Using Resampling Techniques to Measure the Effectiveness of Providers in Workers' Compensation Insurance

David B. Speights, Ph.D.,
Terry J. Woodfield, Ph.D., and Deborah Becker

David B. Speights is a research statistician at HNC Insurance Solutions. He holds a doctorate in Biostatistics from the University of California, Los Angeles. His current research is in the area of individual claims reserving, resampling techniques, and neural networks for right-censored data.

Terry J. Woodfield is Chief Statistician at HNC Insurance Solutions. He holds a Ph.D. in statistics from Texas A&M University. His research interests include statistical forecasting, statistical computing, and applications of advanced statistical modeling techniques to problems in insurance, finance, and retail sales.

Deborah Becker is a Product Manager at HNC Insurance Solutions. She has over 20 years of experience in the property/casualty insurance industry. Her research for the past three years has been provider profiling for the workers' compensation industry.

HNC Insurance Solutions is a business unit of HNC Software Inc. and was formed in the 1998 merger of CompReview Inc. and Risk Data Corporation. HNC Insurance Solutions is an insurance information service and statistical research company which creates and markets solutions for the insurance industry.

Abstract

We use resampling techniques to analyze the impact of providers on workers' compensation costs taking into consideration inherent differences in claim populations between providers. Resampling techniques provide a nonparametric determination of a statistic's distribution and a measure of effectiveness that is not sensitive to deviations from the assumptions underlying most parametric statistical procedures. These techniques are applied to a subset of an extensive nationwide database of workers' compensation claims to demonstrate the methods.

1 Introduction

A major cost saving method for the workers' compensation industry is to refer injured workers' to the most cost-effective provider. Often, inefficient providers dramatically increase costs by over treating or performing ineffective treatments which reduce the quality of care, prolong the length of disability, and increase both the potential for litigation and permanent disability. The goal of treatment should be to return the worker to suitable gainful employment as soon as possible, reducing costs and increasing patient satisfaction. It is important to be able to compare medical costs and indemnity costs between providers when determining the cost outcome of a workers' compensation claim.

When comparing providers, considerations must be made for the differences in claim populations represented by each provider. For example, we may be interested in comparing the total claim cost for patients served by providers A and B. If provider A services a large number of severe injuries, and provider B services no severe injuries, then we will most likely conclude that provider B is less expensive than provider A, even if the two providers are equivalent. Therefore, it is difficult to identify the provider with the lowest costs without accounting for inherent differences in characteristics indicative of claim severity. Throughout this paper, we refer to characteristics indicative of claim severity as comorbidity factors. In section 3, we describe techniques to risk adjust the data so that the comparisons are based on "like to like" factors. The type of risk adjustment used in this paper is also known as normalization.

Bootstrap resampling is a relatively new statistical technique that allows for nonparametric or semiparametric estimates of a statistic's distribution. Traditionally, statistical methods sought to determine analytically the distribution of a statistic. For example, the asymptotic and small sample distribution of statistics needed to compare population means or variances is well known. However, these distributional properties are often rooted in unrealistic assumptions about the population. In Section 2 we give a brief introduction to the idea of resampling. The idea is very straightforward and is applicable to a wide variety of situations. In addition, bootstrap techniques allow us to form complicated statistics that would normally have asymptotic and small sample properties that are difficult if not impossible to derive.

We apply the basic resampling concept for comparing the distribution of ultimate claim cost on two populations while adjusting for inherent differences in claim severity. The statistical methods are an extension of the methods presented in Efron & Tibshirani (1986) to deal with normalized populations. We show two different techniques for comparing two populations while adjusting for claim characteristics that are indicative of the severity of a claim. In addition, we form complicated statistics for comparing the two populations that would normally have asymptotic distributions that are difficult to obtain.

In this paper, two examples using data from HNC Insurance Solutions' Provider Compare® database are given to demonstrate how the methods can be applied to comparing claim costs between providers. First, it has become standard to refer injured workers to a provider network. It makes sense to compare providers not in a network to providers in a network. The outcome of profiling providers in and out of networks is outlined in section 4.1. The second example compares the total claim costs of one provider to a group of providers while accounting for the differences induced by 13 separate claim characteristics. This example can be found in section 4.2.

2 Introduction to Resampling Methods

In this section we give an introduction to resampling methodology. Resampling is a simple technique that was developed to serve two basic purposes. First, resampling provides a departure from the rigid assumptions that underlie many statistical procedures. Like many nonparametric and semi-parametric methods, bootstrap resampling provides a framework that is not constrained by assumptions on the data and error distributions. Second, resampling provides a framework for estimating the distribution of very complex statistics. Many times, a procedure is developed for estimating model parameters, but the distribution of the estimate is either too difficult to derive or requires unrealistic assumptions. Resampling techniques provide a straight-forward method for determining the distribution of any statistic.

Define our data as a sample of size n, $X_1, ..., X_n$, where X can represent a vector or a scalar. Assume the data arises from an unknown distribution function F. Based on the data, it is typically of interest to estimate a population parameter. We can usually denote a population parameter as a function of the distribution function, $\Theta = \Theta(F)$. For example, the population mean is defined as

$$\Theta(F) = \int u dF(u). \tag{2.1}$$

Analogously, we can define a corresponding estimate of that parameter as $\Theta(\hat{F}) = \Theta(X_1, ..., X_n) = \hat{\Theta}$, where $\hat{F}$ represents the empirical distribution function. For an estimate of the population mean, we would use the sample mean

$$\Theta(\hat{F}) = \int u d\hat{F}(u) = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.2}$$

Often times, the distribution of $\Theta(\hat{F})$ is difficult if not impossible to obtain. In these situations, we can use repeated samples from the original data set to obtain the distribution of $\Theta(\hat{F})$.

Let $X_1^{(*)}, ..., X_n^{(*)}$ represent a simple random sample taken with replacement from the original data $X_1, ..., X_n$. Using the data set $X_1^{(*)}, ..., X_n^{(*)}$ we can obtain an estimate of the population parameter Θ with $\Theta(X_1^{(*)}, ..., X_n^{(*)})$. The estimate of Θ using this procedure is a single bootstrap estimate and $X_1^{(*)}, ..., X_n^{(*)}$ is known as a bootstrap sample.

In order to obtain an estimate for the distribution of $\Theta(\hat{F})$, we must take repeated bootstrap samples. Denote the k^{th} bootstrap sample by $X_1^{(k)}, ..., X_n^{(k)}$ and the corresponding estimate of the population parameter Θ by $\Theta(\hat{F}^{(k)}) = \hat{\Theta}^{(k)}$. Repeat the procedure and obtain B bootstrap samples. From the B samples we obtain a set of estimates for the parameter Θ, $\{\hat{\Theta}^{(1)}, ..., \hat{\Theta}^{(B)}\}$. The distribution of the parameter estimate $\hat{\Theta}$ can be estimated with

$$\hat{F}_{\hat{\Theta}}(x) = \frac{1}{B}\sum_{k=1}^{B} I(\hat{\Theta}^{(k)} \leq x), \tag{2.3}$$

where $I()$ is the indicator function defined as

$$I(A) = \left\{ \begin{array}{ll} 1 & \text{A is true} \\ 0 & \text{A is false} \end{array} \right\}. \tag{2.4}$$

The estimate for the distribution of $\hat{\Theta}$ in equation 2.3 allows us to obtain the mean and standard deviation of our statistic as well as any other relevant measures. The parameter estimate is often taken to be the mean of this distribution. A p-value for testing the hypothesis $H_0 : \Theta = \Theta_0$ versus a two-sided alternative can be obtained as

$$p = 2\min(\hat{F}_{\hat{\Theta}}(\Theta_0), 1 - \hat{F}_{\hat{\Theta}}(\Theta_0)). \tag{2.5}$$

Another way to test this hypothesis is by constructing a 95% confidence interval

$$[\hat{F}_{\hat{\Theta}}^{-1}(0.025), \hat{F}_{\hat{\Theta}}^{-1}(0.975)]. \tag{2.6}$$

If this interval contains the point Θ_0, we would not reject $H_0 : \Theta = \Theta_0$. The two methods are equivalent if the distribution of $\hat{\Theta}$ is symmetric.

The introduction to bootstrap resampling methods given in this section is not meant to be exhaustive. We are simply providing the foundation of resampling techniques so that we may develop methods for comparing providers while controlling for comorbidity factors. For further reference to this topic, consult Efron & Tibshirani (1986), Efron (1982), and Efron & Tibshirani (1993) . For an insurance application see Derrig, Ostaszewski & Rempala (1998).

3 Resampling Techniques for Comparing Two Populations

In this section, we present two applications of the general bootstrap technique for comparing two populations in the presence of covariates. The methods presented can be used in many analysis situations, but we restrict our attention to comparing the effectiveness of providers in lowering the cost of workers' compensation insurance claims.

Assume there are two distinct populations we are interested in comparing. Let the cost of a claim, C, from the two populations have distribution functions $F_1(c)$ and $F_2(c)$ respectively. Define $Z = 1$ if we are in population one and $Z = 2$ if we are in population two. We can rewrite the distribution of claim costs conditional on Z as $F(c|z = 1) = F_1(c)$ and $F(c|z = 2) = F_2(c)$. Further assume that there is a set of extraneous variables in these populations that influence the ultimate claim cost. Denote this set of claim characteristics by $X = \{X_1, ..., X_n\}$. We present two methods to compare the distribution of claim costs in the two populations while removing the effects of the extraneous variables.

3.1 Method 1: Normalized Comparisons for Two Populations

The first method of comparison assumes very little about the structure of the data, but requires X to consist of categorical variables exclusively. Techniques for normalizing populations are used to account for the differences in the distribution of X. The first step in doing normalized comparisons is to write the distribution of claim costs for population one adjusted for the distribution of X in population two. This distribution is written

$$F^{(2)}(C|Z = 1) = \sum_{i=1}^{R} F(C|Z = 1, X = x_i)P(X = x_i|Z = 2). \tag{3.1}$$

where R is the number of different possible values the vector X can represent. If we have 5 categorical variables with 3 levels each, then there will be $R = 3^5 = 243$ possible combinations. Due to this limitation, the methods in this section are limited to only a few covariates. The superscript (2) is added to the distribution function to indicate that it has been normalized to the distribution of X in population two. The corresponding distribution function for population two is

$$F(C|Z = 2) = \sum_{i=1}^{R} F(C|Z = 2, X = x_i)P(X = x_i|Z = 2). \tag{3.2}$$

3.1.1 Comparing Two Population Means

A simple comparison that can be made on the two populations is a comparison of means. We would compare

$$\mu_1^{(2)} = \int uF^{(2)}(du|Z = 1) \tag{3.3}$$

to

$$\mu_2 = \int uF(du|Z = 2). \tag{3.4}$$

A comparison of the two populations can be made using the bootstrap distribution of the statistic $\Theta(F) = \Theta = \mu_1^{(2)} - \mu_2$.

Assume that we have a sample of data from population one, $(C_1, X_1)_i, i = 1, ..., n_1$, and a sample of data from population two, $(C_2, X_2)_i, i = 1, ..., n_2$. To make the comparison of these populations we would resample from each population B times. Let the k^{th} bootstrap sample be denoted as $(C_1, X_1)_i^{(k)}, i = 1, ..., n_1$ and $(C_2, X_2)_i^{(k)}, i = 1, ..., n_2$. To obtain an estimate of Θ we would need to estimate $F(C|Z = 1, X = x_i)$, $F(C|Z = 2, X = x_i)$, and $P(X = x_i|Z = 2)$. Estimates of each of these functions can be obtained with

$$\hat{F}^{(k)}(u|Z = 1, X = x) = \frac{\sum_{i=1}^{n_1} I(C_{1i}^{(k)} \leq u, X_{1i}^{(k)} = x)}{\sum_{i=1}^{n_1} I(X_{1i}^{(k)} = x)}, \tag{3.5}$$

$$\hat{F}^{(k)}(u|Z = 2, X = x) = \frac{\sum_{i=1}^{n_2} I(C_{2i}^{(k)} \leq u, X_{2i}^{(k)} = x)}{\sum_{i=1}^{n_2} I(X_{2i}^{(k)} = x)}, \tag{3.6}$$

and

$$\hat{P}^{(k)}(X = x|Z = 2) = \frac{1}{n_2} \sum_{i=1}^{n_2} I(X_{2i}^{(k)} = x). \tag{3.7}$$

For each bootstrap sample, the estimates in equations 3.5, 3.6, and 3.7 can be used to obtain an estimate of $\hat{\Theta}^{(k)}$ with

$$\begin{aligned} \hat{\Theta}^{(k)} &= \textstyle\sum_{i=1}^{R} \int u\hat{F}^{(k)}(du|Z = 1, X = x_i)\hat{P}^{(k)}(X = x_i|Z = 2) \\ &\quad - \textstyle\sum_{i=1}^{R} \int u\hat{F}^{(k)}(du|Z = 2, X = x_i)\hat{P}^{(k)}(X = x_i|Z = 2). \end{aligned} \tag{3.8}$$

This equation reduces to

$$\hat{\Theta}^{(k)} = \sum_{j=1}^{R} \bar{C}_{1j} \frac{n_{2j}}{n_2} - \bar{C}_2, \tag{3.9}$$

where $\bar{C}_{1j}$ is the mean cost of sample one for the j^{th} category of the vector X and n_{2j} is the number of observations in sample 2 for the j^{th} category of the vector X.

Using the B bootstrap estimates of $\hat{\Theta}^{(k)}$, we can obtain the distribution of $\hat{\Theta}$ using equation 2.3. In addition, confidence intervals and tests of hypothesis can be constructed using the methods described at the end of Section 2.

3.1.2 Comparing Two Populations Percentiles

Since the bootstrap procedure is very flexible with respect to the form of the statistic, we can estimate the distribution of statistics that may otherwise be very difficult to estimate. One example is found when comparing two population percentiles. For example, we may be interested in comparing normalized distribution functions from equations 3.1 and 3.2 for a percentile p. The statistic for this comparison is

$$\Theta(F) = (F^{(2)})^{-1}(p|Z = 1) - F^{-1}(p|Z = 2) \tag{3.10}$$

The k^{th} bootstrap estimate of $\Theta = \Theta(F)$ is obtained from the equations

$$\hat{F}^{(2k)}(C|Z = 1) = \sum_{i=1}^{R} \hat{F}^{(k)}(C|Z = 1, X = x_i)\hat{P}^{(k)}(X = x_i|Z = 2), \tag{3.11}$$

and

$$\hat{F}^{(k)}(C|Z = 2) = \sum_{i=1}^{R} \hat{F}^{(k)}(C|Z = 2, X = x_i)\hat{P}^{(k)}(X = x_i|Z = 2), \tag{3.12}$$

where the estimates indicated on the right-hand side of equations 3.11 and 3.12 are found from equations 3.5, 3.6, and 3.7. Combining equations 3.11 and 3.12 the k^{th} bootstrap estimate of Θ is

$$\hat{\Theta}^{(k)} = (\hat{F}^{(2)})^{-1}(p|Z = 1) - \hat{F}^{-1}(p|Z = 2). \tag{3.13}$$

Using the B bootstrap estimates of $\hat{\Theta}^{(k)}$, we can obtain the distribution of $\hat{\Theta}$ using equation 2.3. With this distribution, confidence intervals and tests of hypothesis can be constructed using the methods described at the end of Section 2.

3.2 Method 2: Bootstrapping Linear Regression

In this section, methods for bootstrapping in a linear regression model are used to control for both continuous and categorical variables within the covariate vector X. We only provide an overview of the topic, for a more detailed description see Freedman (1981) and Freedman & Peters (1984). The techniques used in this section assume that the log of the claim cost in the population follows a linear model

$$\log(c) = \alpha + \gamma I(Z = 2) + X'\beta + \epsilon = R'\eta + \epsilon, \tag{3.14}$$

where ϵ is random error term with distribution function F, $R = (1, I(Z = 2), X')'$ is a $p + 2$ dimensional vector of covariates, and $\eta = (\alpha, \gamma, \beta')'$ is a $p + 2$ dimensional vector of parameters.

We can estimate η with the standard least-squares estimate. Since we do not want to disturb the distribution of X in each population, we resample from the set of residuals $e_i = \log(c_i) - R_i'\hat{\eta}, i = 1, ..., n$. Denote the k^{th} bootstrap sample of the residuals with $\{e_1^{(k)}, ..., e_n^{(k)}\}$. The corresponding bootstrap sample of $c_i, i = 1, ..., n$ is found with $\{\exp(R_1'\hat{\eta} + e_1^{(k)}), ..., \exp(R_n'\hat{\eta} + e_n^{(k)})\}$. Using this setup we allow the distribution of X to remain constant and we reconstruct the bootstrapped values of c_i from the residuals.

With B bootstrap sample obtained from the above procedure, an estimate of η from the k^{th} bootstrap sample is the least squares estimate from the regression of $\{\exp(R_1'\hat{\eta} + e_1^{(k)}), ..., \exp(R_n'\hat{\eta} + e_n^{(k)})\}$ on $\{R_1, ..., R_n\}$. Denote the k^{th} bootstrap estimate of η as $\hat{\eta}^{(k)}$. From the B bootstrap estimates of η we can estimate the distribution of $\hat{\eta}$ using equation 2.3. Confidence intervals and tests of hypothesis can be constructed using the methods described at the end of Section 2.

4 Comparing Providers with Resampling Methods

To demonstrate the use of the preceding resampling methods, we give two examples that analyze subsets of HNC Insurance Solutions' Provider Compare ® Database. The first example compares the quantiles of providers in one network to all other providers while controlling for claim severity. The second example uses bootstrap regression techniques to compare one provider to all other providers while controlling for 13 variables.

4.1 Example 1: Comparing Quantiles

In order to measure the effectiveness of a network of providers, we compare the median, seventy-fifth percentile, and ninety-fifth percentile to all other providers outside of the network. To control for the severity of a claim, a grouping of ICD9 code and NCCI injury type are used. We singled out one network of providers from the rest of the providers and measured the effectiveness of that network in lowering the total cost of a workers compensation claim. We refer to the providers in the network of interest as "Network A" and the remaining providers as "Other Providers"

The distribution of claim costs in the "Other Providers" group in each sample generated in the bootstrap process is normalized to the distribution of claim severity in "Network A" as outlined in section 3.1.2. We computed the median, 75^{th} percentile, and 95^{th} percentile from the distribution determined by each bootstrap sample. Figures 1, 2, and 3 show the bootstrap distributions for the median, 75^{th} percentile, and 95^{th} percentile respectively.

The upper left-hand corners of Figures 1, 2, and 3 show a histogram representing the distribution of the median, 75^{th} percentile, and 95^{th} percentile respectively as calculated from each bootstrap sample of the claim costs in the "Other Providers" group. The upper right-hand corners of each figure demonstrate the same statistic as calculated on the "Network A" providers. The lower-left hand corner of Figures 1, 2, and 3 shows the bootstrap distribution of the difference between the "Network A" group and the "Other Providers" group for the median, 75^{th} percentile, and 95^{th} percentile respectively.

From the graphs shown in Figures 1, 2, and 3, we can conclude that the providers in Network A have significantly lower median claim costs. However, upon closer inspection, the difference of 75^{th} and 95^{th} percentiles are not significant at the 0.05 Type I error rate level. Looking at the distributions, we can see a trend of the two populations towards one another as we approach the 95^{th} percentile. This finding implies that Network A may not be as effective on the more severe claims.

4.2 Example 2: Bootstrap Regression Techniques for Comparing One Provider to All Other Providers

This example is based on specific client feedback about a suspicious provider. We will identify the suspicious provider as Z. Using data from our workers' compensation provider database, lost time claims where provider Z is listed as the primary provider have mean indemnity costs of $10,317. The combined sample without provider Z produces a mean indemnity cost of $7,228. The unadjusted estimate of the increase in indemnity costs associated with provider Z is 100%(10317-7228)/7228=42.7%.

A model for the natural logarithm of total indemnity cost regressed against thirteen claim characteristics and the provider Z dummy variable (1 if provider Z, 0 otherwise) was identified through a standard variable selection/model building process. Predictor variables included body part, nature of accident, cause of accident, industry class code, age, gender, injury type, and a reduction of ICD9 code that indicates body region and injury severity through a ten level variable. We used a bootstrap regression technique as outlined in section 3.2 to compute the distribution of the parameter estimate associated with provider Z. Figure 4 shows a histogram representing the bootstrap distribution of the parameter estimate.

Using the bootstrap hypothesis testing strategy developed in this paper in section 2, we determined that the one-sided p-value for a test of no significant difference between provider Z and all other providers in the database was approximately 0.15. The mean increase in indemnity cost associated with provider Z is 54.5%, and the median increase is 53.8%. The unadjusted statistics suggest that provider Z produces 42.7% higher indemnity outcomes than the remaining providers as a group. With p=0.15 we have only marginal evidence of an effect and would not reject our null hypothesis using traditional significance levels of 0.05, and 0.10. Nonetheless, the bootstrap results provide more compelling evidence that provider Z is worth watching compared to standard, unadjusted statistics.

5 Examples Using Simulated Data

Since the methods demonstrated in the last section were applied to proprietary data, we will show additional examples in this section that utilize a randomly generated data set that can be found in Appendix A. The data set was generated from the linear regression equation

$$Y = 5 + 1.0Z + 1.5G + \epsilon, \tag{5.1}$$

where Z is a random deviate from a normal distribution with mean 5 and variance 1 and represents a continuous covariate, G is a Bernoulli random variable with the probability of success equal to 0.5, and ϵ is a random error term with mean 0 and variance 1. The random variables Z, G, and ϵ were generated independently. In addition a categorical variable, C, was generated from a binomial distribution with two trials and a probability of success equal to 0.5. The variable C was generated independently of all variables. A listing of the data can be found in Appendix A. This equation basically represents two straight line equations between Y and Z with a slope of one and an additive error term. The Equation for Group 0 is Y=5+Z and the equation for Group 1 is Y=6.5+Z.

5.1 Example 1: Comparing the Medians of Two Groups

Using the same techniques applied to the data sets in Section 4, we compare the medians of the two groups defined by G. We use C in this example as the normalization variable. We used 500 bootstrap samples to compare the medians of group 0 and group 1. Table 1 represents the sample and bootstrap estimates of the medians normalized for C. In addition, the standard deviation of the bootstrap estimates is also given. Figure 5 shows a histogram of the difference in medians for each bootstrap sample. From the histogram we can conclude that there is a statistically significant difference between Group 0 and Group 1 since no bootstrap sample had a difference less than or equal to zero. This is consistent with the model given by equation 5.1 that was used in generating the sample. The mean difference is 11.576-9.901=1.675 which compares to the difference in means of 1.5 represented in equation 5.1.

Table 1: Results from Simulated Data

Group	Raw Sample Mean	Bootstrap Mean	Bootstrap Std. Dev.
0	9.893	9.901	0.232
1	11.659	11.576	0.137

5.2 Example 2: Bootstrapping Regression

Following the techniques presented in section 3.2, we fit the regression equation

$$E(Y|G,C,Z) = \beta_0 + \beta_1 G + \beta_2 I(C=0) + \beta_3 I(C=1) + \beta_4 Z \tag{5.2}$$

to the simulated data listed in Appendix A. Table 2 shows the estimates from the initial model fit to the whole data set. We took 200 bootstrap samples from the residuals to generate the bootstrap samples as describes in Section 3.2. For consistency, we present a histogram of the bootstrap estimates for β_1 in figure 6, the estimated effect of the binary variable G. From the histogram, we can conclude that there is a statistically significant difference in Y between Group 0 and Group 1 since none of the bootstrap estimates were less than or equal to zero. This histogram is similar to the results presented in Section 4.2 on provider, where G=1 represents a provider with higher cost claims and G=0 refers to a provider with lower cost claims. The mean bootstrap estimate of β_1 is 1.577 and the standard deviation is 0.181. This is in the same range to the least squares estimates that are shown in Table 2. To make stronger statements on how these quantities relate, a simulation study would be needed.

Table 2: Regression Results from Simulated Data

Parameter	Estimate	Std. Err.
β_0	4.984	0.366
β_1	1.625	0.142
β_2	-0.049	0.202
β_3	0.083	0.178
β_4	0.985	0.069

6 Conclusions

We have demonstrated how bootstrap techniques are applicable to comparing providers in workers compensation insurance. The methods outlined in this paper provide a powerful set of tools for assessing the effectiveness of individual providers or a group of providers. The examples presented in section 3 are only a few methods that can be constructed using bootstrap techniques. Since the distribution of any statistic is attainable using resampling methods, it is possible to construct a wide range of meaningful tests about our populations.

From the examples presented, we can determine the effectiveness of subsets of providers while controlling for various comorbidity factors. While the methods presented here are the most effective methods for retrospective study, a word of caution is in order. If we are unable to represent all of the comorbidity variables, it is possible to show a difference between providers that can be explained from unaccounted for comorbidity variables. For example, suppose factor A is present in provider X 20% of the time and factor A is present in provider Y 60% of the time. If the presence of factor A is associated with higher claim costs, then factor A is a comorbidity factor. Even if provider X and Y perform equally, if we do not take factor A into account in the analysis, we will likely show that provider X is less expensive than provider Y. This example shows the importance of considering all comorbidity factors in the analysis.

One way to avoid incorrect assessment in the analysis of providers is to randomly assign claims to providers to create a balance of the claim characteristics between the two samples. This type of study design requires careful planning and execution. Since random assignment is often impractical or costly, the methods in this paper should be used to best account for the differences that may exist.

A Appendix

The following data set was randomly generated using the methods described in section 5.

Y	G	C	Z	Y	G	C	Z	Y	G	C	Z	Y	G	C	Z	Y	G	C	Z
6.58	0	0	2.64	7.65	0	0	3.73	8.10	0	0	5.56	8.25	0	0	3.14	8.39	0	0	4.79
8.61	0	0	2.91	8.63	0	0	2 13	9.33	0	0	5.15	9.38	0	0	4.20	9.41	0	0	4.20
9.42	0	0	4.07	9.42	0	0	4.13	9.48	0	0	5.86	9.56	0	0	4.38	9.58	0	0	4.83
9.68	0	0	4.99	9.79	0	0	4.69	9.89	0	0	4.20	9.94	0	0	5.17	9.95	0	0	5.15
10.04	0	0	4.45	10.10	0	0	4.85	10.21	0	0	5.38	10.46	0	0	6.32	10.70	0	0	6.24
10.84	0	0	5.56	11.06	0	0	6.83	11.21	0	0	6.10	11.42	0	0	6.00	12.10	0	0	5.86
12.13	0	0	7.10	12.98	0	0	7.28	6.30	0	1	3.08	6.33	0	1	3.42	6.38	0	1	3.54
6.41	0	1	3.13	6.77	0	1	3.87	7.93	0	1	4.71	8.08	0	1	3.44	8.32	0	1	5.17
8.34	0	1	5.68	8.59	0	1	4.15	8.72	0	1	3.25	8.89	0	1	4.80	8.91	0	1	3.66
8.93	0	1	5.30	9.04	0	1	3.59	9.05	0	1	4.74	9.17	0	1	4.76	9.22	0	1	5.41
9.47	0	1	4.34	9 50	0	1	5.95	9.52	0	1	5.02	9.53	0	1	5.28	9.68	0	1	5.05
9.73	0	1	5.30	9.75	0	1	4.18	9.82	0	1	5.35	9.86	0	1	4.22	9.94	0	1	5.22
10.03	0	1	5.22	10.16	0	1	6.27	10 18	0	1	5.54	10.22	0	1	4.85	10.28	0	1	5.63
10.40	0	1	5.59	10.46	0	1	5.24	10.46	0	1	5.09	10.63	0	1	5.10	10.65	0	1	5.50
10.70	0	1	5 84	10.70	0	1	3.45	10.72	0	1	4.18	10.79	0	1	5.10	10.83	0	1	5.59
10.95	0	1	6.10	11.02	0	1	5.67	11.08	0	1	5.44	11.21	0	1	5.13	11.26	0	1	4.43
11.29	0	1	6.70	11.38	0	1	4.86	11.61	0	1	6.56	11.70	0	1	5.79	11.73	0	1	6.85
12 00	0	1	4.76	12.28	0	1	5.95	12.31	0	1	5.16	12.45	0	1	6.05	12.78	0	1	6.28
13.00	0	1	6.16	7.69	0	2	3.81	8.22	0	2	3.59	8.52	0	2	3.98	8.53	0	2	3.32
8.63	0	2	4.70	8.67	0	2	3.37	8.71	0	2	4.07	8.90	0	2	5.08	9.05	0	2	4.81
9.14	0	2	3.85	9.38	0	2	3.84	9.81	0	2	5.14	9.88	0	2	4.71	9.90	0	2	4.81
10.17	0	2	4.29	10.18	0	2	5.34	10.22	0	2	4.70	10.30	0	2	5.00	10.66	0	2	5.93
10.69	0	2	6.90	10.74	0	2	5.14	11.79	0	2	5.96	11.79	0	2	4.90	12.72	0	2	6.94
9.00	1	0	4.08	9.86	1	0	3.73	9.97	1	0	4.64	10.10	1	0	3.49	10.44	1	0	3.65
10.47	1	0	4.46	10.49	1	0	5.19	10.62	1	0	4.34	10.89	1	0	5.50	11.32	1	0	4.43
11.66	1	0	4.81	11 68	1	0	4.77	11.97	1	0	5.33	12.21	1	0	5.28	12.30	1	0	5.76
12.31	1	0	5.47	12.49	1	0	6.54	13.05	1	0	6.55	13.51	1	0	5.85	14.55	1	0	5.99
9.03	1	1	3.61	9.25	1	1	2.99	9.39	1	1	4.85	9.47	1	1	3.96	9.52	1	1	4.96
9.99	1	1	2.94	10.44	1	1	3.94	10.58	1	1	3.51	10.74	1	1	5.53	10.75	1	1	4.24
10.80	1	1	5 45	11.06	1	1	4.68	11.09	1	1	5.11	11.22	1	1	3.32	11.26	1	1	5.40
11.26	1	1	4.22	11.29	1	1	3.43	11.33	1	1	4.00	11.59	1	1	4.37	11.67	1	1	5.46
11.68	1	1	4.78	11.75	1	1	5.64	11.97	1	1	5.60	11.99	1	1	5.01	11.99	1	1	3.98
12.05	1	1	5.41	12.08	1	1	6.01	12.12	1	1	5.92	12.13	1	1	6.84	12.42	1	1	5.40
12.50	1	1	5.03	12.50	1	1	6.12	12.62	1	1	6.66	12.76	1	1	5.67	12.83	1	1	4.39
12.95	1	1	6.55	13.03	1	1	4.81	13.24	1	1	6.46	13 53	1	1	7.10	14.20	1	1	7.17
14.39	1	1	6.53	14.46	1	1	7.94	14.67	1	1	5.44	15.38	1	1	6.56	9.78	1	2	3.80
9.85	1	2	5.14	9.93	1	2	4.89	10.23	1	2	4.57	10.57	1	2	4.09	10.79	1	2	4.31
10.79	1	2	4.91	10.96	1	2	3.88	11.05	1	2	6.15	11.13	1	2	5 16	11.19	1	2	4.76
11.32	1	2	5.35	11.43	1	2	4.49	11.78	1	2	4.69	11.98	1	2	4.55	11.99	1	2	5.84
12.13	1	2	5.29	12.15	1	2	4.24	12.25	1	2	5.34	12.44	1	2	5.50	13.70	1	2	5.35

References

Derrig, R., Ostaszewski, K. & Rempala, G. (1998), 'Applications of Resampling Methods in Actuarial Practice', *CAS Forum, Summer.*

Efron, B. (1982), *The Jackknife, the Bootstrap, and Other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. & Tibshirani, R. (1986), 'Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy', *Statistical Science* **1**, 54–77.

Efron, B. & Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

Freedman, D. (1981), 'Bootstrapping Regression Models', *Annals of Statistics* **9**, 1218–1228.

Freedman, D. & Peters, S. (1984), 'Bootstrapping a Regression Equation: Some Empirical Results', *Journal of the American Statistical Association* **79**, 97–106.

Figure 1: Bootstrap distribution of the Median Claim Cost in Network A and in Other Providers and Bootstrap Distribution of the Difference of Medians

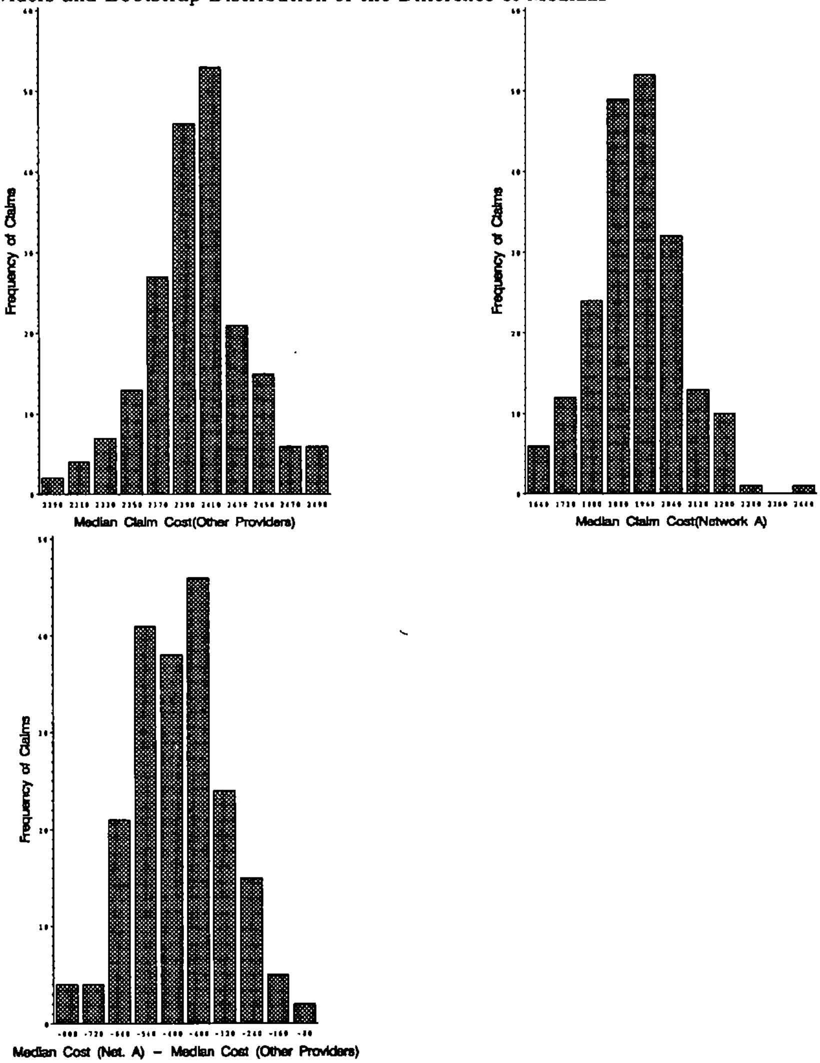

Figure 2: Bootstrap distribution of the 75th Percentile of Claim Costs in Network A and in Other Providers and Bootstrap Distribution of the Difference of 75th Percentiles

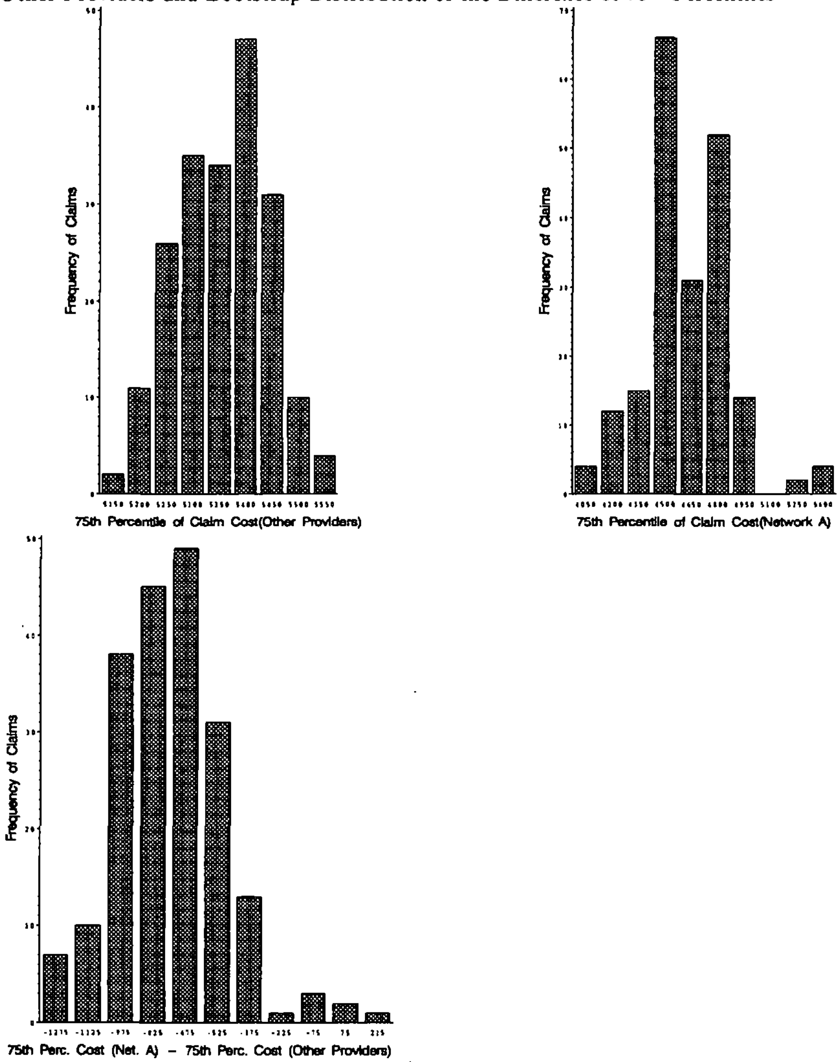

Figure 3: Bootstrap distribution of the 95th Percentile of Claim Costs in Network A and in Other Providers and Bootstrap Distribution of the Difference of 95th Percentiles

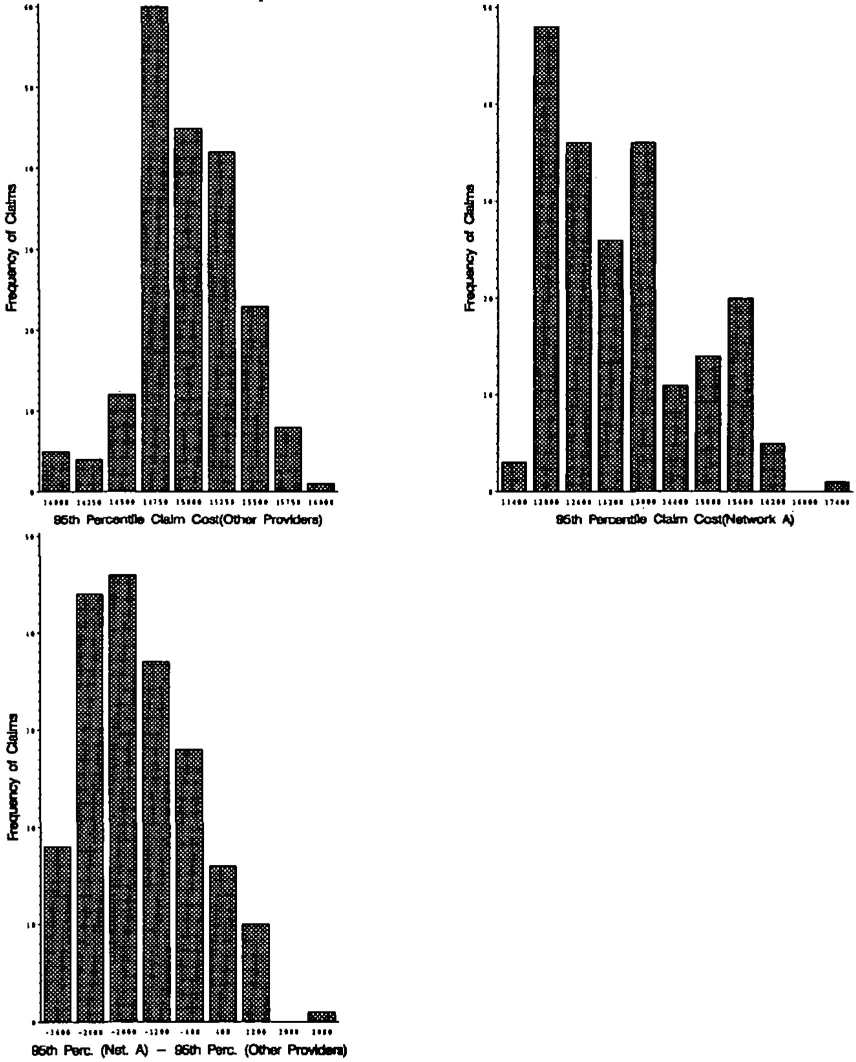

Figure 4: Bootstrap distribution of Regression Coefficient Measuring the Relative Change in Log-Cost for Provider Z versus all other Providers

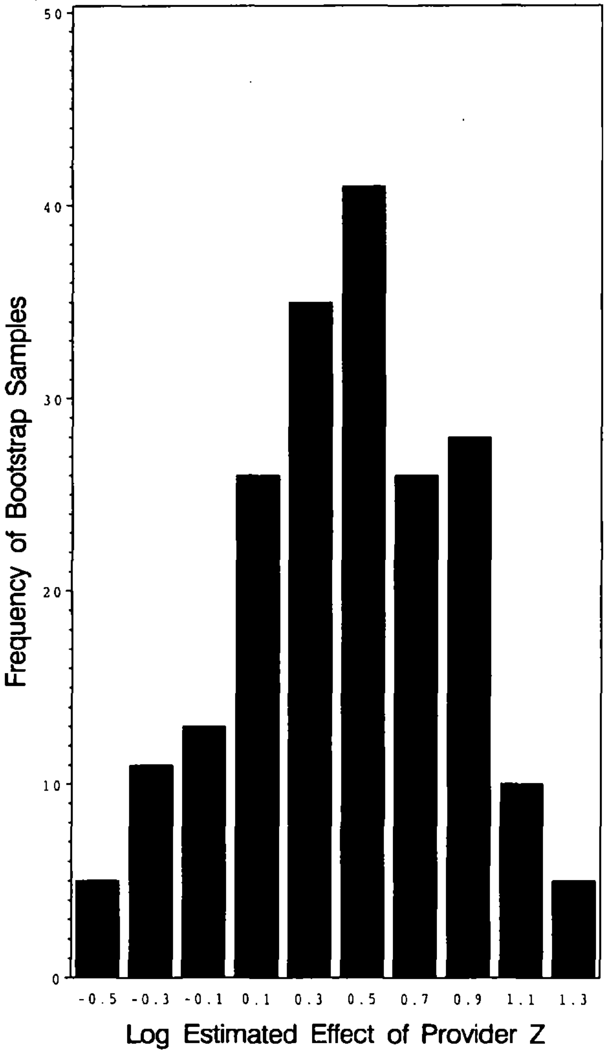

Figure 5: Bootstrap distribution for the Difference of Medians for Group=0 and Group=1

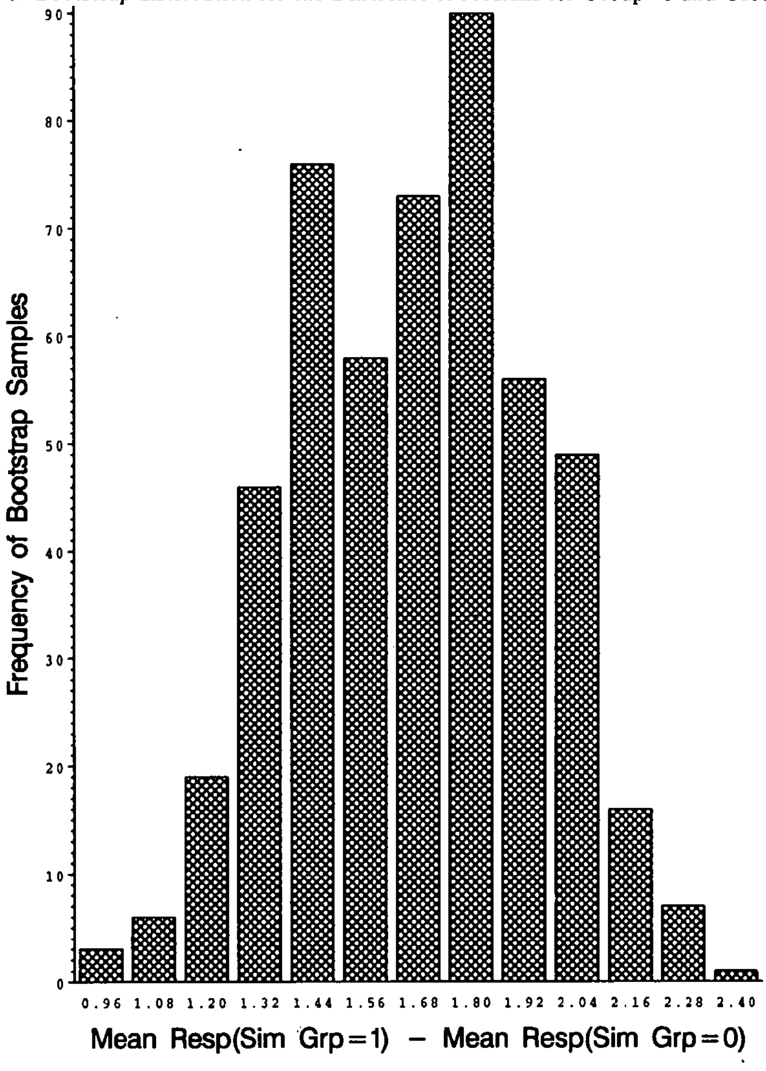

Figure 6: Bootstrap distribution of Regression Coefficient for Simulated Data