

According to this formula, when the experience is considered fully credible the premium rate is based entirely on the experience rate; when the experience is not credible at all the premium rate is based

entirely on the manual rate; and when the experience is considered partially credible the premium rate is a weighted average of the experience and manual rates.

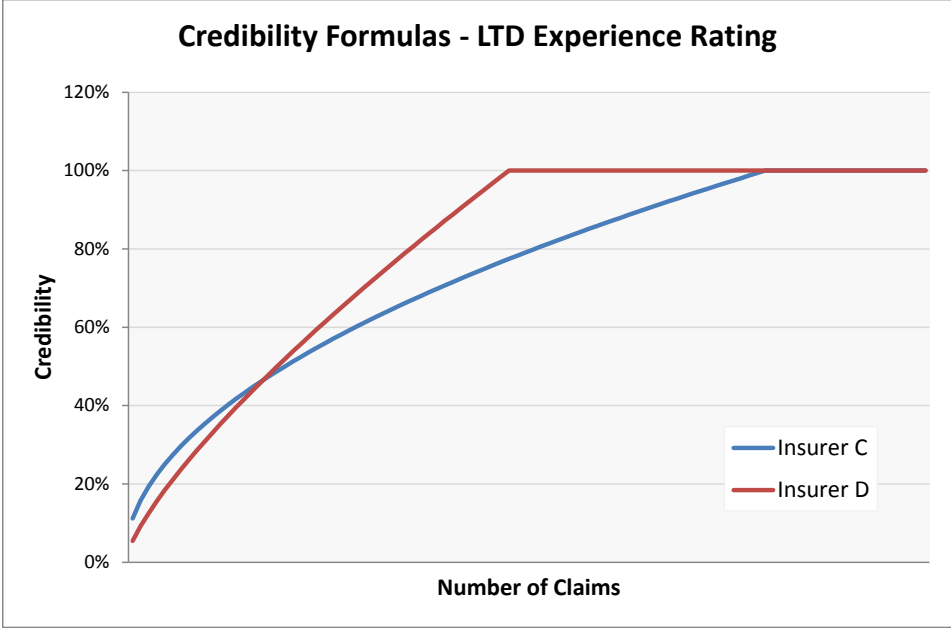
Credibility formulas used for experience rating in LTD vary significantly across the industry. Some of them are empirical, having been derived from experience studies using company-specific data, while others are based on more traditional approaches such as limited fluctuation credibility. The decision of whether to use an empirical model or a more traditional model may be influenced by a number of factors, such as limited claims experience, regulatory requirements, or simply because there are many factors that complicate the application of traditional credibility models in LTD (these are discussed in the next section).

Most credibility formulas used in experience rating take into consideration the number of life years of exposure in the historical experience period. However, the required number of life years of exposure for full credibility does vary from insurer to insurer. There are also differences in the shapes of the credibility curves used by different LTD insurers. These differences can be seen in the following graph which shows the credibility curves used by two different insurers for experience rating. The formulas that generate these curves were obtained through surveys of LTD insurers, conducted as part of this research project. A comprehensive discussion of survey results is provided in the third section of this report, starting on page 16.



One obvious difference between the two curves is that the credibility thresholds used by Insurer A are significantly less than those used by Insurer B. Another difference is that the credibility curve used by insurer A is concave up whereas the credibility curve used by Insurer B is concave down. Also, Insurer B's credibility formula does not require a minimum number of life years of exposure for the experience to be considered partially credible, whereas the formula used by Insurer A does. These differences point to very different pricing strategies between the two companies, possibly indicating a greater confidence in manual rates on the part of Insurer A.

Some of the credibility formulas used in experience rating are based on claim volume as opposed to (or in addition to) number of life years of exposure. One interesting observation is that some of these credibility formulas use *expected* claims whereas others use *actual* claims. The following graph compares credibility curves used by two different LTD insurers for experience rating, both of which are based on claim volume. The formulas that generate these curves were obtained through surveys of LTD insurers. As with the credibility curves based on life years of exposure, significant differences between the credibility formulas shown below are apparent.



Some LTD carriers use multiple credibility formulas in experience rating, for example to differentiate between voluntary and employer paid products, or to differentiate by geographic location. Also, some insurers have different formulas depending on whether the historical experience was favorable or unfavorable.

Manual Ratemaking

Actuaries analyze LTD experience by gender, age, industry, and other variables in order to calculate base rates and pricing factors for their LTD manuals. The choice of variables, and therefore the decision of how finely the experience should be divided, is a subjective choice that varies from one LTD insurer to another. To inform this decision, actuaries use concepts in credibility to determine if experience is suitable for calculating manuals.

There are many different ways to estimate the credibility of experience used for manual rate development. These methods range from subjective judgment to formal procedures using credibility formulas. When the method is strictly subjective, pricing actuaries make educated judgments to decide if the experience is credible enough for manual rate development. For example, if results are stable or constant across multiple time periods, or if results are consistent with expectations based on experience from other sources such as industry experience and/or professional judgment, then the experience may

be deemed credible. Thresholds could be incorporated to make the process more objective, for example by requiring a minimum number of policies or a minimum number of claims for the experience to be deemed credible. The advantages of using subjective methods include simplicity and flexibility. One obvious drawback is that it may be difficult to justify subjective decisions on credibility.

One example of a formal process that may be used to determine if LTD experience is credible for manual rate development is provided below. The example demonstrates applications of limited fluctuation credibility concepts. The procedure generally begins with a predetermined (and subjective) confidence interval. For example, a block of LTD experience could be considered fully credible for ratemaking if observed LTD claims are within 5% of expected claims 95% of the time. In mathematical notation:

$$\text{Probability } \{| \text{Observed Claims} - \text{Expected Claims} | < 0.05 \times \text{Expected Claims}\} = 0.95$$

Based on limited fluctuation credibility theory, this confidence interval can be translated into an equation for estimating the minimum number of claims required for the experience to be considered credible for manual rate development:

$$\lambda = (1.96^2 / 0.05^2) \times (1 + (\sigma / \mu)^2)$$

In the above equation, the value of λ is equal to the minimum number of claims required for full credibility, and the values of μ and σ are equal to the mean and standard deviation of the claim amounts.

To see how this formula can be used, consider the following table which shows values of μ and σ for three different blocks of LTD experience:

| LTD Block | Expected Claims (μ) | Standard Deviation (σ) |
|-----------|---------------------------|---------------------------------|
| 1 | \$40,000 | \$25,000 |
| 2 | \$40,000 | \$10,000 |
| 3 | \$100,000 | \$25,000 |

From the above equation, LTD Block 1 is considered fully credible for ratemaking if there are at least $(1.96^2 / 0.05^2) \times (1 + (\$25,000 / \$40,000)^2) = 2,137$ claims. LTD Blocks 2 and 3 are both considered fully credible at 1,633 claims because both blocks have the same ratio of σ to μ .

If the experience is not considered fully credible for ratemaking, then a decision needs to be made as to how much weight, if any, should be given to the experience. Continuing with limited fluctuation credibility, the weight could be determined as follows:

$$\text{Credibility Weight} = \text{Minimum of } (100\%) \text{ and } \left(\sqrt{\frac{\text{Number of Expected Claims in the Experience}}{\text{Number of Claims Required for Full Credibility}}} \right)$$

In the previous example, suppose that 500 claims are expected in LTD Block 1. Then the weight given to the experience would equal $\sqrt{500 / 2,137} = 0.483$.

Once the weight is determined, the partially credible experience can be combined with experience from similar risk classes, for example by using industry tables or by using experience from a reinsurer.

The advantages of using formal procedures for evaluating credibility are that these procedures are objective and theoretically justifiable. The disadvantages are that they may be difficult to implement, and they may not be applicable depending on the experience and the underlying credibility model.

Challenges in Applying Credibility in LTD

There are many factors that complicate the application of traditional credibility models in LTD. Some of the challenges in applying credibility stem from group dynamics, like having to deal with heterogeneous risk classes and non-independent exposures. Other issues in applying credibility are linked to the volatile nature of LTD benefit payment patterns, which can vary considerably in terms of duration and size. On top of everything, LTD insurance is sold in a very competitive marketplace and this makes it difficult to apply traditional credibility models in the experience rating of LTD products. Following is a discussion of these challenges.

Non-Independence of Claims

Most traditional credibility models, including limited fluctuation and least-squares models, operate under an assumption that exposures in the experience are independent random variables. However, LTD exposures, which are often measured in terms of claims, are not believed to be completely independent. For instance, external factors such as work conditions or the state of the economy can affect many if not all of the members of a group, and these factors are correlated with disability incidence.

Work conditions contribute, in part, to the health and well-being of employees. Physically demanding work conditions can actually trigger the onset of disability for several employees of the same group. For example, suppose a group of firefighters suffer injuries during an unfortunate accident while combatting a fire. The resulting disabilities are not completely independent since they were caused by the same event.

More generally, similar types of disabilities may be prevalent in some groups due to the nature of work. Back injuries are common in industries where lifting heavy objects is part of the work routine, such as construction work. Group-specific dynamics like these contribute to the non-independence of disability claims.

Economic factors also contribute to non-independence. Unfavorable trends in disability incidence often coincide with economic recessions. One possible explanation is that certain employees file claims during a recession (and would not file for claim otherwise) because they fear being laid-off and having a total loss of earnings. This is another example of why LTD claims are not completely independent: multiple claims can be motivated by a single event – in this case, a downturn in the economy.

The examples above illustrate non-independence of claim occurrences (i.e. incidence). Claim terminations are also believed to be non-independent events. Recovery trends are impacted by external factors like changes in claims management practices or changes in the economy. Spikes in recoveries are often observed when benefits administration resources are increased.

Interestingly, the jury is still out on how non-independence of LTD claims impacts credibility. On the one hand, one could argue that a concentration of non-independent risks can bolster the credibility of the experience. Consider again a group of construction workers. Disability trends among members of this group are likely to persist due to the nature of their work. Then the risk class of construction workers may be considered “credible” due to the individual risks being correlated.

On the other hand, one could argue that non-independence reduces credibility because it increases the volatility of the expected outcome. For example, claim terminations are affected by changes in claims management practices. If they were unaffected by changes in claims management practices then one could reasonably expect the experience to be less volatile. This was the approach that was taken in the development of the 2012 GLTD valuation standard.

The important point is that there are many ways to look at non-independence of LTD claims. Economic factors, employment characteristics, industry-specific dynamics, and claim management practices all contribute to non-independence in their own ways.

Heterogeneous Claims

Historical LTD claims are often used to estimate future claims by assuming the experience is homogeneous. The underlying assumption is that claims experience will emerge similarly as it had in the past. There are a number of reasons why this assumption may not be true:

- Changes in the demographic mix of employees over time can alter disability experience trends
- External factors like economic recessions can impact the experience
- Changes in underwriting or claim management practices can shift the experience
- Changes in plan design may result in different claims experience

Some ways to address these issues include factoring in potential demographic changes in the employee population, or constant-rating historical premium in the experience rating calculation when there is a change in plan design. These types of adjustments are not trivial.

Heterogeneous risk characteristics of the individual employees in a group can also complicate the application of credibility in LTD. Experience rating relies on using past experience to predict future experience. On the one hand, groups with more homogenous demographics could be viewed as being more credible than groups with a greater demographic mix of employees, due in part to their uniform risk characteristics. On the other hand, groups with a greater demographic mix of employees may be considered more credible because there is greater spread of risk, presumably causing lighter swings in the experience when compared to groups with more homogenous demographics.

Credibility models used for experience rating that are based on expected claim volumes may be able to capture some of these group dynamics. However, credibility models that are based solely on life years of exposure may not be capable of dealing with these issues.

Competitive Pricing Pressures

Competitive pricing pressures make it challenging to apply credibility in LTD pricing. LTD carriers operate in a highly competitive market. There is pressure to give the experience more credibility than what would be prescribed if a purely theoretical approach were taken, because it is difficult to pursue rate increases on cases that have had good experience in a competitive market environment.

Claim Duration

LTD claim durations can range from one year to several decades long, depending on diagnosis, definition of disability, limitations, and many other factors. This in itself creates challenges with applying credibility in LTD. On top of this, LTD claim experience tends to be more volatile in the early durations of claim for several reasons:

- LTD claim terminations in early durations are dominated by recoveries. There is a strong correlation between recoveries and cause of disability in early durations of claim, resulting in recovery patterns in early durations that can vary significantly by cause of disability. Furthermore, LTD recoveries are impacted more by claim management practices than deaths, resulting in termination experience in early durations that may differ from insurer to insurer.
- Benefits from other sources are typically awarded within the first few years of claim, creating irregular payment streams in early durations.
- The change in definition from an “own occupation” to an “any gainful occupation” definition usually occurs within the first few years of claim, resulting in a spike in recoveries at the change in definition.
- The maximum benefit period for mental & nervous claims is usually limited to 24 months.

Claim terminations and benefit payments are more stable in the later durations of claim, which results in reduced volatility.

These dynamics make it difficult to apply credibility in LTD. Some credibility models address the issue by using factors that explicitly lower the number of exposures required for full credibility in later durations of claim, like the limited fluctuation model in the 2012 GLTD valuation standard, for example.

Benefits from Other Sources

LTD benefit payment streams are irregular because monthly benefit amounts typically change over time. The approval of SSDI benefits or the loss of Workers' Compensation benefits can significantly impact net benefit amounts. The resulting variable payment stream adds yet another layer of complexity to the credibility framework, because the expected severity of claims is volatile.

Another connection between benefits from other sources and credibility can be described in terms of claim termination rates. Let's assume that claimants who receive SSDI awards are more morbid than claimants who are not awarded SSDI benefits, because the SSDI definition of disability is more stringent than most contractual definitions in LTD. Then the termination experience for claims with SSDI benefits could be significantly different than the experience from claims that have not been awarded SSDI benefits. With this being said, the credibility of a block of LTD experience could depend in part on how many claims in the experience are SSDI recipients.

Outlier Claims

Outlier claims may complicate the application of credibility when the underlying credibility formula is based on claim amounts, by artificially bolstering the credibility of experience that includes one or more very large outlier claims. These issues raise the question of how claim amounts should be handled in determining credibility.

Even when the formula is not a function of claim amounts, outlier claims can complicate applications of credibility. In the pricing of LTD products, very large claims can shift the experience rate by a significant amount. Depending on the credibility of the experience, these outlier claims can create serious pricing challenges. The reason why these types of claims are challenging to work with is that the statistical probability of a claim of the same magnitude happening again could be very small, because maybe only a fraction of the group are highly paid individuals, but nonetheless these claims do happen.

Very large outlier claims can be dealt with in a number of ways in experience rating. They can either be left in the experience or completely removed from the experience. These two approaches are somewhat subjective, but the decision of whether to remove large claims from the experience or leave them in can be supported by concrete analysis, such as thorough analyses of historical experience and of future risk exposure. A key decision, however, is whether or not to adjust the credibility of experience that includes or excludes outlier claims. By definition, outliers are unusual occurrences, so perhaps the experience that includes outliers should be considered less credible.

Another way of dealing with outlier claims is to implement risk pooling mechanisms in the experience rating framework. Risk pooling typically involves simultaneously capping claims in the experience that are above a predetermined threshold, and loading the rate in proportion to the exposure to risk. This process could include adapting the credibility formula to take pooling into account. In a paper called "Some Applications of Credibility Theory to Group Insurance" by Charles Fuhrer (included in the annotated bibliography), the author demonstrates how to synchronize credibility models with pooling mechanisms using least-squares criteria.

Regulatory Requirements

Some states have adopted credibility requirements as part of the supervisory process. These requirements may apply to manual rate development, for example modifying existing pricing factors, or to experience rating of LTD products. It can be a challenge for some carriers to meet the required credibility criteria. For example, in rate filings that involve changing a pricing factor for a provision that has only modest historical experience, there may not be enough experience to support a rate change based solely on the company's own data. These situations can be challenging from both the regulatory agencies and the insurance carriers' perspectives, and may require companies to justify rate changes by other means if historical experience is not credible.

Estimating Parameters

Estimating the parameters of a credibility model (like confidence intervals and full credibility thresholds) is often based on a combination of subjective opinion and empirical testing. When the credibility is based on an underlying confidence interval, then the confidence factor (e.g. 85%) and allowable error (e.g. 5%) are usually determined subjectively. The choice of confidence factor and allowable error may depend on several different factors, like the credibility application (manual rate development versus experience rating versus valuation), or the quality of the experience data.

Empirical methods are typically employed for estimating credibility thresholds and determining the factors that can influence the credibility of LTD experience. Predictive modeling can be used to identify the key drivers of experience, which can then be incorporated into the credibility framework. These approaches can be challenging to implement.

Credibility Survey for Actuaries Practicing in LTD

Summary of Responses to Survey Questions for Actuaries Practicing in LTD

Detailed survey results are provided in the following section. The responses are intended to develop information that could be used to support more effective use of credibility in group LTD applications. To this end, we have included responses directed at group products other than LTD, namely group life and short-term disability. These are intended to identify potential opportunities where those practices might be extended to LTD.

The following is a brief summary of the survey responses pertaining to LTD:

- Life years of exposure are used by most carriers in credibility formulas for experience rating. When claims are used, the formulas are fairly evenly split between actual and expected claims.
- There is a wide range of minimum and full credibility thresholds used for experience rating.
- A few companies have formal methods for dealing with outliers in experience rating, but most deal with outliers on a case-by-case basis.
- Most carriers responded that underwriters can modify the credibility of LTD case experience. Some of the companies have guidelines for how underwriters can affect the credibility estimate.
- The responses are fairly evenly split between companies that have tested their credibility formulas and companies that have not done any testing.
- The responses are fairly evenly split between companies in which the quality of claim data affects credibility in experience rating applications and companies in which it does not.
- The responses are fairly evenly split between companies that employ formal credibility processes in calculating manual rates and companies in which no formal process exists.
- Most carriers have not had issues in justifying credibility methods for rate filings.
- Many companies have not had the opportunity to implement the 2012 GLTD valuation standard. Those that have begun to use it have not had any issues with the prescribed limited fluctuation credibility model.
- Many respondents commented that they would like to see an industry standard for applying credibility in LTD.

- Some respondents commented on the problems in applying the same formula to every situation.
- Some respondents commented on the challenges created by the interaction of frequency and severity components of LTD claims.

Credibility Survey for Actuaries Practicing in LTD Insurance

Part I: Case Rating

- (a) What factors are considered in your credibility formulas for experience rating new or renewal cases? (check all that apply)

The numerical values in the table represent the number of affirmative responses. For example, from the table, 2 carriers responded that they factor lives into their LTD credibility formula.

| | LTD | STD | Group Life |
|------------------------|--------------------------|-----|--|
| Lives | 2 | 2 | 1 |
| Life Years of Exposure | 10 | 9 | 10 |
| Actual Claims | 3 | 1 | 2 |
| Expected Claims | 4 | 2 | 2 |
| Premium | | | |
| Elimination Period | 4 | 7 | |
| Benefit Period | | | |
| Demographic Mix | 1 (average age) | | 1 |
| Type of Product | 1 | | 1 (basic vs. additional) |
| Occupation Class | 1 | | |
| Industry | | | |
| Diagnosis | | | |
| Other (specify) | 1 (months of experience) | | 1 (months of experience) 1 (Average certificate, maximum certificate, pooling level and number of months of experience) |

- (b) Please describe your formulas in as much detail as possible, including at a minimum the underlying theory, basic structure, minimum credibility level and full credibility threshold.

| LTD minimum Credibility Level | Number of LTD Carriers |
|----------------------------------|------------------------|
| 0 lives | 3 |
| 1 – 249 lives | 1 |
| 250 – 499 lives | 2 |
| 500+ lives | 3 |
| 1 – 249 life years of exposure | |
| 250 – 499 life years of exposure | 1 |
| 500+ life years of exposure | 1 |

| LTD Full Credibility Threshold | Number of LTD Carriers |
|--|------------------------|
| 1 – 24,999 life years of exposure | 6 |
| 25,000 – 34,999 life years of exposure | 1 |
| 35,000+ life years of exposure | 1 |
| 1 – 99 claims | |
| 100+ claims | 1 |
| Varies by experience | 2 |

Comments pertaining to LTD:

- Items such as plan design, demographics, or industry may be considered by the underwriter but no formal practice for implementing these factors into the credibility formula exists.
- We use a 2D table, with credibility factors being dependent on life years of exposure and number of months of experience. Greater credibility is assigned for a greater number of months of exposure, assuming the same number of life years. The formula is biased towards simplicity by design, although we do recognize that other factors such as elimination period can affect credibility.
We struggled with the limited fluctuation model. We looked at our business mix in terms of broad case size buckets for cases that had been in force for four years, and studied how two years of experience correlated with two years of subsequent experience. We calculated the level of credibility which would minimize the error between the two-year periods, keeping in mind market considerations. This was done by calculating correlation factors and then fitting a curve through the correlation points, which was the basis for generating a table of credibility factors. We feel that the market is closer to the appropriate level of credibility than some actuaries give credit. For STD, there is market pressure to assign higher credibility (maybe too much).
- Credibility is based on a weighted average of actual and expected number of claims. While not explicitly part of the credibility formula, other factors (e.g. EP, BP, demographic mix, type of product, occupation class, industry) have an indirect impact.
- We use a table of values derived from a formula. Theory suggests that the credibility is significantly lower than what is produced by our formulas. In our actual formula we include industry considerations.
- LTD credibility is based on a table with incidence rates that vary by qualifying period. The lower the qualifying period, the higher the incidence rate.
- Many factors get into expected claims: EP, demographic mix, product, industry.
The structure of the credibility formula reflects an application that has been around for many years and is believed to be common across the industry.
The LTD credibility formula reflects practical considerations and has been tested.
- We have two formulas – one is the primary formula for calculating credibility and is based on life years of exposure. The second is based on claims and is only used for benchmarking against the primary formula.

- In the formula for calculating expected claims, there are adjustments for elimination period, benefit period, product type, demographics, occupation and industry. So these factors make it into the credibility formula indirectly.
- Elimination period and demographic mix are implicitly part of the credibility formula through the claims experience. Certain demographics have more claims and are therefore considered to have more credible experience. Credibility is dampened if the relationship between expected severity from manual and actual severity is off (i.e. lower credibility is assigned when actual severity deviates from the expected severity, in effect increasing the credibility threshold based on variance of severity). The initial credibility factor is based on frequency (incidence), then we factor in the claim severity component. We treat frequency and severity separately in the formula.

Comments pertaining to group life and STD:

- STD Full credibility at 750 LYE
GL full credibility at 25,000 LYE (basic) / 35,000 LYE (supplemental)
- Simplified incidence based experience rating formulas and credibility are used for prospects with 250 – 999 lives that meet certain criteria. Otherwise, standard experience rating formula and credibility are used.
- Note that for STD, expected number of claims is based on elimination period and lives, so these make it into the credibility formula indirectly.
- Credibility in STD starts at 100 life years, 400 life years is fully credible (although it varies by EP). Credibility in group life generally start at 500 life years of exposure, 18,000-20,000 is fully credible.
- In group life, cases with fewer than 500 lives get no credibility. In STD, cases with fewer than 100 lives get no credibility.
- Experience rating starts at 300 lives for group life and 100 lives for STD. For group life, at 30 months of experience, full credibility is reached at 25,000 life years of exposure. At 12 months of experience, 25,000 life years of exposure are given 54% credibility. STD experience is considered fully credible at 530 life years.
- For STD, credibility is based on elimination period. Minimum = 100 life years. Full credibility depends on elimination period. For group life, the minimum requirement is 500 lives. Technically, you cannot reach full credibility.
- For STD, the minimum level is 100 life years and full credibility is reached at (1) 600 life years for ≤ 8 day EP and (2) 750 for > 8 day EP. We use a modified square root formula for interpolation. For group life, the minimum level is 300 life years and full credibility is reached at 21,000 life years. We use a modified square root formula for interpolation.

2. What is the maximum time period for capturing life years of exposure? What is the minimum time period?

| Carrier | LTD Minimum | LTD Maximum |
|---------|---|--|
| A | 2 years | 5 years |
| B | 1 year | 5 years |
| C | 1 year | 5 years |
| D | 1 year | 5 years |
| E | NA | 5 years |
| F | 2 years | 5 years |
| G | 2 years: 200-499 lives 1 year: 500-999 lives 3 years: 1000+ lives | 3 years: 200-499 lives 5 years: 500-999 lives 5 years: 1000+ lives |
| H | 1 year | 5 years for prospects 6 years for renewals |
| I | No minimum | Up to 5 years for prospects Up to 5 years depending on time in force for renewals |
| J | No minimum | Typically 4 years, but UWs can use more years |
| K | 3 Years | 5 Years |

- In practice, underwriters will use their judgment

3. (a) How do experience outliers affect credibility? For example, in the pricing of LTD insurance, the presence of large claims in the experience period can swing the experience rate by a significant amount. How are these types of situations dealt with?

Note that some respondents chose more than one option in the table below.

| | LTD | STD | Group Life |
|--|-----|-----|------------|
| Outliers are left in the experience and the credibility is unaffected | 2 | 7 | 3 |
| Outliers are left in the experience and the credibility is reduced | 1 | | |
| Outliers are removed from the experience and the credibility is unaffected | 4 | 3 | 5 |
| Outliers are removed from the experience and the credibility is reduced | 1 | | |
| Pooling points (e.g. floors and ceilings on the experience rate) are used and the credibility is a function of pooled claims | | | |
| Pooling points are used and the credibility is not a function of pooled claims | 4 | | 4 |

(b) Are such adjustments formal, informal, or a combination of both? If pooling is used, do you use a formal or informal pooling charge approach?

- Formal: 3
- Informal: 8

Additional comments pertaining to LTD:

- For LTD, we use pooling points and a pooling charge (formal process). Separately, we also have floors and ceilings, which are a ratio of the manual rate. The pooling charge is based on a typical group, though it varies by case size range and maximum range. We take out the excess over the pooling point.
- Pooling is based on the manual claim cost algorithm. If we have claim-level detail, we will remove the excess reserve above the pooling point and replace with a pooling charge.
- This is up to UW discretion, we have no formal process. For example, the UW may remove a particular year of experience. There would be no change to the credibility for such adjustments (except that there would be fewer life years in the experience).
- Right now, the system doesn't automatically pool the claims, UWs use their judgment. Our goal is to have the system do formal pooling. We are still trying to determine the threshold. In practice, underwriters may modify using underwriting judgment to make informed decisions to remove outliers, pool, or modify credibility.
- The credibility and formula rates produced on experience rated cases do not adjust for the impact of outliers. However, underwriter discretion is allowed to adjust proposed rates

based on the presence of outlier claims. We have a small case focus, so only a few cases are experience-rated. We do not use any pooling methods for claims on experience rated cases. Underwriters can use judgment to make informal adjustments on cases with outlier claims.

- There is no formal protocol for dealing with outliers in LTD. Outliers are handled on a case-by-case basis by underwriters and pricing actuaries. The starting point is to leave them in the experience with no change to credibility. We currently do not have any pooling methods to address this situation.

On the flip side, what if there are NO claims in the experience period? For example, a law firm with no claims – how do you deal with that? Outliers go both ways.

- A consistent approach for LTD does not exist for removing outliers from the experience. Sometimes large LTD claims are removed, sometimes they're not.
- Underwriters have the option to leave outliers in or remove them. The decision whether or not to leave them in is based on extensive underwriting. For example, if they know a claim is closing within the next couple of months then maybe they'll remove the claim.
- Pooling in LTD is formula driven.
- All adjustments are informal. This is the art of underwriting.

Additional comments pertaining to group life and STD:

- Pooling in group life is formula driven.
- Underwriters have the option to leave outliers in or remove them.
- In group life, a tool is used to analyze the census and calculate pooling points based on expected claims. The tool also provides underwriters with the charge to add back into the pricing.
- We have a formal pooling process for group life experience rating. A pooling level is based on average certificate and case size. The pooling charge percent is calculated as part of total expected claims that are on volumes in excess of the pooling level. Pooled claims equal the amount of paid, pending or pro-rated waiver claim reserve amounts on individual claimants that exceed the pooling level.
- The adjustments for STD and group life are formal. For group life, the system determines the pooling level based on volume.
- This is up to underwriter discretion, we have no formal process. For example, the underwriter may remove a particular year of experience. There would be no change to the credibility for such adjustments (except that there would be fewer life years in the experience).
- In group life, pooling is based on the manual claim cost algorithm. If we have claim-level detail, we will remove the claims above the pooling point and replace with a pooling charge. For STD, there is no adjustment.
- For STD and GL, our processes are informal. We remove large claims (they are either left out or normalized). This is done at the underwriter's prerogative. Group life is too complicated to set a pooling point.

4. (a) Are underwriters allowed to modify the formula-based credibility of a block of experience?

| | LTD | STD | Group Life |
|-----|-----|-----|------------|
| Yes | 8 | 7 | 6 |
| No | 3 | 4 | 5 |

(b) If yes, please describe, in as much detail as possible, how underwriters may affect the calculation of credibility.

- Underwriters prefer to increase credibility. We have a points system looking at case characteristics which centers around stability. For example, we give points for the quality of the data or more detailed data. These points can as much as double the table credibility, though the head underwriter can exceed this limit. The underwriter can also ignore the floor and ceiling.
- We have documented guidance for UW staff regarding experience adjustments. Underwriters will adjust the credibility for a case in some situations, such as when the experience is not stable (i.e. exposure of lives or loss ratio changes significantly during the experience period), if there is poor data or an incomplete or inaccurate census, or if there was a catastrophic event. The underwriters will use their judgment to decide the adjustment. They may allow more than 5 years of data to be included in the experience if it is consistent. At certain premium thresholds, any adjustments would be reviewed with the actuaries. Generally UWs would check in with actuaries whenever adjustments are made. Another potential adjustment would be to reweight the experience periods. When such reweighting is used, there is an automatic reduction in the credibility.
- There is no formal guidance, generally if experience is significantly different from the baseline the UW may choose to rely solely on one or the other.
- The underwriter can modify credibility for a plan, a combined case where plans are legally tied, or a broker block. For example, if there are 5 years of experience, the UW can remove a particularly bad year. UWs can also change the credibility itself. I am not sure how often they make such changes.
- UWs do not adjust credibility in the formulaic rate for a case. Underwriters are allowed to adjust credibility in determining the “underwriting risk rate.” There aren’t rules on how to adjust the credibility, but if the “underwriting risk rate” diverges from the formula-based rate significantly, then they need pricing approval.
- Underwriters cannot change the credibility formula, however they are allowed discretion for the proposed rating action.
- There are guidelines for how underwriters can affect the credibility calculation for LTD. They will look at (1) movement in LYE across the experience period and (2) claims distributions across the experience period to inform their decision on whether or not to change the credibility of a case. Underwriters are limited to making a 15% change in the credibility value (both ways...i.e. +15% and -15%).
- Underwriters can remove some claims from the experience, but they cannot touch the credibility formula.

- Underwriters will use lower levels of credibility if the data is unreliable. No formal, mechanical process exists for adjusting credibility.

5. How often are your credibility formulas updated?

| | LTD | STD | Group Life |
|-----------------------------|-----|-----|------------|
| Annually or More Often | | | |
| Every 2 – 5 Years | 5 | 4 | 4 |
| Every 5 Years or Less Often | 6 | 7 | 7 |

6. Have you done any testing of your credibility formulas and methods? If yes please describe (in as much detail as possible).

- Yes: 6
- No 5

[Description]

- By analyzing loss ratio variance by case size and across differing experience periods
- We looked at large cases in force for four years and studied how two years of experience correlated with two years of subsequent experience. We calculated the level of credibility which would minimize the error between the two-year periods. We also factored in market considerations.
- We have not yet done any formal testing, but we have started forming a plan for testing this year. We plan to look at predicted costs vs. actual costs for each case.
- We have run a Monte Carlo simulation (using assumed claim incidence rates to see how many lives we need to get credible experience). This testing stemmed from FL objections. When we do manual rate reviews, we assume more conservative credibility.
- We have not tested our credibility formulas recently. A consultant did some testing years ago. $A \times (\text{first year}) + B \times (\text{second year}) + (1 - A - B) \times \text{manual} = \text{third year actual}$ then solve for A and B. We plan to test Life this year.
- We used historical company data to evaluate the credibility approaches which, when used to combine historical experience and manual rates, provide the best fit for subsequent observed experience. This was considered to be an appropriate approach to testing the strength of alternative credibility formulae, since in practice credibility is used to balance experience and manual rates.

The implication is that credibility should not only be a function of the predictive quality of the experience rate, but also a function of the quality and fit of the manual rate. Our findings from the study suggested that our prior credibility formula, which we believe was in-line with industry averages, generally assigned too much credibility to cases. This was particularly the case for groups having lower claims volumes.

The testing involved using 5+ years of company experience, using the first 3 years to see how well the formulas predict the next 2 years. Both the structure and the levels were tested. For example, prior formula used expected claims only, but testing revealed that MAX(expected claims, actual claims) was a better predictor.

- No testing done in the past, however we have been leveraging predictive modeling resources in an effort to change the way we think about credibility in LTD. The predictive modeling serves as a way to understand which items are most predictive in determining credibility.
 - Normal peer review. Large case experience provides good case studies. For life, we performed a statistical analysis on our block of business using member level data with various benefit levels to determine full credibility. Results were around 2 million life years.
7. (a) Does the quality of the claim data affect credibility? For example, if LTD claim data is missing offset information, then is the data less credible than if offset information was known?

| | LTD | STD | Group Life |
|-----|-----|-----|------------|
| Yes | 5 | 4 | 5 |
| No | 6 | 7 | 6 |

(b) If yes, please describe, in as much detail as possible, how your credibility formulas are affected by considerations of data quality.

- We use minimum data requirements to make sure there is a certain level of data quality. The underwriter can also add points for good data.
- This is handled by underwriter judgment
- This is theoretically left to UW discretion (cannot think of a real example where this happened).
- Underwriters look at all experience information to see if it's within expected levels. If it's not, they will use judgment to limit credibility (informal).
- In the reserving process, best estimate assumptions are used to fill the gaps, but these don't make it into the credibility formula.

Five years ago, a study was performed to see how data quality affects predictive capacity. What was found was that missing information does not really affect the accuracy of the prediction. In fact, one could argue for a claim count based credibility formula, because for mid-sized cases more detail does not necessarily improve estimates of future experience.

- There are safe assumptions in the reserve algorithm to fill gaps in data. The quality of the data may play into how UWs view the experience, and sometimes they will take a more conservative stance in the pricing if the experience data is poor.
- Data quality issues do not affect the credibility calculation. In the reserves, there are formulas that adjust the experience based on best estimates when information is missing.
- Underwriters are allowed to adjust credibility subjectively. The challenge is how to quantify every situation that arises in which data quality is relevant.

- In group life, sometimes data quality issues are taken into account in estimating credibility, but to a lesser extent than LTD.

8. Have you had any issues with filing your credibility methods for experience rating in any states? If yes, please describe.

- Yes: 5
- No: 6

- Yes, some states have asked about credibility and justification for the credibility that is used. Some states also have very conservative credibility standards (i.e. provides very low credibility) that they commonly reference.
- Yes, we have had problems in the past. We handled the issues by filing separate forms for groups of 50+ lives, for which we don't file the rating details.
- Some states use a more conservative, theoretical-based credibility formula than what is generally used in practice – we have to comply.
- In some states, we have been told that we had not adequately supported our credibility formula for experience rating. We called the state agencies and ultimately sent a sales representative to meet with them face-to-face. Finally, we filed two contracts: (1) cases with fewer than 50 lives and (2) cases with 50 lives or more.
- One of the biggest challenges is that there is the hypothetical (i.e. statistical) measure of credibility and then there's the practical measure. That being said, we would like to have a better understanding of regulators' true concerns. Are regulators concerned that the formulas are not actuarially sound? How does so much drilling into credibility affect the policyholder? If anything, there is a bigger risk of undercharging as opposed to overcharging.
- We've had issues filing in some states in the past. We ended up working with a consulting group and eventually split the policy forms between small case and large case, and this helped resolve the issues. Sometimes, regulators drill into certain items and sometimes they don't, and it seems like an inconsistent process. As a result, sometimes you can fly through a rate filing and other times it seems like everything gets challenged.
- Don't recall having any issues. Some states tend to ask more detailed questions on experience rating.
- Some states mandate their own credibility tables.

Part II: Manual Rates and Pricing Factors

9. (a) Do you employ credibility techniques in setting manual rates (explicitly or implicitly)?
 Note that some respondents chose more than one option in the table below.

| | LTD | STD | Group Life |
|--|-----|-----|------------|
| Using credibility formulas or tables | 5 | 3 | 3 |
| Using credibility thresholds (specify) | 1 | 1 | 1 |
| No formal process exists / educated judgment | 6 | 8 | 7 |

(b) Please describe your formulas in as much detail as possible, including at a minimum the underlying theory, basic structure, minimum credibility level and full credibility threshold. (Please note that these formulas will be used to draw broad conclusions regarding current industry practices. The details of company-specific formulas will not be shared outside of the research team.)

- Very crude credibility formulas based on claim counts have been applied to highlight areas where experience might differ from current rate expectations. This highlighting is followed by the application of an educated judgment which may or may not result in a rate change. Rate changes may also be executed in non-highlighted cells
- We use credibility only for industry, occupation, salary and area factors. We use a limited fluctuation (i.e. classical) credibility model. We determine credibility levels for our experience and market experience based on adjusted lives exposure. Market exposure is determined from survey data and BLS statistics. The exposure is adjusted to give less weight to larger employer groups.
 We set up α and a confidence interval. The formula is biased to give the company's own experience more credibility.
 Once company and market credibility are determined they are adjusted in proportion to sum to 100%.
 The starting point is (company credibility) x (company experience indicated change) + (market credibility) x (market factors indicated change). The resulting factors are smoothed and adjusted using judgment.
 The market factors used are based primarily on rate filings with adjustments for outliers.
- We look at the number of policies in each pricing cell and use judgment to determine the level of credibility. For example, if there are less than 100 policies, we would deem the information non-credible. One thing we might do, for example with SIC, is to broaden the category and use the data at a higher level.
- The formulas used for manual rating are significantly more conservative than those used for experience rating. They are closer to theoretical credibility, with a tighter confidence interval. LTD is considered fully credible at 2,000,000 life years

Credibility is used for calculating industry factors, area factors, case size factors, EP factors, but not for calculating base rates.

- LTD is considered fully credible when expected claims = \$100,000,000
- We changed LTD rates in 2011, and switched our credibility to be based on claim counts and incurred claims metrics. Previously, we looked at total incurred claims only. For the claim count metric, we use only non-maternity claims. The full credibility threshold is 2,000 claims in 3 years. For incurred claims credibility formula, our threshold is \$22 million of incurred claims for 50% credibility. This ratio is applied to the actual to expected ratio for incurred claims to blend each cell with the same ratio for all business.
Credibility is used for everything now, even basic claim cost.
- We create segments that are credible (based on actuaries' opinions), we are conscious of over-segmentation, and we track claim counts in the cells used for ratemaking.
For some plan provisions, there's simply not enough historical claim experience to be considered credible, so other methods are applied. For example, interpolating/extrapolating, or using the experience from similar provisions.
- Credibility is determined subjectively, with the recognition that if there is a lot of business underlying the experience, then the experience is generally considered to be fully credible. We will apply subjective weighting of credibility in certain situations. In other words, a formula may not be used to determine the partial credibility of sparse experience, but we do recognize when business is not credible and will apply weightings accordingly.
- We utilize limited fluctuation credibility in determining the credibility of data used in ratemaking. The confidence intervals are based on parameters of 95% and 10%.
We look at claim amounts. Partial credibility is assigned based on the limited fluctuation square root rule. We try not to make buckets too small, and it is usually a subjective decision whether or not partially credible experience is credible enough. For example, even if the experience is only 20% credible, if it's really bad then the experience will probably get used.
We also look at other competitors' pricing factors for benchmarking.
We look at sales versus target to inform decisions in ratemaking.
Basically, ratemaking takes into account a lot of different things.
- Certain cells are believed to be very credible. There are other cells where we know the experience is not credible (e.g. industry) and so we look to outside sources for rate relativities. When manual rates are calculated from first principles, they are based, in part, on claim termination ratios, which are more credible in early durations than in later durations of claim (because there are more observations). So credibility does come into play, and will vary with claim duration.
- Experience studies contain 95% confidence interval, not a credibility measure.

(c) If you use credibility, what is your baseline?

- The starting point is (company credibility) x (company experience) + (market credibility) x (market factors), then the resulting factors are smoothed and adjusted for judgment. For example, if company credibility = 50% and market credibility = 25%, then we use 66.6% and 33.3% for company credibility and market credibility, respectively.
- We might use other companies' rate manuals, and also work with our reinsurance partner.
- For industry factors, we blend our experience with the results for the total block. We use linear interpolation.
- Blend with nationwide total experience
- Either the current factor (use credibility to decide whether to make a change), or external sources such as public rate filings and reinsurer input.

10. Have you had any issues with filing your credibility methods for manual rate development in any states? If yes, please describe.

- Yes: 2
- No: 9

- Yes, in some states the assumption is that currently filed rates are fully credible and that any change to filed rates must be justified with credible experience. Lacking credible experience in a cell results in existing rates remaining in effect, regardless of the basis or credibility of the originally filed rates. Some states have also adopted a very conservative credibility standard (i.e. provides very low credibility) that they commonly reference.
- Sometimes with area factors we need to show the particular state's area factors and justify not giving those factors full credibility. We have been successful in providing these justifications and getting our filings approved.
- Some states use a conservative method – we still have problems with these states.

Other issues:

1. If loss ratios in are favorable, they want their rates reduced. They don't care about credibility on a nationwide basis. They don't have specific credibility methods specified in their regulations, so we can't cite such methods.
2. In our manual rate development, we exclude groups over 1,000 lives. Some states have a problem with that. We feel that experience-rated groups should not affect the manual rates.

Some states do not want any credibility applied. In other words, they're basically saying that "Nationwide experience is fine, so don't use state specific experience even if it's been poor." Regulators seem to have been focusing more on credibility in the last couple of years.

- Questions on credibility are standard from regulators in certain states. They have become more practical, however, and less mechanical.

Part III: Valuation

11. (a) How do you use credibility in the setting of morbidity assumptions for experience that gets used for LTD claim reserves?

| | LTD | STD | Group Life |
|--|-----|-----|------------|
| Using credibility formulas or tables | 3 | | 1 |
| Using credibility thresholds (specify) | 2 | | |
| No formal process exists / educated judgment | 7 | 6 | 5 |

(b) Please describe your formulas in as much detail as possible, including at a minimum the underlying theory, basic structure, minimum credibility level and full credibility threshold. We understand that these may vary by reserve basis (GAAP vs. Stat) and by claim duration.

- We do not have any explicit credibility-based reserving formulas or methods
- We are now updating our methods and implementing the structure described for the 2012 GLTD table. We have modified the method to give full credibility to termination experience within the change in definition transition period. We assume our experience is fully credible through five years of claim.
- GAAP: Standard normal value with 90% confidence interval. Roughly 270 observations for full credibility. Our experience is fully credible for the first 5 years.
Stat: 5,000 observations for full credibility (this is prescribed in PA and NY).
- No formal process exists but we make sure the cells are not sparsely populated with data. We have begun thinking of credibility more with the new valuation table for LTD. For example, we have performed A/E studies against the table and in some cases there are segments that are fully credible based on the prescribed credibility thresholds.
We've created regression models to identify variables that drive significant deviations from the baseline. Based on the regression analysis, there can be cells that are partially credible for which a credibility weighted average would be used to calculate the termination rate.
- Termination studies are performed every 2 years to update the STAT basis (i.e. first 5 years of claim).
- We have updated our LTD termination rates twice since the SOA 2008 Experience Report. Adjustments were made by durational buckets; i.e we followed the procedures prescribed by the 2008 Experience Report. Credibility is based on 95% / 10% confidence intervals. We have developed credibility tables that are similar to the 2012 GLTD Valuation credibility tables, but they're slightly different. Full credibility = 384 terminations.
- For a given duration, there exists a tabular termination rate (e.g. 2012 GLTD). We use a binomial distribution to model how rates are expected to deviate from their observed value. If the deviation is significant, then we will keep the tabular value. If the rates do not deviate very much, then we use the observed value. Binomial distribution uses 15% / 85% range. If the model produces values that are outside of the 15th and 85th percentile, then we revert to the tabular value.

- For LTD, we previously used judgment but will use the new LTD guideline related to new table going forward. For life waiver we use the waiver actuarial guideline credibility formula

(c) What is your baseline?

- Our baseline is the 2008 basic table. The methodology is identical to the 2012 table except that there are no margins.
- 1987 CGDT for GAAP and Stat
- We don't have a baseline reserve standard. We assign 100% credibility to terminations from our own experience for the first 5 years of duration, and to the industry tables for future durations. Hence, it is critical that we update our study every 2 years to capture most recent trends.

(d) By what parameters do you vary your credibility?

- Credibility is based on number of terminations
- Credibility is based on the number of terminations in each duration bucket.

12. Describe any challenges or requirements imposed by regulators or auditors when applying credibility to reserve calculations.

- The topic of credibility in reserve calculations per se hasn't arisen
- One state has rejected the 5,000 threshold for stat in the past
- No significant issues. A while back, when a new GAAP reserve basis was developed, there was some noise from auditors about using sparse data, and now we pay closer attention to this issue.
- No issues. We do not apply credibility to reserve calculations. Our experience is considered fully credible and we use company-specific assumption in the first 5 years.
- There haven't been any issues. We have changed our termination basis twice since 2008 and haven't had resistance from regulators
- No issues. Statutory minimum reserve requirements require a minimum number of closures in order to adjust the 1987 CGDT tables beyond 24 months. We don't have enough experience to extend beyond 24 months, so there have not been any issues.
- Artificial "floors" or "ceilings" that don't allow for recognition of full company level experience basis

13. With regard to the 2012 GLTD Table, please describe any issues that have arisen as you have applied the limited fluctuation credibility model prescribed by the GLTD Work Group.

- We have begun to attempt to understand the limited fluctuation credibility model but have not yet done any work to apply it
- We have had no major issues in applying this method. We have modified the method to give full credibility to termination experience within the change in definition transition period. We assume our experience is fully credible through five years of claim.
- No issues. In the first 5 years, our experience is fully credible. Thereafter, we have partial credibility.
- We are just starting to look at this
- No issues.
- This is still a work in progress. We have not evaluated the 2012 GLTD Table and the prescribed model.
- We haven't had a chance to work with it yet.
- We have not tried it yet.
- None to date.
- We are still in the development phase of using the new LTD proposed guideline.
- Still evaluating the impacts of this new regulation.

Part IV: General

14. What are your most significant challenges and issues in applying credibility in LTD?

- The market's expectation of credibility is the most significant challenge. We see carriers frequently applying too much credibility to experience. This creates the expectation amongst brokers and customers that favorable experience is credible and thus a rate decrease is expected
- We would like to see one industry-wide empirical test
There are trade-offs to make things simpler. We can't reflect too many variables because the method becomes difficult to explain to underwriters. The underwriters have a lot of freedom to modify the credibility, so we need to have them on board with our method. It is a challenge to find a method that works and is executable.
There tends to be a bias toward assigning higher credibility to company experience (a good case and a case with good experience are not necessarily the same thing).
It is difficult to explain why we would need a rate guarantee for a 100% credible case. This just means that past experience alone is the best predictor of future experience, not that future costs are 100% guaranteed.
- For case rating, based on feedback from UW and sales, our credibility levels are lower than what the competition would assign. We believe that the claim count basis is the right way to go, but we get pushback because it naturally assigns higher credibility for bad experience and lower credibility for favorable experience.
LTD is a low frequency event, so the number of claims would have to be very high to have true statistical credibility.
There is a lot of variability in the marketplace from carrier to carrier. Carriers have inconsistent approaches.
How underwriting impacts credibility presents a challenge.
- The biggest challenge is that theoretical credibility is very different from what is used in practice. If we want a tight range around our expected variance, we need a significant amount of experience. A claims-based approach seems more correct theoretically (vary by expected incidence).
We can see that there are entire blocks of business that are not credible, because earnings bounce around significantly from quarter to quarter.
- From a pricing perspective, even with a very credible block (cell), our experience is still volatile. LTD is impacted by economy and interest rates.
From an underwriting perspective:
Mergers and acquisitions – we continue to question whether historical view is indicative of future outlook with changing and evolving demographics.
Fewer people retiring – again, plays to the change in demographics, aging workforce with the attrition normally associated with hiring practices.

Competitive intelligence – carriers are more reluctant to release detailed information to use in predictive modeling.

Carrier loyalty – groups don't stay with incumbent carrier for more than 3 years, which makes it hard to establish a credible, predictable claim picture.

Employers are shifting to voluntary products, making the experience less _____ going forward (word was missing from the response).

The impact of the recent economic downturn.

- One of the greatest challenges is that there is no good place to go to get an industry idea on best practices. Approaches are subjective. Also, with LTD, all credibility formulas seem to give too much credibility to any one case.
- What is the right way to think about and calculate credibility? The same formula applying to all situations seems to be a problem.

There certainly exist opportunities for innovation: the interplay with the manual rating basis makes the whole credibility framework a complex process.

Claim experience is a key component in credibility.

Pooling is an important consideration.

- Credibility is often applied as a single number. I believe there may be something missing in this approach. Experience is a function of (1) premium, (2) incidence and (3) severity. Perhaps different credibility standards should apply to incidence and severity separately. Predictive modeling may help to better understand and identify how to work with incidence and severity separately in the credibility framework. In the past, we have tweaked our methodology for estimating claims (frequency), but we've never targeted the other components of claim.
- It would be nice if there were different methods to look at, to see some of the different approaches that can be used for determining credibility in LTD, because there has not been very much research nor is there very much published material on what other models are out there.
- Bringing in underwriters from different environments and with different ideas can be challenging to work with. We would prefer to see a more unified approach to applying credibility across the industry.

Data quality issues: there is so much subjectivity that surrounds the credibility process when it comes to data quality.

Manual rates: should credibility of experience increase when the manual basis is not very good? In other words, should there be a credibility factor for the manual component and a separate credibility factor for the experience component, and then the blended experience could reflect both factors.

It is human nature to think that good experience is more credible. Is there any truth to this?

- Market view of credibility vs. actuarial theory.

Credibility Survey for State Regulators

Summary of Responses to Survey Questions for Regulators

Detailed survey responses from LTD regulators are provided in the following section. State regulators have unique and valuable perspectives on this subject. We believe that their responses to survey questions, coupled with survey responses submitted by actuaries who practice in LTD insurance, may help define the primary issues and challenges in applying credibility in LTD.

The following is a brief summary of survey responses from state regulators.

- Credibility is considered in the context of experience rating and manual rate development in a few states, but not by the majority of state regulatory agencies.
- Regulators look for consistency in the application of credibility from filing to filing. Significant variations in the credibility standards used from one filing to the next can raise red flags.
- Lack of a uniform standard for applying credibility in LTD is problematic.
- The complement of credibility (i.e. the use of experience from other sources when the underlying experience is only partially credible) is equally as challenging to apply.

Credibility Survey for State Regulators

1. In what context do you consider credibility in the regulation of long-term disability insurance?

- Manual rate development
 - Experience rating formulas
 - Reserving
 - Other (specify)
 - Not at all
-
- On the product review side, we consider credibility in the context of manual rate development and experience rating formulas. The solvency area pays more attention to reserves.
 - We consider credibility in the context of ratemaking and experience rating.

2. Are the LTD credibility standards unique to LTD, or does your office use common standards across all accident and health product lines?

- In the state of Florida, rating rule 690-149 defines credibility criteria for manual rate development and experience rating. This rule distinguishes between low frequency and high frequency benefits products. The credibility standards for LTD apply to other low frequency benefits products, like long-term care and critical illness insurance.
- Our credibility standards are more unique to LTD.

3. This question pertains to credibility applications in experience rating.

(a) Do you explicitly review credibility factors in LTD experience rating formulas? If you do, are there formal or informal regulatory standards (or minimum requirements) that these formulas should meet? If so, what are these standards and what are some ways to determine if there are issues with any one formula?

- Yes, we do review credibility factors in LTD experience rating formulas. Our overriding concern is that LTD insurers need to demonstrate the predictive ability of the experience rating process, which includes credibility. Rule 690-149 puts forth criteria for applying credibility in experience rating. Satisfying these criteria is sufficient, but if there are other justifiable methods that differ from the credibility criteria outlined in rule 690-149 then they will be considered as well.
- Yes, we do review credibility in experience rating but there are no formal standards. Demonstrating consistency is one of the biggest things that we look for. We don't think it's appropriate for companies to have one standard one year and a very different standard in another year. We also look at the reasonableness of credibility standards.

(b) What (if any) are the most significant challenges in the application of credibility in LTD experience rating?

- The inherent volatility in LTD experience requires a very large block for the experience to be predictive. Not many groups are big enough to have low enough volatility in the experience to be predictive.
- Many companies have their own credibility standards for experience rating that differ from ours, and they will argue that theirs is better than ours.
- Lack of a uniform standard for applying credibility.

4. This question pertains to credibility applications in manual ratemaking.

(a) For companies who justify manual rates on their own LTD experience, do you consider the credibility of the experience data? If so, how does your office determine the minimum experience requirements for manual ratemaking? Please describe in as much detail as possible, including (if applicable) the underlying credibility model used to determine the minimum thresholds.

- Yes, we consider the credibility of experience that gets used for manual rate development. Rule 690-149 puts forth criteria for applying credibility in manual rate development. These criteria must be satisfied for rates to be approved.
- Yes, we consider the credibility of the experience data. We leave it to the insurer to define the credibility standard, but we look at it in terms of reasonableness and consistency. We want to make sure the standards are consistent from filing to filing.

(b) What (if any) are the most significant challenges in the application of credibility in manual ratemaking?

- Very often, companies that do not have credible experience will want to use the experience from a reinsurer. The challenge is in showing that the reinsurer's experience is applicable. For example, differences in market distribution systems can lead to differences in experience. If companies use reinsurers' experience then they basically need to show that the underlying risk classes are similar.
- Lack of a uniform standard.

General Remarks:

- Our state needs a methodology for applying credibility that every company can easily adopt, regardless of systems restraints. There is a wide range of systems capabilities out there. Some have intricate data capabilities and some have poor data capabilities. Given the wide range of system skill levels, the credibility model prescribed by our office is kept simple. It would be nice to, say, differentiate frequency and severity components on an elimination period basis within the credibility model, but many insurers would not have the capacity to do this.

We recently held a workshop with actuaries from various insurance companies and asked if they could support more intricate credibility models. Only a few of the people in attendance responded that they could support more complex models.

- Consideration should be given to the complement of credibility in manual ratemaking, when the experience is only partially credible. Maybe there could be a hierarchical system that that would rank other sources of experience starting with best choice (e.g. national experience) to next-best choice (e.g. statewide experience) and so on. This complement-of-credibility model could then be applied consistently from filing to filing, making the process more scientific and less open to cherry-picking.

Credibility Survey for Actuaries Practicing in Medical Insurance

Summary of Responses to Survey Questions for Actuaries Practicing in Medical Insurance

The survey for actuaries practicing in medical insurance is intended to develop information that could be used to support more effective use of credibility in group long-term disability (LTD) applications. One of the objectives is to identify opportunities that may exist in medical insurance that could be extended to LTD.

The following is a brief summary of survey responses from actuaries who practice in medical insurance.

- Most credibility formulas used in medical insurance are based on the number of member months in the experience period.
- Actuaries are beginning to use predictive modeling techniques to inform decisions on credibility.
- Competitive pressures present the most significant challenge in applying credibility to medical insurance.

Credibility Survey for Actuaries Practicing in Medical Insurance

Section 1: Credibility Applications in Case Underwriting and Experience Rating

15. (a) How is volume measured in your credibility formulas (check all that apply)?

- Lives
- Life years of exposure
- Member months
- Actual claims
- Expected claims
- Premium
- Other (specify)

(b) What additional segments, if any, do your credibility formulas vary by?

- Occupation class
- Benefit levels
- Type of product
- Other (specify)

(c) Please describe your formulas in as much detail as possible, including at a minimum the underlying theory, basic structure, minimum credibility level and full credibility threshold. (Please note that this information will be used to draw broad conclusions regarding current industry practices. Company-specific formulas will not be shared outside of the research team.)

Participant 1:

(a) Our formula is based on member months and number of subscribers (i.e. covered employees). Most credibility formulas used in medical insurance are based on member months.

(b) Our formula varies by occupation class.

(c) Our credibility formula is not based on a theoretical model. It is based primarily on the generally accepted credibility level in the marketplace. Minimum credibility starts at 51 subscribers. Experience is deemed fully credible at 500 subscribers and 10,000 member months.

Participant 2:

(a) Our formula is based on member months.

(b) Type of product. We vary the threshold by which a group is considered fully credible by medical versus pharmacy benefits.

(c) No minimum really. Experience rating begins at 50 employees (roughly 90 members on average, which translates to about 1,000 member months because we use 12 months to rate).

Sometimes it dips below because statutory regulations require large groups (i.e. groups of 50 or more employees) to be experience rated. But you could get groups in which 40 employees are enrolled and 10 are eligible, which would qualify the group for experience rating. Or sometimes there are groups made up of mostly single members (i.e. no family members), and you could have 60 members which translates to only 720 member months.

Full credibility for medical benefits is attained at 12,000 member months (i.e. 1,000 members).

Full credibility for pharmacy benefits is attained at 7,200 member months (i.e. 600 members).

16. Has predictive modeling been used to inform the structural design of your credibility model, for example to identify segments that are key drivers of experience? If yes, please explain how predictive modeling has been used in the credibility framework.

Participant 1:

We use predictive modeling but not in the context of credibility. Predictive modeling is used in the manual component of rating.

Participant 2:

Not currently, but we are beginning to take predictive modeling approaches into account. This year, with the ACA, there is a large focus on small group and individual products, and we have standardized practices in these markets. Eventually, hopefully in the next year or two, we plan on revising and standardizing practices for large groups, and at that time we will take advantage of predictive modeling applications.

17. Have you done any testing of your credibility formulas and methods? If yes please describe (in as much detail as possible).

- Yes
- No

[Description]

Participant 1:

We haven't tested the credibility formula explicitly, but when we implemented a new rating method we tested the new method, and credibility was looked at indirectly because it is part of the method. Basically, we looked back in time to see how well the new rating method predicts known experience. The test was based on two years of historical experience. The new rating method was applied to the first year of experience to see how well it predicts the second year. We did end up tweaking credibility formulas for smaller sized cases as a result of the testing.

Participant 2:

No.

18. How often are your credibility formulas updated?

- Annually or more often
- Every 2 – 5 years
- Every 5 years or less often

Participant 1:

They are not updated on a regular basis. However, we do evaluate how well our credibility formula stacks up against the competition on a regular basis, about once every year. Our perception of market credibility is based on feedback from Underwriting and Marketing.

Participant 2:

Every 5 years or less often.

Section 2: Credibility Applications in Ratemaking and Valuation

19. Do you employ explicit credibility techniques in developing manual rates? For example, are credibility formulas or thresholds part of the manual ratemaking process, or is the process driven more by subjective decisions? Please describe in as much detail as possible.

Participant 1:

Yes, we have explicit thresholds that tend to be much higher for manual rate development when compared to experience rating. Due to health care reform, we have had to make changes to the pricing of small case group business, and credibility has been a key component of manual rate development. I'm not sure of the exact details though.

Participant 2:

We don't use manual rates for non-credible large groups. Instead, we use a pooling market loss ratio methodology. This process involves adjusting non-credible cases for pricing trends and a target loss ratio. Fully credible groups are excluded from the pool. Partially credible groups get weighted by the credibility factor, so for example if a group is 50% credible then they are weighted by a factor of 50% in the pool. The approach is formulaic.

20. How is credibility applied in developing reserve assumptions? Is a formal approach used to determine if the experience is credible enough for deriving assumptions, or is the process driven more by subjective decisions? Please describe in as much detail as possible.

Participant 1:

Not explicitly. It's based on subjective judgment.

Participant 2:

The manner in which we determine whether a segment is big enough to result in its own reserving segment is typically driven by membership. The lowest membership I've seen in its own reserving cut is about 1,000 members.

With regards to segmentation for reserving, that's typically the result of studies (e.g. logistic regressions, etc.) resulting in recommendations which we either accept and use for several years, or simply keep reserving with our current segments. Right now, our segmentation is by line of business (e.g., small group, large group, etc.) and product (CDHP vs. non-CDHP). CDHP means "consumer-driven health plan". Basically it's a higher deductible plan with an H.S.A. or H.R.A. attached to it.

Section 3: General

21. What are the most significant challenges and/or issues in applying credibility in medical insurance?

Participant 1:

No real challenges or issues. I would say, however, that the levels of credibility used in the market are not theoretically correct.

Participant 2:

Competitive pressures from the market present the most significant challenges. Credibility is intended to balance rates over time. If a group is on the lower end of the credibility curve and has had favorable experience, but the average pricing trend is 8 or 10%, that group would get a rate increase. But the competition will see the favorable experience on that group and produce a more attractive quote. This is a big challenge because the credibility formula is intended to balance the pooled loss ratio, but the competitive market dynamics complicate the process. If the same group was fully credible you might not increase rates.

Credibility Adjustment Factors for Medical Loss Ratio Refund Calculation

The following are notes from a conversation with the actuary who developed the credibility factors used in the federal Medical Loss Ratio formula that all commercial medical insurers must use.

1. How are concepts in credibility applied in the medical loss ratio (MLR) formula?

The MLR requirement applies in every state and within three blocks of business: (1) the individual, (2) the small group, and (3) the large group medical insurance markets. There are concerns about splitting medical experience data among 50 states and three business segments within each state, because the data as such may not be credible, especially for smaller sized insurers. Even large companies may not have substantial blocks of policies in some states. So there are credibility issues with MLR reporting, which requires insurers to meet a minimum 80% loss ratio in the individual and small group markets and an 85% loss ratio in the large group market.

Historically, loss ratio fluctuations from year to year would tend to average out over time, and insurers would hit business targets over the long term. With the new MLR requirement, however, insurers cannot offset losses (loss ratios above target) with profits (loss ratios below target) as they were accustomed to doing in the past.

The new law recognizes these issues and allows some companies to make credibility adjustments to their reported loss ratios when the loss ratio is less than the MLR standard. Basically, an insurer will calculate its loss ratio, and then may be permitted to add a credibility adjustment to the loss ratio based on the number of life years in the experience period, per the adjustments shown below.

| Life Years | Additive Adjustment |
|-----------------|---------------------|
| < 1,000 | No Credibility |
| 1,000 - 2,499 | 8.30% |
| 2,500 - 4,999 | 5.20% |
| 5,000 - 9,999 | 3.70% |
| 10,000 - 24,999 | 2.60% |
| 25,000 - 49,999 | 1.60% |
| 50,000 - 74,999 | 1.20% |
| 75,000 + | 0.00% |

As can be seen from above, no adjustments are permitted for insurers with experience that includes 75,000 life years (i.e. member years) or more. The theoretical statistical fluctuations differ somewhat

those actually used. For example, there continues to be decreasing variability well over 75,000 lives, but a regulatory decision was made to allow adjustments only under the 75,000 threshold.

2. *What were some of the key considerations when you first began to develop the credibility framework?*

We spent a lot of time thinking about how to model it. We developed a stochastic model that uses Milliman’s health care guidelines probability distributions, which contain probability distributions for a standard set of benefits. In the model, we simulated 10,000 trials (i.e. medical insurance loss ratios) for different sized insurers, starting with an insurer with 1,000 insured lives, followed by an insurer with 2,000 insured lives, and so on up to an insurer with 150,000 insured lives. We then looked at the variability that was evident on the downside (since MLR credibility adjustments only apply to situations when the loss ratio is less than the MLR standard).

Another key consideration was what plan design to use in the modeling. We modeled different types of plans and noticed significant deviation of results by plan design (i.e. there was a larger variability of loss ratios for certain plan types). This suggested that not only does size of company matter, but benefit design also matters. We ended up using variation in deductibles, which may be imperfect but it keeps the application simple, and most insurers can easily access this information. So the factors in Table 2 shown below reflect differences in plan design through variation in deductible.

| Deductible Range | Adjustment Factor |
|-------------------|-------------------|
| < \$2,500 | 1.000 |
| \$2,500 - \$4,999 | 1.164 |
| \$5,000 - \$9,999 | 1.402 |
| >= \$10,000 | 1.700 |

The current approach adopted by the NAIC mandates each insurer calculate an average deductible for the entire business, which then specifies the adjustment factor from Table 2 to apply to the Table 1 adjustment. We would have preferred that each deductible bucket in the experience get its own factor, then the companies could take a weighted average of the factors weighted by number of policyholders in the deductible bucket.

In the modeling, we used the probability distributions to generate expected claim costs, and then adjusted resulting claim costs for deductibles, coinsurance, and stop loss insurance. We did this for everybody in that block of business (e.g. block with 1,000 lives, block with 2,000 lives, ... , block with 150,000 lives). The mean of the claims (i.e. pure premium) was known in advance. We then compared results of our simulations to the mean of claims to test variability below the 80% loss ratio threshold.

Age and gender were also important considerations, but they did not make it into the final credibility tables.

3. What are some of the challenges in applying credibility in the pricing of medical insurance?

There are so many variables that can affect claim costs in medical insurance. For example, geographic area: there exist significant variations in costs across jurisdictions. Demographics are another key variable that complicates credibility analysis. There are significant differences in claim costs between males and females at young ages, for example, making it difficult to apply credibility when rates are unisex. In group insurance, industry and group size really matter. There are differences in claim costs that have more to do with adverse selection than random processes. Larger groups have better dispersion of risks, and the fact that the employer doesn't know about everyone's medical backgrounds, reduces the incidence of adverse selection. However, this is not true of smaller sized groups.

Actuarial Guidelines Pertaining to the Use of Credibility

Accident and health actuaries use credibility now more than ever, due to the fairly recent ideological shift to principle-based approaches for calculating reserves and capital. For instance, credibility plays a key role in the latest group life waiver of premium and LTD valuation standards. As credibility becomes increasingly more prominent in accident and health insurance applications, new actuarial guidelines have cropped up to support the use of credibility in specific areas. The following is a list of published actuarial guidelines that pertain to the application of credibility in group insurance:

- Actuarial Standard of Practice No. 25
- Actuarial Guideline for the 2012 Group Long-Term Disability (GLTD) Valuation Table
- Group Term Life Waiver of Premium Disabled Life Reserves Actuarial Guideline

Although these guidelines were published at different times and for different purposes, there are common themes that run through each of the documents. For example, they all describe the importance of carefully analyzing the experience in order to determine if the experience is relevant or should be segmented into homogeneous risk classes.

The actuarial guidelines for the 2012 GLTD and Group Term Life Waiver of Premium valuation standards both prescribe the credibility models that should be used for blending company specific experience with the industry tables. These two guidelines also specify an acceptable length for the experience period used to develop company-specific assumptions.

In addition to the three sources listed above, the state of Florida has published regulatory guidelines for applying credibility in LTD. Rule 69O-149 defines credibility criteria for the pricing of low frequency products, like long-term disability insurance.

Following is an outline of each of the documents described above.

Actuarial Standard of Practice No. 25: Credibility Procedures Applicable to Accident and Health, Group Term Life, and Property/Casualty Coverages

The Actuarial Standards Board adopted Actuarial Standard of Practice No. 25 (ASOP 25) to govern the use of credibility in actuarial disciplines. The document includes guidelines and recommendations for using credibility. The following key points were taken directly from ASOP 25.

- The actuary should be familiar with and consider various methods of determining credibility. The models selected may be different for different applications. The selection process involves testing the tentatively selected model and possibly revising the model. The actuary should select credibility procedures that do the following:
 1. Produce results that are reasonable in the professional judgment of the actuary,
 2. Do not tend to bias the results in any material way,
 3. Are practical to implement, and
 4. Give consideration to the need to balance responsiveness and stability.
- The actuary should use care in selecting the related experience that is to be blended with the subject experience. Such related experience should have frequency, severity, or other determinable characteristics that may reasonably be expected to be similar to the subject experience. If the proposed related experience does not or cannot be adjusted to meet such criteria, it should not be used. The actuary should apply credibility procedures that appropriately reflect the characteristics of both the subject experience and the related experience.
- In carrying out credibility procedures, the actuary should consider the homogeneity of both the subject experience and the related experience. Within each set of experience, there may be segments that are not representative of the experience set as a whole. Credibility can sometimes be enhanced by separate treatment of these segments.

Actuarial Guideline for the 2012 Group Long-Term Disability (GLTD) Valuation Table

This actuarial guideline governs the calculation of claim termination rates used for estimating LTD claim reserves under the 2012 GLTD valuation standard. The following points were taken directly from the guideline and specifically address applications of credibility in the 2012 GLTD valuation standard:

- If not invoking the small company exception, a company must use a credibility-weighted combination of its own claim termination experience with the 2012 GLTD Valuation Table to create its specific valuation table.
- Credibility weighting factors (shall be) developed for each duration group.
- The appointed actuary shall:
 1. Segment the company claim termination experience into any major subgroups that may produce significantly different results (e.g., market niches, claims operations, unique benefit designs, etc.);
 2. Combine affiliated statutory entities and assumed reinsurance, where claims management is under a common structure, when considering company experience. It is also appropriate to evaluate experience separately when specific blocks of company business have distinct claims-management practices or significantly different risk characteristics.
 3. Include all relevant experience the company is capable of providing for as many of the last five years as possible, not including (a suitable) lag period.
- If, at the time of valuation, a company has fewer than 50 open claims disabled within two years of the effective date of the valuation, and fewer than 200 open claims disabled more than two years prior to the effective date of the valuation, the carrier is exempt from the requirement that the 2012 GLTD Valuation Table be modified by the company's own experience. Said company will use 100 percent of the 2012 Valuation Table for calculating claim termination rates in order to comply with the minimum valuation standard.

Group Term Life Waiver of Premium Disabled Life Reserves Actuarial Guideline

This guideline governs the calculation of group term life waiver of premium disabled life reserves under the 2005 GTLW valuation standard. Prior to the publication of this guideline, there had been no formal recommendations for calculating waiver of premium reserves. This document includes guidance on blending company-specific experience with the 2005 GTLW Mortality and Recovery Tables. The following points were taken directly from the guideline and specifically address applications of credibility in the 2005 GTLW valuation standard:

- Company experience shall:
 1. Be segmented into policies with similar benefits, on individuals of each gender;
 2. Be experience-specific to the company;
 3. Include all relevant experience in the past three most recent years;
 4. Exclude experience that is not in the past six most recent years;
- A company may use a credibility-weighted combination of company mortality experience with the 2005 GTLW Mortality Tables and/or of company recovery experience with the 2005 GTLW Recovery Tables to create its specific valuation tables.
- The blended tables for each gender and type of experience (mortality and recovery) shall be computed using the (prescribed) formula Blended Table.

State of Florida Rule 69O-149.0025

This rule provides guidance on the application of credibility in manual ratemaking for insurance policy forms with low expected claims frequency, such as LTD. According to the Florida Office of Insurance Regulation, the criteria for applying credibility in rule 69O-149.0025 can also be extended to applications in experience rating. The following points were taken directly from the rule and specifically address credibility:

- For policy forms with low expected claims frequency, the data from the fewest number of entire calendar years, starting with the most recent experience year and looking back year by year as necessary, to the calendar year in which the accumulated claims first equal or exceed a total of 1,000 claims, shall be assigned 100 percent credibility; 200 claims shall be assigned 0 percent credibility. If 100 percent credibility is not achieved by using the most recent five year period, the data from the most recent five year period only shall be used.
- Florida only experience shall be used if it is 100 percent credible. If Florida experience is not 100 percent credible, a combination of Florida and nationwide experience shall be used. The Florida data shall be given the weight of the ratio of the Florida credibility to the nationwide credibility.
- The data is combined using the indicated weights. The combination of the two weights will always equal 100 percent. A rate change is determined from the blended data. If the nationwide credibility is less than 100 percent, the indicated rate change is weighted by the nationwide credibility and medical trend, if applicable, by the compliment of the nationwide credibility. If nationwide credibility is 100 percent, there would be no trend component.

Annotated Bibliography

The following sources discuss credibility in a way that can be applied to LTD.

- **American Academy of Actuaries' Life Valuation Subcommittee, *Credibility Practice Note*, July 2008**

This paper discusses common practices in applying credibility and issues that often arise when applying credibility in life and health insurance. Specific examples of practical applications are provided to highlight the strengths and weaknesses of the approaches taken in applying credibility in each of the examples. The examples mostly focus on credibility standards upheld by state regulators, credibility applications for estimating future claim experience, and credibility applications for updating experience assumptions. There is also a section of this paper that covers historical background and prominent theoretical results in credibility.

- **Greg Barn, *Credibility in Group Insurance Pricing*, Risk Matters, July 2012, pgs. 1-3**

This is a brief article addressing some of the issues in applying credibility in the context of pricing group insurance. The author provides a list of considerations that actuaries should take into account when thinking about credibility in the pricing of group insurance products, including product design, quality of manual rates, loss outliers, and the demographic mix of plan participants.

- **Joseph A. Boor, *The Complement of Credibility*, Proceedings of the Casualty Actuarial Society, Volume LXXXIII, Part 1, No. 158**

This paper discusses the complement of credibility in broad enough terms that seem applicable to LTD. The author provides several examples of common applications of the complement of credibility in actuarial practices. Included in the discussion are criteria to assess the effectiveness of each of the applications, such as whether or not the application contains a bias, or if the application may produce inaccurate results. The discussion is easy to follow and the examples are meaningful.

- **Curtis Gary Dean, *Topics in Credibility Theory, Construction and Evaluation of Actuarial Models Study Note*, 2005**

This study note discusses applications of the Buhlmann-Straub credibility model in actuarial mathematics, including procedures for estimating parameters using non-parametric and semi-parametric methods. The parameters are defined in terms of risk, i.e. the variance of hypothetical means is described in terms of homogenous and heterogeneous risk classes, and the expected value of the process variance is described in terms of variation in individual risk

experience. The Buhlmann-Straub credibility model is developed both conceptually and through the use of examples in this study note.

- **Charles Fuhrer, *Credibility with Incomplete Information in Group Insurance*, Actuarial Research Clearing House, Vol. 1, 1996, pgs. 593 – 595**

This paper analyzes the impact of incomplete claim information on credibility in group insurance, arising from the existence of claims incurred but not reported (IBNR) as of the valuation date. Starting with a credibility model that uses linear least squares approximations, the author develops credibility formulas that actually produce lower credibility estimates when the experience is believed to be incomplete due to IBNR claims. The examples provided pertain to group life and group long term disability insurance.

- **Charles Fuhrer, *Credibility Theory*, Record of Society of Actuaries, Vol. 19 No. 1B, 1993, pgs. 863 – 881**

Although credibility is discussed in terms of group medical insurance, many of the concepts can be extended to other group insurance products such as long term disability. The paper includes a comparison of credibility applications in manual ratemaking and experience rating, a review of some standard credibility formulas, and a discussion of why these formulas may or may not apply well in group insurance. The paper also describes credibility models that are more applicable to group insurance, both in terms of group dynamics and the nature of claims. A section at the end of the paper provides a list of interesting questions and answers pertaining to applications of credibility.

- **Charles Fuhrer, *Some Applications of Credibility Theory to Group Insurance*, Transactions of the Society of Actuaries, 1988, Vol. 40 Pt. 1**

Methodologies for developing credibility formulas applicable to group insurance are given in this paper, based on best estimate approximations using least-squares criteria. Several credibility formulas are developed that can be applied specifically to the following scenarios:

- Variable group sizes
- Changes in the number of participating group members
- Outlier claims
- Competitive pricing pressures

The paper also includes a description of group insurance dynamics, a summary of statistical concepts, and a comparison of the derived credibility formulas to other credibility models.

- **Roy Goldman, *Pricing and Underwriting Group Disability Income Coverages*, Transactions of the Society of Actuaries, 1990, Vol. 42**

Beginning on page 197, there is a discussion of the challenges in evaluating LTD experience. Classical and Bayesian credibility concepts are discussed starting on page 209, along with detailed examples of how to apply theoretical concepts when experience rating LTD products. Goldman's credibility discussion ends on page 213. At the end of the paper, there are comments made by other actuaries on the credibility topics introduced in this paper.

- **Vincent Goulet, *Principles and Application of Credibility Theory*, Journal of Actuarial Practice, Vol. 6, 1998**

This paper contains a thorough review of traditional credibility theory, focused on applying limited fluctuation, Bayesian, and Buhlmann-Straub credibility models in experience rating insurance products. The examples provided help shed light on practical applications and estimating model parameters. There is also a discussion regarding to which situations the models should and should not be applied. At the end of the paper, hierarchical and crossed-classification credibility models are proven to work well when applying credibility in experience rating large portfolios.

- **Group Long-Term Disability Valuation Standard Report of the American Academy of Actuaries' Group Long-Term Disability Work Group, American Academy of Actuaries, August 10 2012**

This report describes the limited fluctuation credibility model that has been prescribed by the 2012 GLTD Working Group for blending company-specific experience with the 2012 GLTD Claim Termination basis. In the description of the model's parameters, issues pertaining to non-independence of group LTD risks and volatility by claim duration are discussed. The methods for calculating full and partial credibility thresholds are discussed in detail.

- **Thomas N. Herzog, *Introduction to Credibility Theory*, 4th Edition, 2010**

This text book traces through key historical developments in credibility theory and provides a comprehensive analysis of credibility models from both theoretical and practical points of view. A wide range of topics in credibility are covered, including limited fluctuation credibility, Buhlmann's contributions along with the Buhlmann-Straub model, Bayesian approaches to credibility, and the least squares methods that pertain to credibility. Some examples that are provided illustrate practical applications of credibility concepts in health insurance, and there are case studies that focus on health care issues that relate to credibility.

- **Stuart A. Klugman, Harry H. Panjer and Gordon E. Wilmot, *Loss Models: From Data to Decisions –Third Edition, 2008***

This text book includes a section on credibility theory (Chapter 20, pages 555 – 636). Limited fluctuation credibility theory is discussed in section 20.2. Greatest accuracy credibility theory is covered in section 20.3, focusing on Bayesian and Buhlmann credibility concepts. Section 20.4 covers techniques for estimating the parameters used in credibility models, including parametric, semi-parametric and non-parametric estimation techniques.

- **Stuart Klugman, *Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model, Proceedings of the Casualty Actuarial Society, Volume LXXIV, pgs. 272-321***

This paper includes a comprehensive discussion of Bayesian methods for applying credibility in ratemaking. The first section provides several different ways that Bayesian credibility can be applied in ratemaking, namely by using empirical, parametric empirical, and hierarchical Bayesian analyses. The remaining sections focus exclusively on hierarchical normal linear models (HNLM). Theoretical support of HNLM is given, along with several examples of common HNLM ratemaking models. These examples relate to Workers' Compensation insurance and use historical experience data to estimate both the frequency of claims and loss ratios for different Workers' Compensation rating classes.

- **Myron H. Margolin, *On the Credibility of Group Insurance Claim Experience, Transactions of the Society of Actuaries, 1971, Vol. 23 Pt. 1 No. 67***

A unique approach for calculating credibility factors that can be applied in group insurance is given in this paper. The approach differs from Bayesian methods and other theories of credibility in that the credibility factors are developed from observed correlation coefficients between loss experience in consecutive policy years. The paper includes an overview of the pertinent statistical methods. At the end of the paper, there is reaction to the topic from several different actuaries practicing in group insurance.

- **Hassett, Matt and Januzik, Brian, *Credibility: Theory Meets Regulatory Practice***

This paper provides examples of regulatory standards for applying credibility, along with a discussion of the positive and negative attributes of each of the standards. There is a section that deals with credibility applications in disability insurance titled "Credit Insurance and Credibility". The following examples are included in this paper:

- Credibility criteria for the Medicare supplement refund calculation
- Medicare supplement rate increase filing projections
- Credibility in credit insurance

- **Howard C. Mahler, *Credibility with Shifting Risk Parameters, Risk Heterogeneity, and Parameter Uncertainty*, Proceedings of the Casualty Actuarial Society, Vol. LXXXV, 1998, pgs. 455 - 653**

This paper examines situations that complicate the straightforward application of credibility concepts in insurance operations. In these types of situations, approaches for extending standard credibility formulas are presented, along with some examples showing how the theory can be applied in practice. Although the examples are not specific to group disability insurance, the material does seem like it could be extended to group LTD. There are several examples that cover credibility applications in Workers' Compensation insurance.

- **Glenn Meyers, *Empirical Bayesian Credibility for Workers' Compensation Classification Ratemaking*, Proceedings of the Casualty Actuarial Society, Volume LXXXIV, pgs. 96-121**

This paper examines the application of an empirical Bayesian credibility model in ratemaking for Workers' Compensation insurance. Both the model and the methods used for estimating the model's parameters are discussed in detail. A demonstration showing how the credibility model can be applied to classification ratemaking is provided by using Workers' Compensation data from the state of Michigan from 1982-1983 (the data and the results are provided in Exhibit 1 at the end of the paper). The strength of the proposed model is tested by using a variety of statistical tools.

- **Hakop Pashayan, *A Modern Approach to Group Risk Pricing and Credibility*, Presented to the Institute of Actuaries of Australia 2009 Biennial Convention, Sydney, Australia**

This paper covers credibility applications in group insurance pricing, focused on lump sum insurance products in Australia. A review of the concepts in credibility theory that are being applied in group insurance pricing is given. There is a comprehensive discussion of the limitations of traditional credibility models as applied in group insurance pricing, including a look at Bayesian, Buhlmann and limited fluctuation credibility models. In sections 4 – 6, a new model is proposed for pricing group products that feature lump sum benefits, and then the model's output is analyzed and compared to the output from more traditional credibility models.

- **Stephen W. Philbrick, *An Examination of Credibility Concepts*, Taken from the CAS Proceedings Volume LXVIII**

This paper uses examples to show how concepts in credibility can be applied in practice. There is also a discussion of how "process variance" and "variance of hypothetical means" affect credibility estimates, and a section on credibility applications in experience rating and

ratemaking. At the end of the paper is a brief discussion of the most widely known theories in credibility.

- **Rene Schnieper, *On the Estimation of the Credibility factor: A Bayesian Approach*, ASTIN Bulletin, 1995, Vol. 25 No. 2, pgs. 137 – 151**

This paper illustrates a technique for estimating the parameters in a credibility model that uses Bayesian methods for calculating credibility. The parameters (often called “variance of hypothetical means” and “expected value of process variance”) are treated as random variables whose values are estimated using statistical methods. A numerical example in section 3 shows how credibility can be applied in calculating credibility-weighted premium, once the model’s parameters have been estimated. Toward the end of the paper, the theory is extended to a family of exponential random variables to cases where exposure units are arbitrary, and examples are given.

- **Drew Tindall and Jess Mast, *A Credible Theory of Credibility, Contingencies*, September/October 2003, pgs. 45 – 47**

Although this paper focusses on issues in applying credibility in life insurance, many of the issues raised can be extended to long term disability insurance. For example, the authors identify potential issues in applying credibility when assuming claims experience is homogeneous, and these issues can also pertain to group LTD. Various examples are provided to show how an assumption of claim homogeneity breaks down in practice, and recommendations are given for handling situations in which historical experience is believed to be inconsistent with future experience.

- **Manuel Tschupp, *Application of Credibility Theory to Group Life Pricing*, Milliman Insight, August 2011**

This paper explores applications of modern credibility theory in the pricing of group insurance products, including disability insurance. At the beginning of the paper, there is a discussion of why the traditional Buhlmann-Straub credibility model may not apply directly to the pricing of group insurance products. Sections 2 and 3 introduce a credibility model for pricing group contracts. The model is based on a risk differentiation system that combines contract-specific and broad portfolio risk components to compute a credibility estimator for use in the pricing. Sections 4 and 5 describe special considerations for applying the proposed model, such as prospective versus renewal pricing.

- **Manuel Tschupp, *Application of Credibility Theory to Group Life Pricing – Extended Techniques*, Milliman, December 2011**

This paper is an extension of the paper entitled “Application of Credibility Theory to Group Life Pricing,” in which various issues pertaining to the pricing of group insurance products (including LTD) are explored. A multi-dimensional credibility model is introduced which takes into account several different sources of claim information, such as claim frequency and claim severity. The model is then extended to situations that commonly arise in the pricing of group insurance, such as situations involving outlier claims and inadequate underwriting. The paper includes a discussion of credibility floors, large-case pricing, and the inclusion of known external risks in the credibility framework.