**Supplementary Online Materials**

**Appendix:**

**Additional Description of Data and Analyses**

Number of Claims per Employee

The observations for this analysis were the total number of claims and total number of man hours among the affected group for the 2 years prior to and after the implementation of the intervention. Below is an example of the data structure:

|  |  |  |
| --- | --- | --- |
| **Original Data** |  | **Calculated Variables** |
| **Company** | **Period Beginning Date** | **Period End Date** | **Intervention Period** | **N of Claims** | **Total Number of Man Hours** |  | **BegYr** | **Int** | **LN(hrs)** |
| 1 | 7/12/2001 | 7/22/2003 | Before | 11 | 76503 |  | 41.53 | 0 | 11.245085 |
| 1 | 11/29/2003 | 11/27/2005 | After | 0 | 61728 |  | 43.91 | 1 | 11.030493 |
| 2 | 7/15/2001 | 7/15/2003 | Before | 7 | 44265 |  | 41.54 | 0 | 10.69795 |
| 2 | 11/14/2003 | 11/13/2005 | After | 1 | 55221 |  | 43.87 | 1 | 10.919099 |
| ... |  |  |  |  |  |  |  |  |  |

There were three calculated variables used in the regression analysis: BegYr – date of beginning of period expressed as number of years after 1/1/1960 (default in SAS), Int – dummy variable indicating whether the period was pre or post intervention, LN(hrs) – natural log of the total number of man hours for the affected group which was used as the offset parameter. The data was analyzed using Poisson regression with repeated measures where Company was the subject id and the claim rate had the following fixed effects structure:

In this model, exp() is the claim rate ratio representing change in claim rate associated with the passage of 1 year, regardless of intervention and exp() is claim rate ratio with and without the intervention in a given year.

Geometric Average Cost per Claim

The observations for this analysis were the paid cost of each observed claim. Below is an example of the data structure:

|  |  |  |
| --- | --- | --- |
| **Original Data** |   | **Calculated Variables** |
| **Company** | **Date of Injury** | **Paid Cost** |   | **lntotal** | **Date** | **Int** |
| 1 | 7/12/2001 | $170.01 |  | 5.165 | 41.529 | 0 |
| 1 | 7/26/2001 | $118.00 |  | 4.812 | 41.567 | 0 |
| 1 | 8/10/2001 | $172.35 |  | 5.178 | 41.608 | 0 |
| 1 | 10/25/2001 | $317.95 |  | 5.778 | 41.816 | 0 |
| 1 | 11/12/2001 | $81.66 |  | 4.462 | 41.866 | 0 |
| 1 | 2/25/2002 | $76.00 |  | 4.394 | 42.153 | 0 |
| 1 | 2/21/2002 | $423.01 |  | 6.059 | 42.142 | 0 |
| 1 | 7/19/2002 | $982.78 |  | 6.896 | 42.547 | 0 |
| 1 | 8/26/2002 | $105.95 |  | 4.709 | 42.651 | 0 |
| 1 | 4/8/2003 | $83.00 |  | 4.477 | 43.267 | 0 |
| 1 | 3/11/2003 | $1,689.40 |  | 7.435 | 43.191 | 0 |
| 2 | 10/17/2001 | $1,128.23 |  | 7.033 | 41.794 | 0 |
| 2 | 11/19/2001 | $196.24 |  | 5.305 | 41.885 | 0 |
| 2 | 8/2/2001 | $8,020.83 |  | 8.99 | 41.586 | 0 |
| 2 | 8/12/2001 | $541.90 |  | 6.304 | 41.614 | 0 |
| 2 | 11/12/2002 | $338.72 |  | 5.84 | 42.865 | 0 |
| 2 | 5/27/2003 | $672.27 |  | 6.518 | 43.402 | 0 |
| 2 | 6/12/2003 | $743.79 |  | 6.619 | 43.445 | 0 |
| 2 | 2/5/2004 | $250.11 |  | 5.542 | 44.097 | 1 |
| ... |   |   |   |   |   |   |

There were three calculated variables used in the regression: lntotal – LN(Paid Cost + 5) (Note: 5 was added to each of the claims to avoid taking the natural log of a $0 claim, of which there were relatively few), Date – the number of years after 1960 (default in SAS) the injury occurred, Int – a dummy variable indicating if the claim occurred during the intervention period.

The lntotal was assumed to follow a normal distribution, and its mean was assumed to be defined as:

E(lntotal) =

The data was modeled using linear regression with repeated measures with Company as the subject id.

In this model, exp() is the ratio of the geometric means representing a 1 year change, regardless of intervention and exp() is the ratio of the geometric means with and without the intervention in a given year.

Cost per FTE

The observations for this analysis were the total cost of all claims during the two years before and two years after the intervention. Below is an example of the data structure:

|  |  |  |
| --- | --- | --- |
| **Original Data** |  | **Calculated Variable** |
| **Company** | **Period Beginning Date** | **Period End Date** | **Intervention Period** | **Total Paid Cost** | **Total Number of Person Hours** |  | **LN(Cost)** | **zero** | **LN(hrs)** | **BegYr** | **Int** |
| 1 | 7/12/2001 | 7/22/2003 | Before | $4,220.11 | 76503 |  | 8.35 | 0 | 11.25 | 41.53 | 0 |
| 1 | 11/29/2003 | 11/27/2005 | After | $0.00 | 61728 |  | - | 1 | 11.03 | 43.91 | 1 |
| 2 | 7/15/2001 | 7/15/2003 | Before | $11,641.98 | 44265 |  | 9.36 | 0 | 10.70 | 41.54 | 0 |
| 2 | 11/14/2003 | 11/13/2005 | After | $250.11 | 55221 |  | 5.52 | 0 | 10.92 | 43.87 | 1 |
| ... |  |  |  |  |  |  |  |  |  |  |  |

There were four calculated variables: LN(Cost) – natural log of the total paid cost of all claims during the 2 year period,” zero” – a dummy variable indicating if the two year period total costs were $0, LN(hrs) – natural log of total number of person-hours, BegYr – date of beginning of the two-year period, expressed as number of years after 1/1/1960 (default in SAS), Int – a dummy variable indicating if the period was pre or post intervention.

The total paid cost had a large proportion of $0 values (35%) due to the large number of observations that had 0 claims. Therefore, the total paid cost had a mixture distribution of both discrete and continuous distributions and as a result, a 2-part regression model was used as described in (REF).

In the first part of the 2-part regression model, the probability of observing a non-$0 cost during a 2 year period is analyzed using a binomial regression model with repeated measures. The details of estimating probability ratios using binomial regression can be found in (Deddens ref). The second part of the 2-part model focuses only on the non-zero cost observations, and models the average paid cost for a 2-year period using linear regression with repeated measures. Since the paid costs were skewed, the values were log transformed (note: since only the non-zero costs were analyzed, there was no issue of taking the log of 0).

The probability of observing a non-$0 total cost was analyzed with the following model:

Among the nonzero total paid costs observations, it was assumed that the LN(Costs) follow a normal distribution, and its mean was defined as:

E(LN(Cost)|(Cost not equal to 0) =

Once the parameters were estimated from each regression analysis, the results were combined to estimate the expected Cost per FTE in the following way. The conditional average paid cost, given the paid cost was not $0, is calculated from the linear regression model incorporating a smear factor to account for converting from a lognormal distribution. Next, this average is multiplied by the probability that a given paid cost observation is not $0, calculated from the logistic regression model. Finally, this estimated paid cost is divided by the number of FTEs. The effect of the intervention can then be estimated by taking the ratio of the expected cost per FTE when INT=1 and when Int=0, while holding all other variables constant.

Therefore, the effect of the intervention would be estimated as exp()\*exp() for a given year and the effect of 1 year of elapsed time would be exp()\*exp(). To obtain a 95% CI for these estimates, a bootstrap method was used as described in (REF).

Model B

For all three outcomes, a second set of models were run where the time effect was not restricted to a constant trend. Instead, the year of the Period Begin Date was entered as a class variable in each model, replacing the BegYr variable.