

Dirty Data

Anathema to Best Estimates

Louise Francis
Simon Sheaf

Casualty Loss Reserve Seminar
Chicago
15 September 2009

Working Party Members

- Robert Campbell
- Louise Francis (chair)
- Virginia R. Prevosto
- Mark Rothwell
- Simon Sheaf

Agenda

- Context
- Literature Review
- Horror Stories
- Survey
- Experiment
- Actions
- Concluding remarks
- Questions and discussion

Agenda

- **Context**
- Literature Review
- Horror Stories
- Survey
- Experiment
- Actions
- Concluding remarks
- Questions and discussion

Cost of Poor Data

- Olson: 15% - 20% of operating profits
- Insurance Data Management Association (IDMA): Cost to US economy is \$600bn a year
- The IDMA believes that the true cost is higher than these figures reflect, as they do not allow for "opportunity costs of wasteful use of corporate assets."

PWC Survey

- Almost 50% of respondents did not believe that senior management placed enough importance on data quality
- Only 18% were very confident in the quality of data shared with third parties
- On average, respondents thought that data represented 37% of value of their companies
- Where data improvement initiatives were undertaken, significant returns on the investment are realized

Agenda

- Context
- **Literature Review**
- Horror Stories
- Survey
- Experiment
- Actions
- Concluding remarks
- Questions and discussion

Literature Review

- The vast majority of the available literature on data quality is directed towards the IT industry
- However, the following sources of information are more actuary- or insurance-specific:
 - Actuarial Standard of Practice #23: Data Quality
 - Insurance Data Management Association (IDMA)
 - CAS White Paper on Data Quality
 - Data Management Educational Materials Working Party
 - Board for Actuarial Standards Exposure Draft: Data (UK)

Actuarial Standard of Practice #23

- The American standard for all practice areas developed by the Actuarial Standards Board
- Provides descriptive standards for:
 - selecting data
 - relying on data supplied by others
 - reviewing and using data
 - making disclosures about data quality
- http://www.actuarialstandardsboard.org/pdf/asops/asop023_097.pdf
- In the UK, the Board for Actuarial Standards is working on an actuarial standard on data

Insurance Data Management Association

- The IDMA is an American organisation which promotes professionalism in the Data Management discipline through education, certification and discussion forums
- The IDMA web site:
 - suggests publications on data quality
 - describes a data certification model
 - contains Data Management Value Propositions which document the value to various insurance industry stakeholders of investing in data quality
- <http://www.idma.org>

CAS White Paper on Data Quality

- Developed by the Casualty Actuarial Society's Committee on Management Data and Information
- Provides guidelines to satisfy ASOP 23
- Describes a system of standardised procedures to insure the integrity of statistical data for personal automobile
- <http://www.casact.org/pubs/forum/97wforum/97wf145.pdf>

CAS Data Management Educational Materials Working Party

- Reviewed a shortlist of texts recommended by the IDMA for actuaries (9 in total)
- Publishing a review of each text in the CAS *Actuarial Review* (starting with the August 2006 issue)
- Combined the reviews into an actuarial introduction to data management
- This was published in the Winter 2007 CAS *Forum*
- Both the reviews and the final paper are available through www.casact.org

Literature Review Summary

- Standards are generally prescriptive but descriptive information is available
- www.idma.org and www.casact.org are good sources for more information, containing papers and other information in addition to those reviewed in the paper

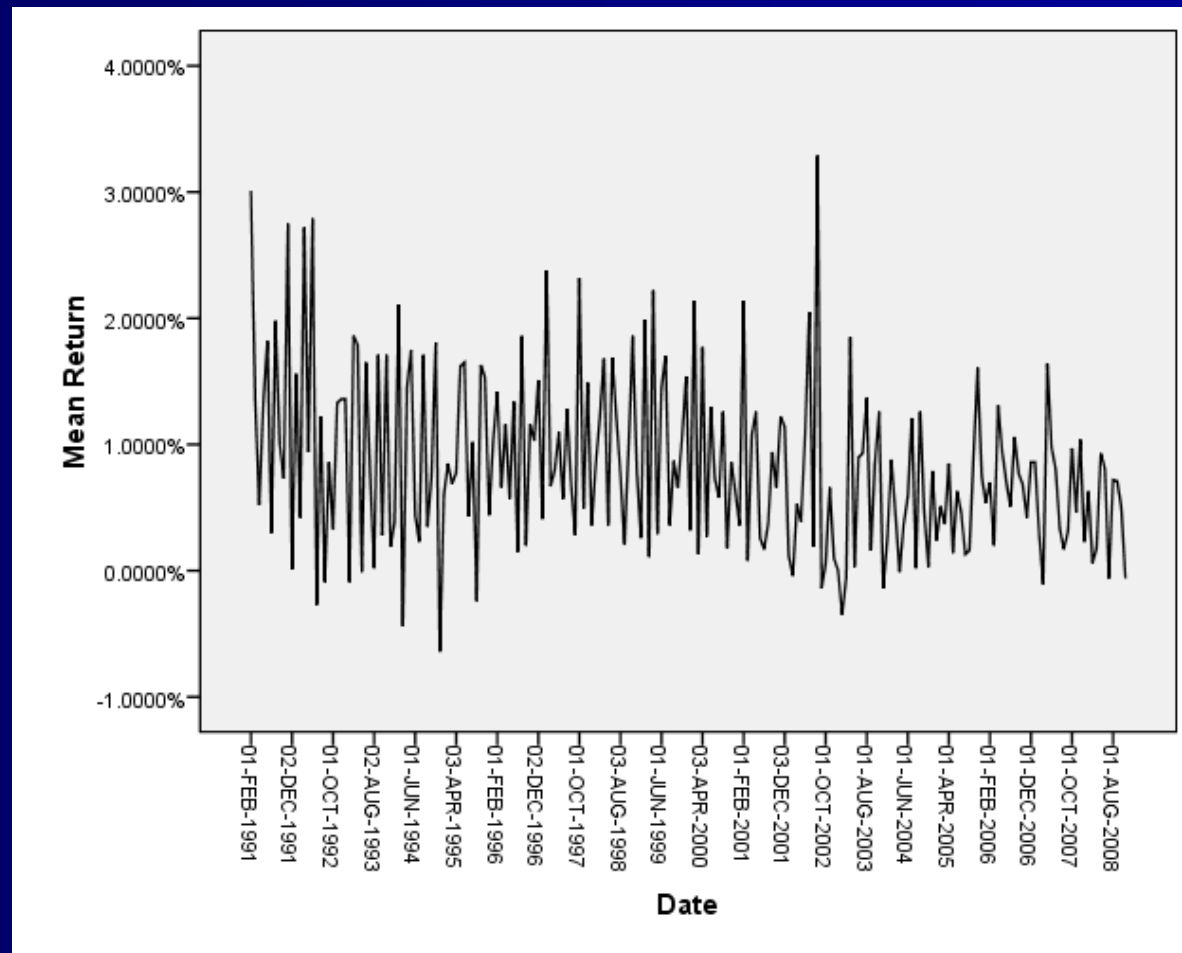
Agenda

- Context
- Literature Review
- **Horror Stories**
- Survey
- Experiment
- Actions
- Concluding remarks
- Questions and discussion

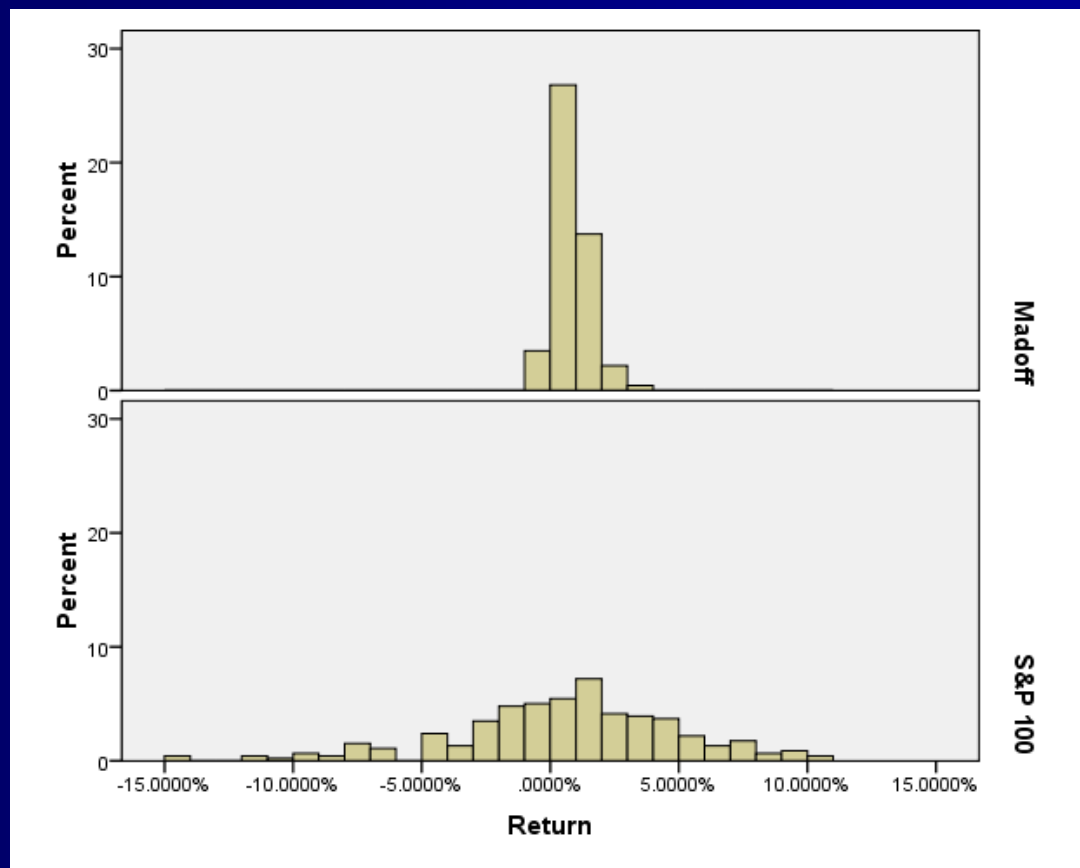
Horror Stories – Non-Insurance

- Heart-and-Lung Transplant – Wrong blood type
- Bombing of Chinese Embassy in Belgrade
- Mars Orbiter – Confusion between units of measurement
- Fidelity Mutual Fund – Withdrawal of dividend
- Porter County, Illinois – Tax bill and budget shortfall

Horror Stories – Madoff Fund - Fake Data



Madoff Fund - Real Return vs Fake Data



Madoff Fund - Descriptive Statistics for Various Assets

Asset	Mean	Std. Deviation	Skewness	N
Balanced Fund	0.46%	2.84%	(0.89)	149
Long Bond Fund	0.60%	2.40%	(0.36)	149
S&P 100	0.31%	4.77%	(0.47)	149
S&P 500	0.30%	4.61%	(0.70)	149
Madoff Feeder Fund	0.75%	0.62%	1.01	149

Note: Data from July, 1996 to Oct, 2008

Horror Stories - Independent Insurance

- Independent Insurance collapsed in June 2001
- A year earlier, Independent's market valuation had reached £1bn
- Independent's collapse came after an attempt to raise £180m in fresh cash by issuing new shares failed because of revelations that the company faced unquantifiable losses
- The insurer had received claims from its customers that had not been entered into its accounting system, which contributed to the difficulty in estimating the company's liabilities

Horror Stories - Reserving

- NAIC concerns over non-US country data
- Canadian federal regulator uncovered:
 - inaccurate accident year allocation
 - double-counted IBNR
 - claims notified but not properly recorded
- Former US regulator – requirement for reconciliation exhibits in actuarial opinions motivated by belief that inaccurate data being used

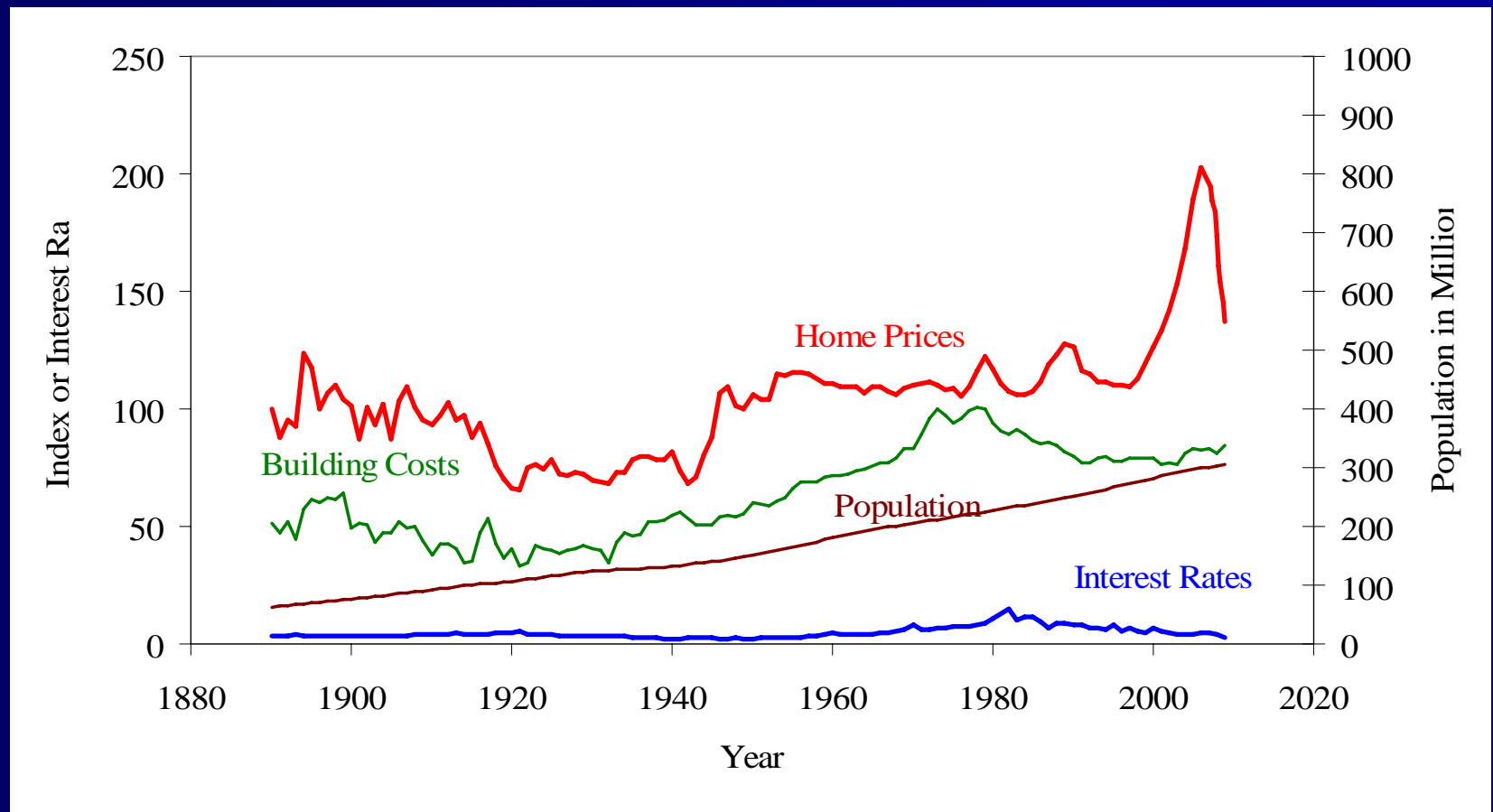
Horror Stories – Rating/Pricing

- Examples faced by ISO:
 - Exposure recorded in units of \$10,000 instead of \$1,000
 - Large insurer reporting personal auto data as miscellaneous and as a result it was omitted from ratemaking calculations
 - One company reporting all its Florida property losses as fire (including in years when they had significant hurricane losses)
 - Mismatched coding for policy and claims data

Horror Stories - Katrina

- US catastrophe models underestimated costs of Katrina by approximately 50% (Westfall, 2005)
- 2004 RMS study highlighted exposure data that was:
 - out-of-date
 - incomplete
 - mis-coded
- Many flood victims had no flood insurance after being told by brokers that they were not in flood risk areas

Possible Error in Assumptions: Do US House Prices Go Down?



Agenda

- Context
- Literature Review
- Horror Stories
- **Survey**
- Experiment
- Actions
- Concluding remarks
- Questions and discussion

Survey of Actuaries

- Purpose: Assess the impact of data quality issues on the work of general insurance actuaries
- Two questions:
 - What percentage of time is spent on data quality issues?
 - What proportion of projects are adversely affected by such issues?

Targeted Distribution

- Members of the Working Party
- Members of CAS Committee on Management Data and Information
- Members of CAS Data Management and Information Educational Materials Working Party
- Working Party members each contacted a number of additional people
- This resulted in 76 responses

Results of Survey

Question 1: Percentage of Time Spent on Data Quality Issues

Employer	Number of Responses	Mean	Median	Minimum	Maximum
Insurer/Reinsurer	40	25.0%	20.0%	2.0%	75.0%
Consultancy	17	26.9%	25.0%	5.0%	75.0%
Other	17	29.6%	25.0%	1.0%	80.0%
All	74	26.5%	25.0%	1.0%	80.0%

Results of Survey

Question 2: Percentage of Projects Adversely Affected by Data Quality Issues

Employer	Number of Responses	Mean	Median	Minimum	Maximum
Insurer/Reinsurer	40	32.5%	20.0%	3.5%	100.0%
Consultancy	17	37.6%	30.0%	5.0%	100.0%
Other	17	35.4%	25.0%	1.0%	100.0%
All	74	34.3%	25.0%	1.0%	100.0%

Survey Conclusions

- Data quality issues have a significant impact on the work of general insurance actuaries:
 - about a quarter of time is spent on such issues
 - about a third of projects are adversely affected
- The impact varies widely between different actuaries, even those working in similar organizations
- Limited evidence to suggest that the impact is less significant for company actuaries

Agenda

- Context
- Literature Review
- Horror Stories
- Survey
- **Experiment**
- Actions
- Concluding remarks
- Questions and discussion

Hypothesis

The uncertainty of actuarial claim reserve estimates based on poor quality data is significantly greater than those based on better quality data

Data Quality Experiment

- Examine the impact of incomplete and/or erroneous data on the actuarial estimates of ultimate claims and claim reserves
- Use real data with simulated limitations and/or errors and observe the potential error in the actuarial estimates

Data Used in Experiment

- Real data for primary private passenger bodily injury liability business for a single no-fault state
- 18 accident years of fully developed data, so actual ultimate claims are known

Actuarial Methods Used

- Paid claim methods
 - Chain Ladder
 - Bornhuetter-Ferguson
 - Berquist-Sherman closing rate adjustment
- Incurred chain ladder method
- Inverse power curve for tail factors
- No judgment used in applying methods

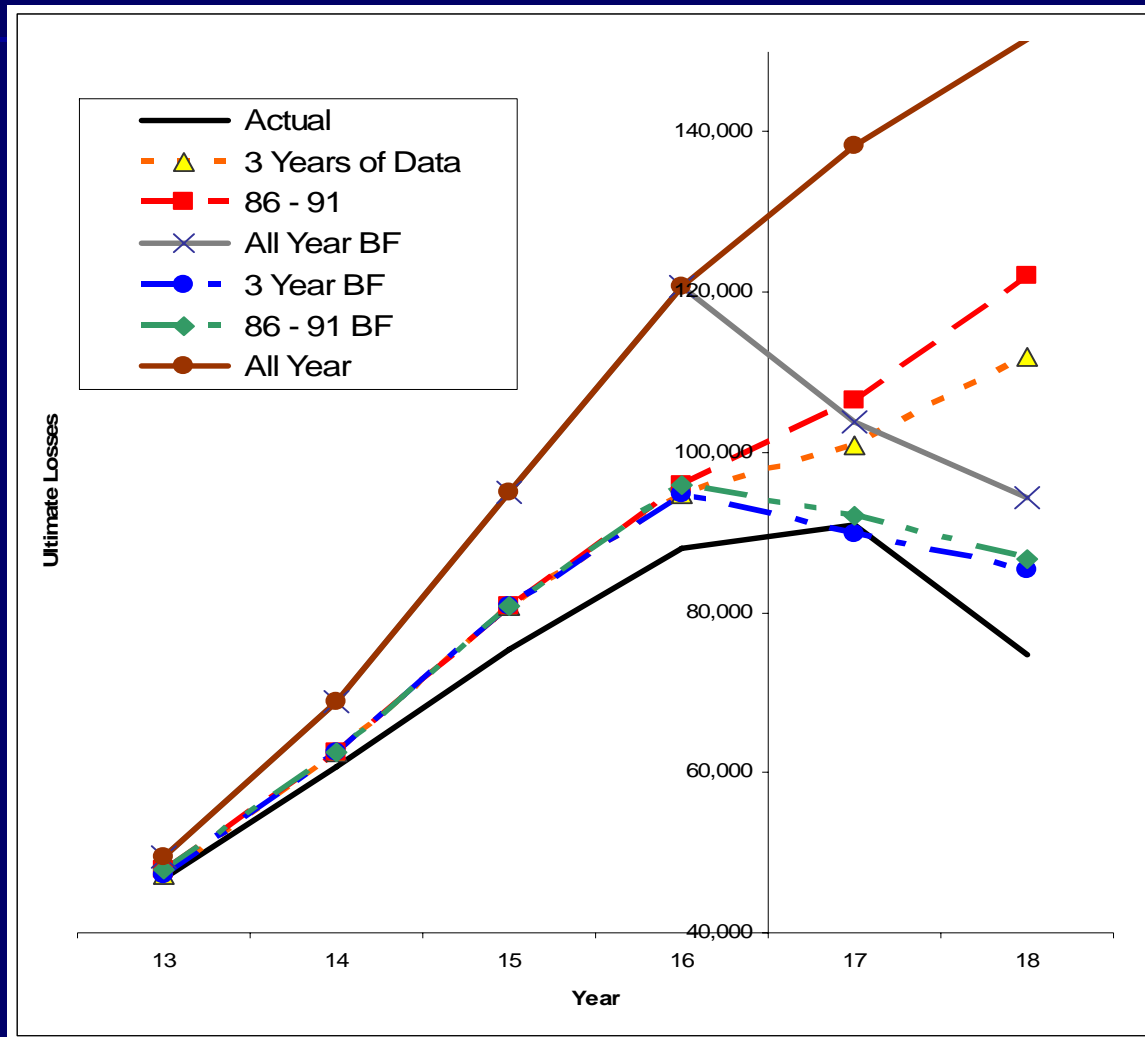
Measure Impact of Data Quality

- Compared estimated and actual ultimate claims
- Used bootstrapping to evaluate the uncertainty of the results

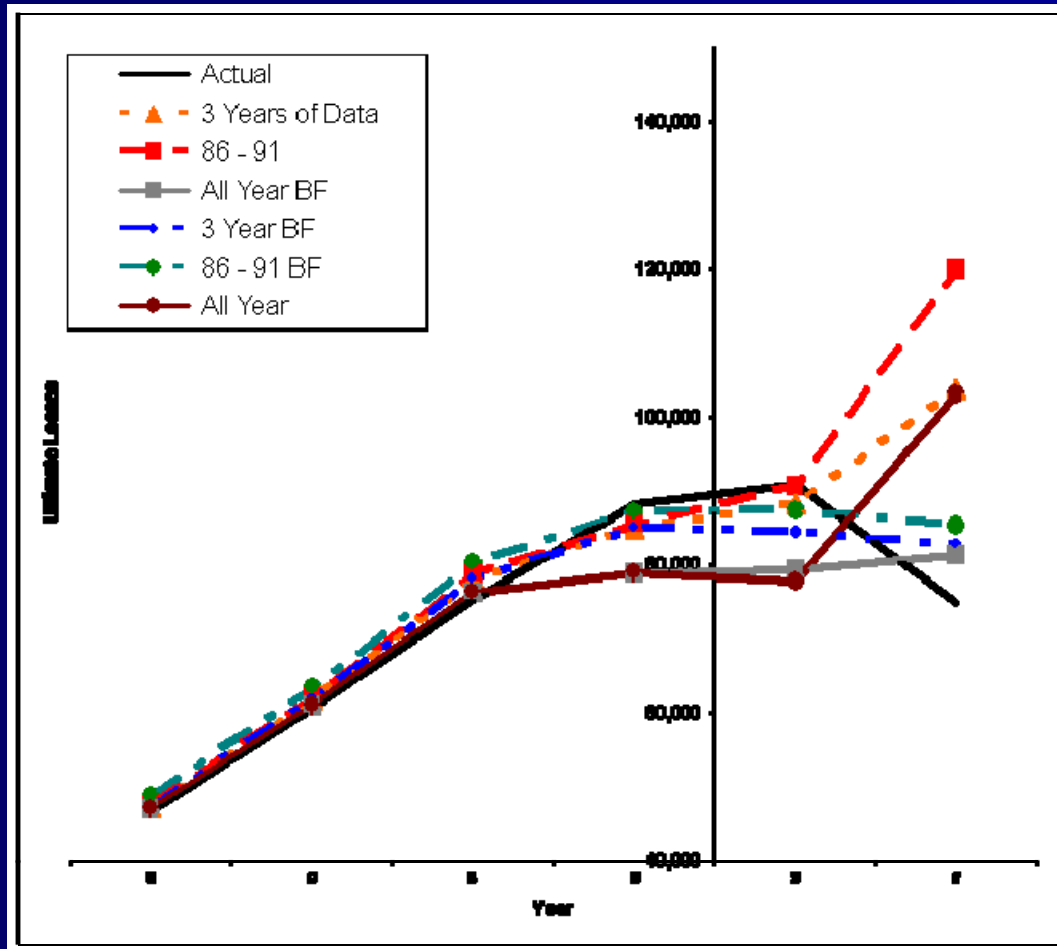
Varying Quantity of Data

- Varied completeness of the data
- Three scenarios:
 - use all accident years and diagonals
 - use only 6 accident years
 - use only last 3 diagonals

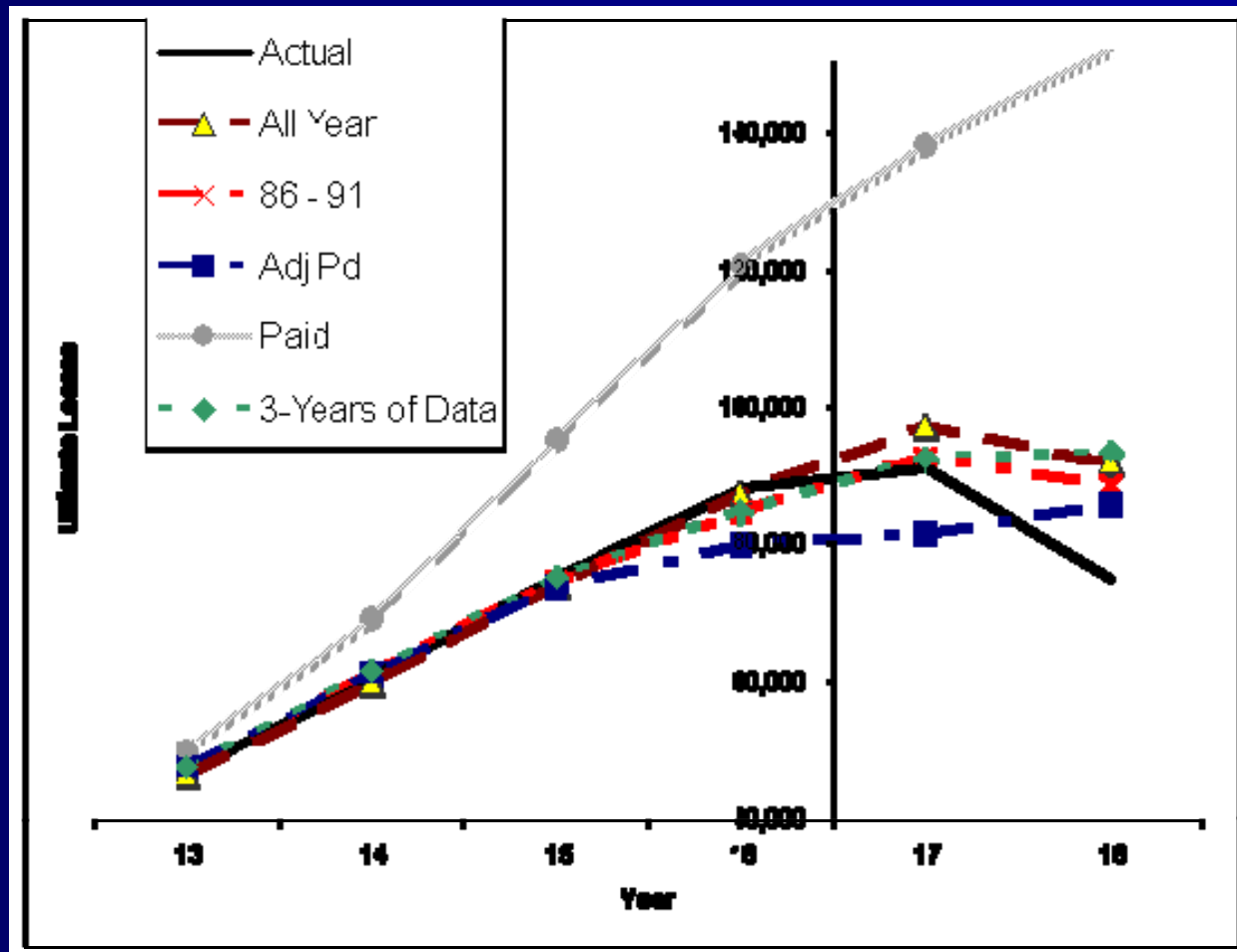
Estimated Ultimate Claims Based on Paid Claims



Estimated Ultimate Claims Based on Adjusted Paid Claims



Estimated Ultimate Claims Based on Incurred Claims



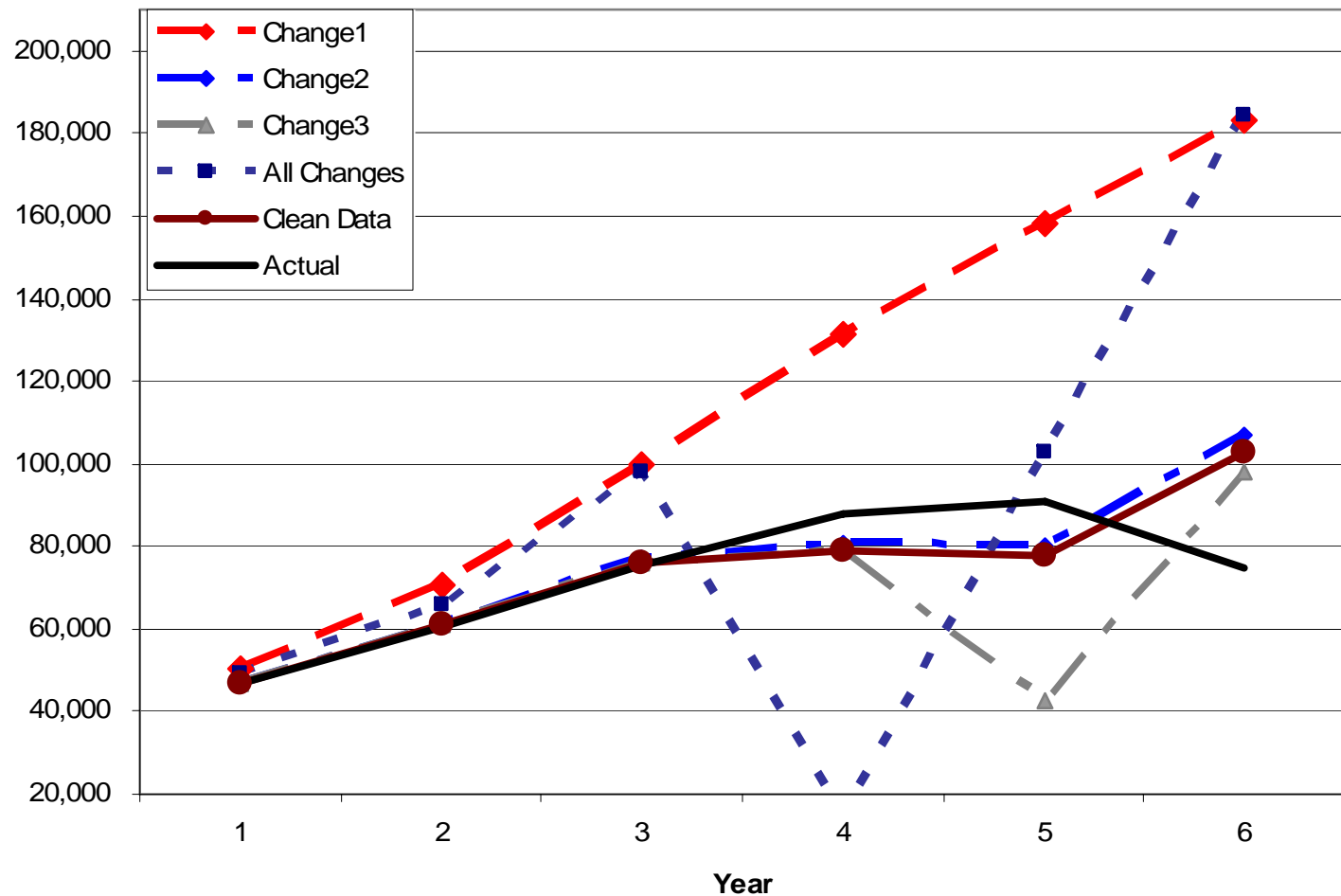
Results of Adjusting Quantity of Data

- More data generally leads to better estimates

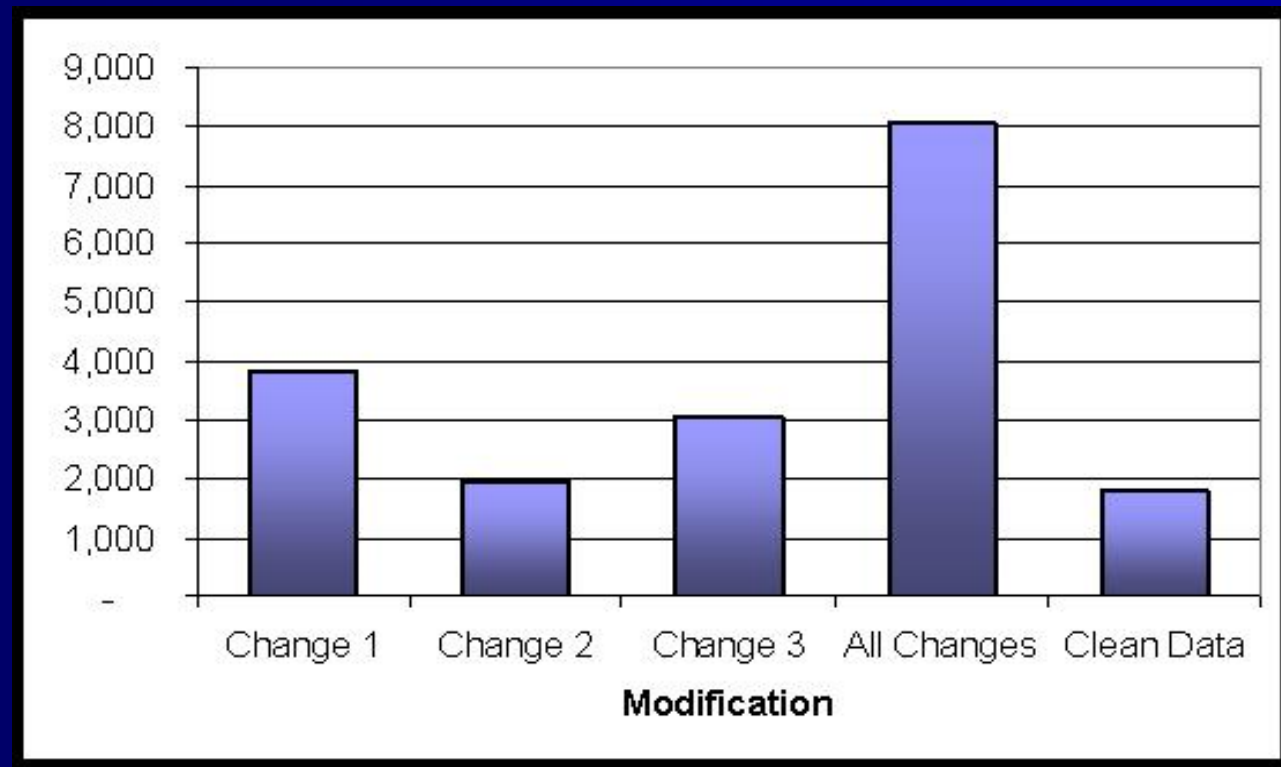
Introducing Data Errors

- Six simulated data errors:
 - Misclassification of losses by accident year
 - Early years not available
 - Late processing of financial information
 - Paid losses replaced by crude estimates
 - Overstatements followed by corrections in following period
 - Definition of reported claims changed
- Applied first three errors individually and all six errors together

Estimated Ultimate Claims Based on Adjusted Paid Claims



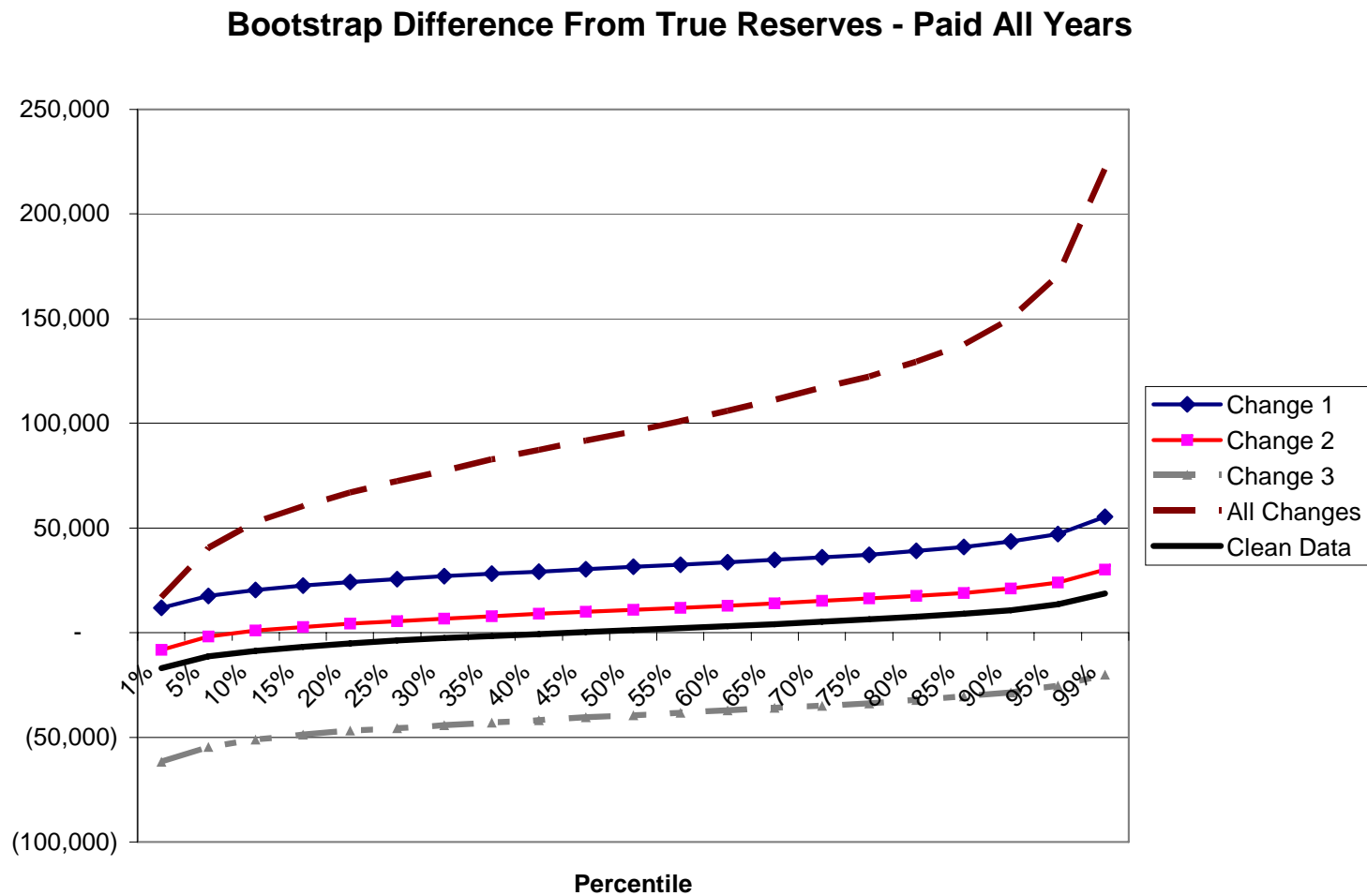
Standard Errors for Adjusted Paid Claims Data



Results of Introducing Errors

- Extreme volatility, especially those based on paid data
- Actuaries ability to recognise and account for data quality issues is critical
- Actuarial adjustments to the data may never fully correct for data quality issues

Distribution of Reserve Errors



Results of Bootstrapping

- Less dispersion in results for error free data
- Standard deviation of estimated ultimate claims is greater for the data with errors
- Confirms original hypothesis that errors increase the uncertainty of estimates

Experiment Conclusions

- Generally greater accuracy and less variability in actuarial estimates when:
 - quality data used
 - greater number of accident years used
- Data quality issues can erode or even reverse the benefits of increased volumes of data
 - If errors are significant, more data may worsen estimates due to the propagation of errors for certain projection methods
- Significant uncertainty in results when:
 - data is incomplete
 - data has errors

Agenda

- Context
- Literature Review
- Horror Stories
- Survey
- Experiment
- **Actions**
- Concluding remarks
- Questions and discussion

Actions – What can we do?

- Data quality advocacy
- Data quality measurement
- Management issues
- Screening data

Data Quality Advocacy - Examples

- The Casualty Actuarial Society:
 - Data Management and Information Committee
 - Data Management and Information Education Materials Working Party

Data Quality Measurement Ideas

- Quantify traditional aspects of quality data such as accuracy, consistency, uniqueness, timeliness and completeness using a score assigned by an expert
- Measure the consequences of data quality problems
 - measure the number of times in a sample that data quality issues cause errors in analyses, and
 - the severity of those errors
- Use measurement to motivate improvement

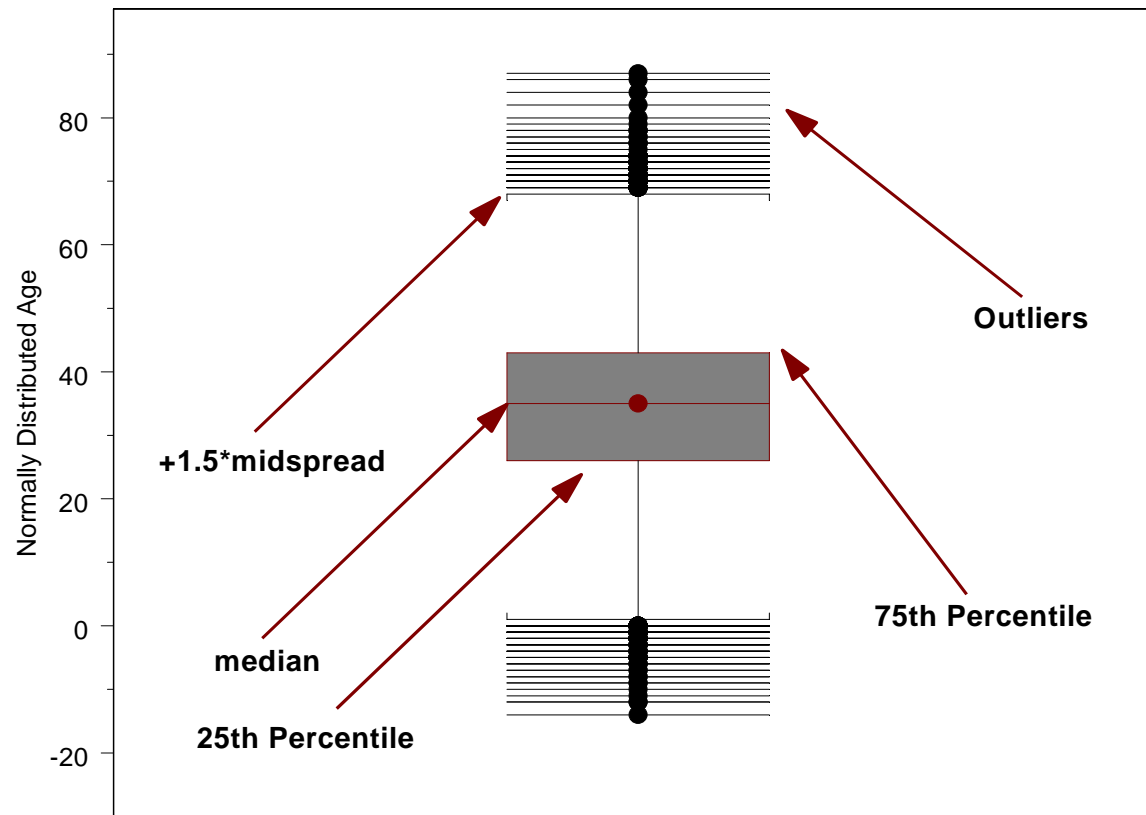
Management Issues

- Redman : Manage Information Chain
 - Establish management responsibilities
 - Describe information flow
 - Understand customer needs
 - Establish measurement system
 - Establish control and check performance
 - Identify improvement opportunities
 - Make improvements

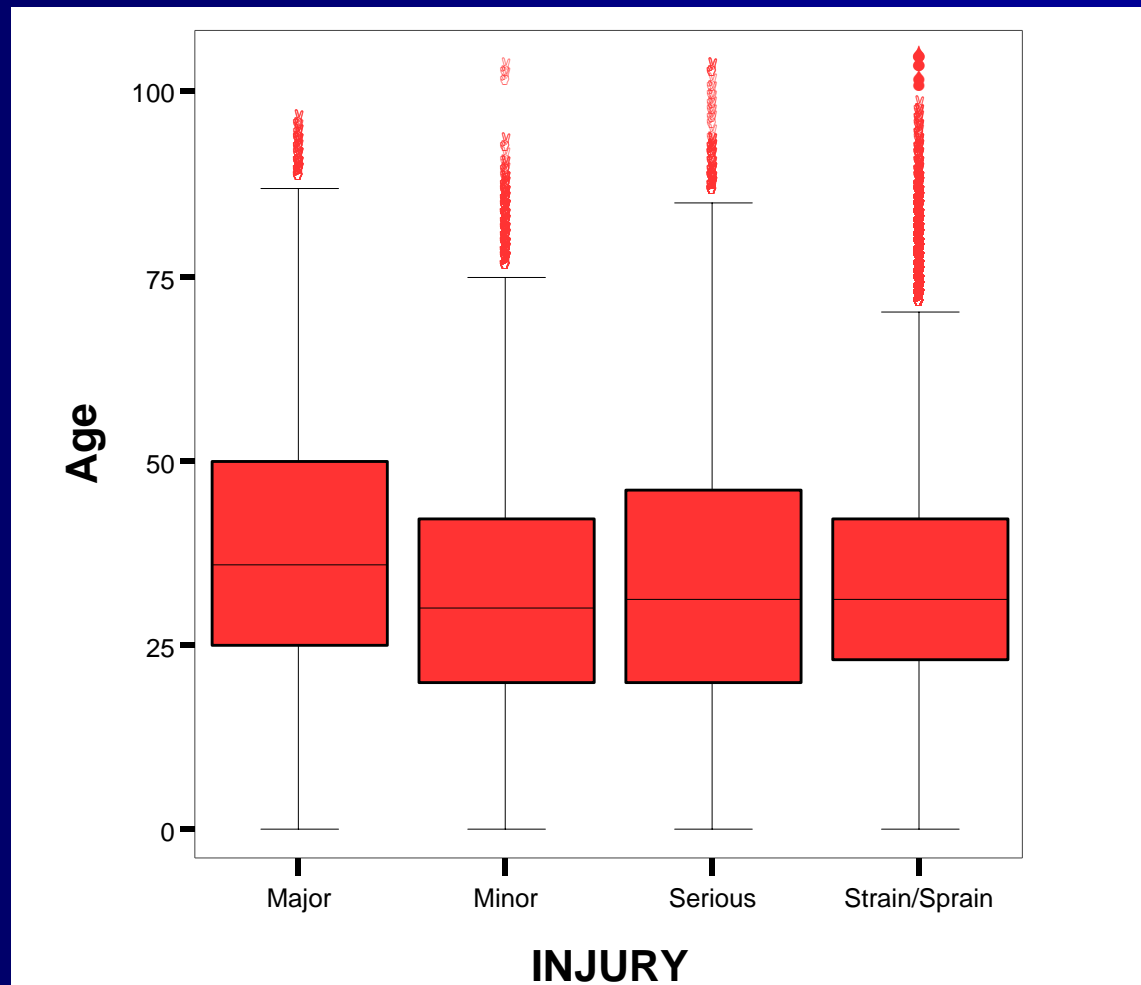
Management Issues

- Data supplier management
 - Let suppliers know what you want
 - Provide feedback to suppliers
 - Balance the following:
 - Known issues with supplier
 - Importance to the business
 - Supplier willingness to experiment together
 - Ease of meeting face to face

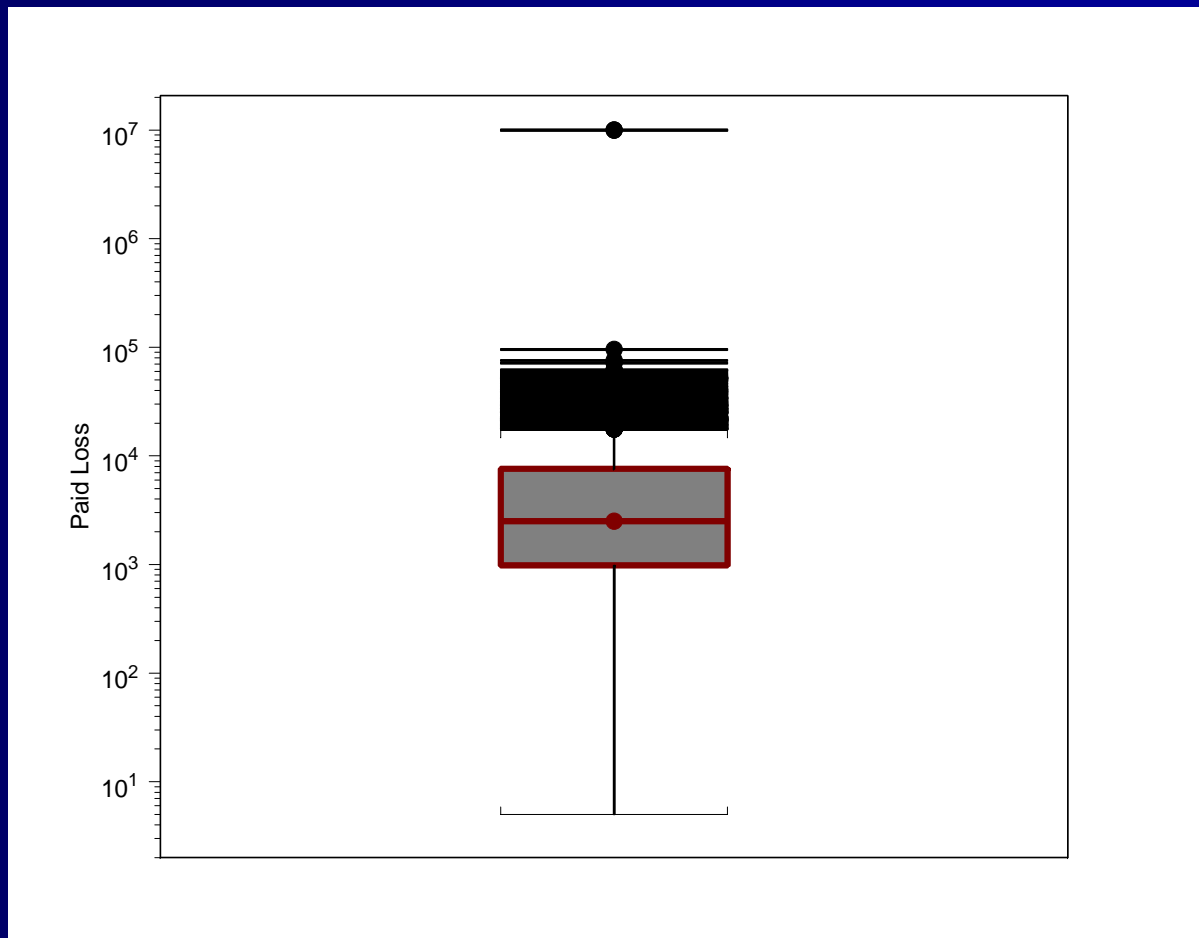
Screening Data – Box and Whisker Plot



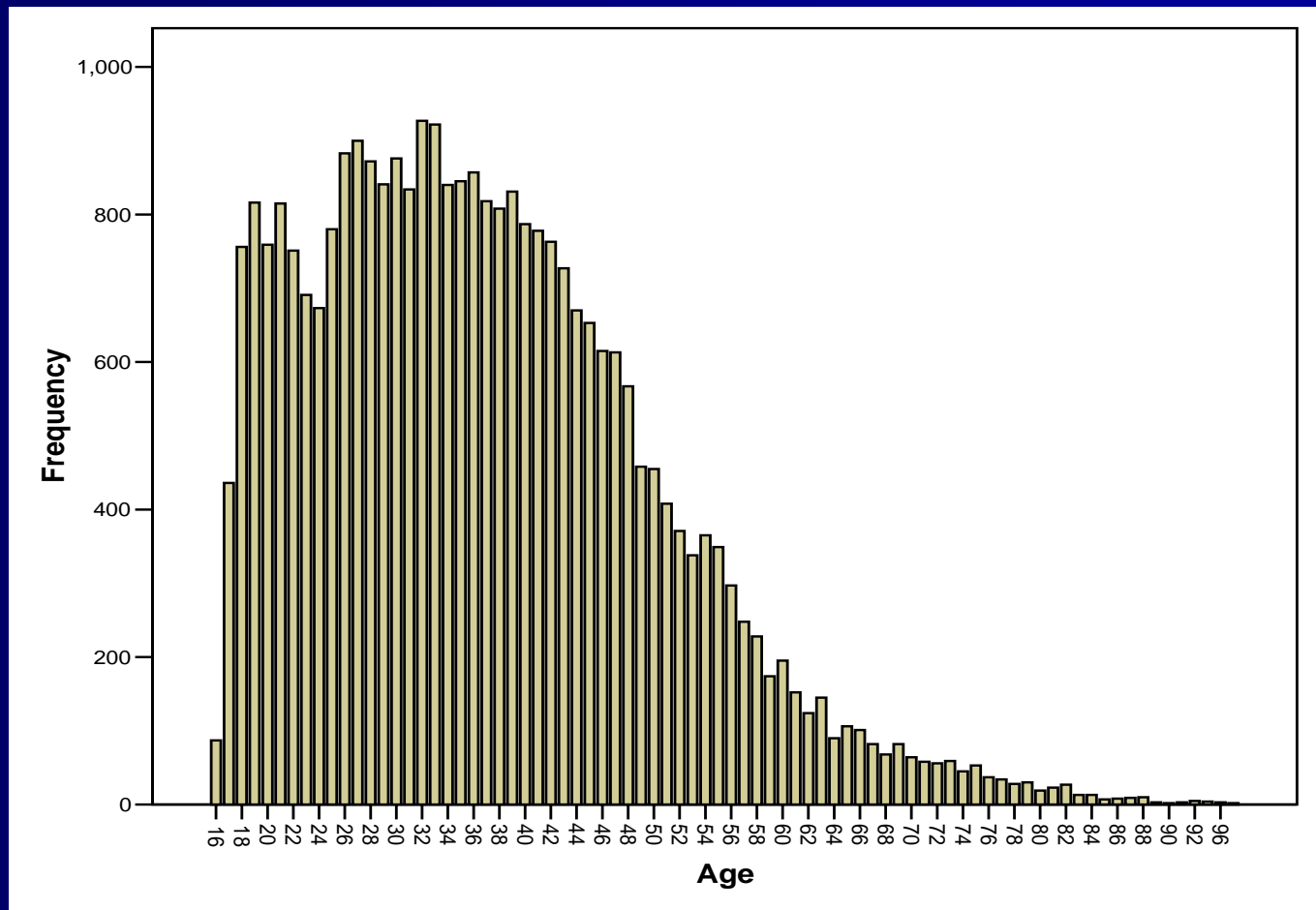
Box and Whisker Plot by Category - Age by Injury



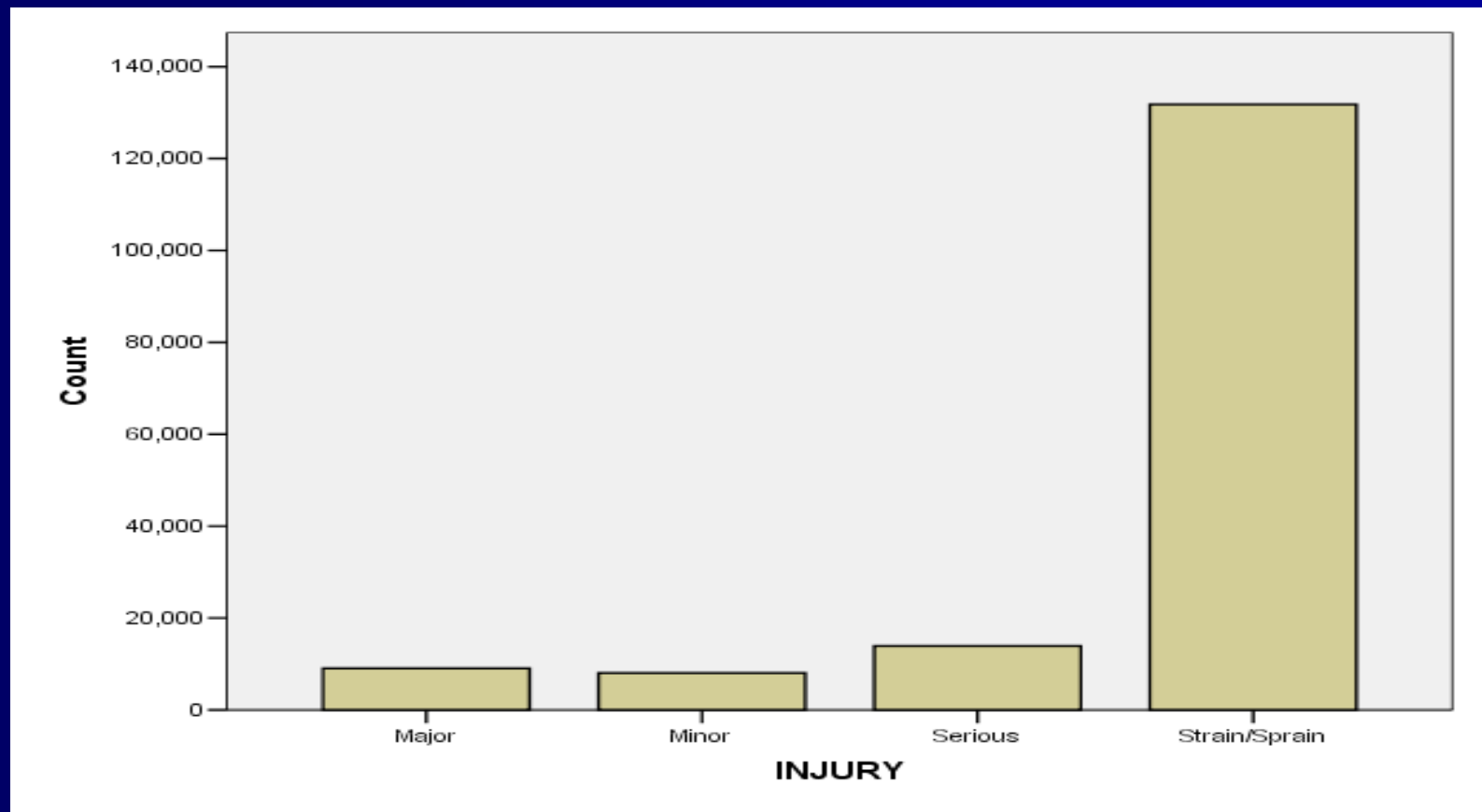
Box and Whisker Plot with Outlier



Screening Data – Histogram



Screening Data – Bar Plot by Category



Screening Data - Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
License Year	<i>30,250</i>	<i>490</i>	<i>2,049</i>	<i>1,990</i>	<i>16.3</i>
Valid N	<i>30,250</i>				

Multivariate Methods

$$\mathbf{MD} = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

\mathbf{x} is a vector of variables

$\boldsymbol{\mu}$ is a vector of means

$\boldsymbol{\Sigma}$ is a variance-covariance matrix

MD is Mahalanobis Distance

Agenda

- Context
- Literature Review
- Horror Stories
- Survey
- Experiment
- Actions
- **Concluding remarks**
- Questions and discussion

Conclusions

- Anecdotal horror stories illustrate possible dramatic impact of data quality problems
- Data quality survey suggests data quality issues have a significant cost on actuarial work
- Data quality experiment shows data quality issues have significant effect on accuracy of results

Conclusions

- The Working Party believes that insurers should devote more time and resources to increasing the accuracy and completeness of their data by improving their practices for collecting and handling data
- In particular, insurers would benefit from the investment of increased senior management time in this area
- By taking such action, they could improve both their profitability and their efficiency.

Questions and Discussion

The paper can be found at:

www.casact.org/pubs/forum/08wforum/Dirty_Data.pdf

Louise Francis

Francis Analytics

+215 923 1567

louise-francis@msm.com

Simon Sheaf

Grant Thornton UK LLP

+44 20 7728 3280

simon.h.sheaf@gtuk.com