Report of the Data Quality Working Party

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

This paper is produced by the GIRO Data Quality Working Party. It is intended to communicate the Working Party views, findings and conclusions to other members of our profession. In no way should anything in this paper be regarded as the views of GIRO, the Institute of Actuaries, the Casualty Actuarial Society or the employers of any working party member. Nor should anything in this paper be regarded as best practice for actuaries in non-life insurance.

Abstract
The Data Quality Working Party was formed because of the view that data quality issues significantly impact the work of general insurance actuaries; and such issues could have a material impact on the results of general insurance companies.

In this paper we will
- describe some actual data quality disasters in non-insurance and insurance businesses;
- present the results of a data quality survey of practicing actuaries in the United States, Canada and Great Britain;
- present the results of a data quality experiment where data was altered to change its quality and the effect on analyses using the data were observed; and
- provide advice on what can be done to improve the state of data quality.

"In just about any organization, the state of information quality is at the same low level"
- Jack Olson, *Data Quality*

I Introduction

In its 2005 report the GRIT (General Insurance Reserving Issues Task Force) working party of GIRO (General Insurance Research Organization) observed that actuaries need to "Facilitate a debate on how to improve data quality available, to support reserving, both claims data and pricing information. Many actuaries and many of our stakeholders think this is weak"[1] The GRIT report also recommends that the actuaries test the reasonableness of their data[2].

Data quality is an important issue affecting all actuaries. Whether one is reserving, pricing, modeling or performing other functions, virtually all actuaries encounter data that is either incomplete or inaccurate. Recently enacted laws in both Europe (Basil II) and the United States (Sarbanes-Oxley) addressing record keeping issues would seem to justify more attention to data quality, but a general increase in concern about data quality is not obvious. The data quality working party was constituted to act as a catalyst to the profession and the industry to improve data quality practices.

In this paper we will

- recount some data quality disasters in non-insurance businesses
- provide data quality "horror stories" from the insurance industry
- present the results of a data quality survey of practicing actuaries in the United States, Canada and Great Britain
- present the results of a data quality experiment where data was intentionally altered to change its quality and the effect on analyses using the data were observed
- provide advice on what can be done to improve the state of data quality

1. Background on Data Quality

The actuarial literature on data quality is relatively sparse. The GRIT working party report recommended more focus on data quality (Copeman et al., 2006) and suggested that UK professional guidance notes incorporate standards from Actuarial Standards of Practice (ASOP) 23. Furthermore the GRIT survey found that many respondents expressed concern

---

[1] Copeman et al., p. 13
[2] Copeman et al., p 17

over data quality.

The American Academy of Actuaries (AAA) Standard of Practice #23 on data quality provides a number of guidelines to actuaries when selecting data, relying on data supplied by others, reviewing and using data and making disclosures about data quality. The guidelines advise actuaries to review data for reasonableness and consistency. The actuary is also advised to obtain a definition of data elements in the data, to identify questionable values and to compare data to the data used in a prior analysis. The actuary is also advised to judge whether the data is adequate for the analysis, requires enhancement or correction, requires subjective adjustment, or is so inadequate that the analysis cannot be performed.

The Casualty Actuarial Society (CAS) committee on Management Data and Information and the Insurance Data Management Association (IDMA) also produced a white paper on data quality (CAS Committee on Management Data and Information, 1997). The white paper states that evaluating the quality of data consists of examining the data for:

- Validity,
- Accuracy,
    - absolute accuracy
    - effective accuracy
    - relative accuracy (inaccurate by consistent over time)
- Reasonableness,
- Completeness.

A typical actuarial review of data consists of balancing totals from the data underlying the analysis to published financial reports and inspecting the data for obviously erroneous values, such as negative amounts for financial variables like paid losses. The data quality white paper describes a number of more extensive activities that could be performed to assure the overall integrity of the data systems serving all the different business users within an insurance company. These include data edits (or checks) to detect impermissible values in the data and periodic data audits to measure the extent of data quality problems. Since actuaries typically use data supplied by others, the white paper advises actuaries to review the extent of checking done by the providers and include their findings in documentation.

Many of these activities described by the white paper are aimed at data managers and IT professionals, as well as actuaries, who are responsible for upstream data that is the original source of the data used by actuaries. The white paper describes the Statistical Data

Monitoring System, a system of standardized procedures used by statistical agents and insurance companies to insure the integrity of data used in statistical filings for personal automobile that are required in the United States. This regulatory requirement was adopted by three states – Connecticut, Rhode Island and New York – in the early 1980s. While only adopted by three states, the requirements applied to the countrywide data of any company writing personal auto insurance in any of the three states. The system includes.

- process description and review of control procedures
- detailed data verification via sampling tests
- summary data verification via reasonability reviews
- financial reconciliation
- annual review and certification
- review and evaluation by state examiners on a periodic basis

As the actuarial analysis in the rate level filing most likely rely on the same data, these data quality procedures also help to assure the quality of that data.

Data quality is core area of concern for the IDMA and their web site, www.idma.org, provides a number of resources on data quality. The IDMA web site provides suggested publications on data quality and describes a data certification model. Another resource is the *Data Management Value Proposition*, where the IDMA documents the value to the insurance industry of investing in data quality. A core belief of the IDMA is that "Data may be the most important resource of the insurance industry" (IDMA, Value Proposition, General Information). The IDMA notes that technological innovation has permitted the insurance industry to integrate data from many sources and share data with customers and other stakeholders via web sites. However, making more data available may not always be beneficial to a company if the quality of the data is suspect. As the IDMA notes, inaccurate data may expose companies to liability. In the next section we review the literature published on the cost of data quality (or lack of it).

Concurrently, the topic of data quality is of increasing importance to the CAS. The CAS has a task force charged with finding educational material on data quality for its membership. As part of that effort, the task force is working on raising the awareness of the profession on the data quality issue via seminars and conferences.

2. Background on the Cost of Poor Data Quality

In the literature on data quality there is a virtually universal agreement that poor data quality imposes a significant cost on companies and on the economy. For instance, Moore predicts that there is a significant likelihood that a data quality error will cause the downfall of at least one large corporation (Moore, 2006). In this section we summarize some of the published findings with respect to the magnitude and cost of data quality problems.

There are various rules of thumb found in the literature concerning the cost of poor data quality. Both the IDMA and Olson cite an estimate that data quality problems cost companies 15% - 20% of operating profits[3]. The IDMA value proposition also [4] cites an estimate that poor data costs the US economy $600 billion a year. The IDMA believes that the true cost is higher than these figures reflect, as they do not depict "opportunity costs of wasteful use of corporate assets." (IDMA Value Proposition – General Information).

According to Eckerson, in many customer databases, 2% of records per month become obsolete because of deaths and address changes (Eckerson). In addition to this, data entry, merging data from different systems, etc. contribute many additional errors. Eckerson mentions that most organizations overestimate the quality of their data. "On one hand, almost half of the companies who responded to our survey believe the quality of their data is "excellent" or "good." Yet more than one-third of the respondent companies think the quality of their data is "worse than the organization thinks."". Eckerson also cites a study done by The Data Warehouse Institute that indicates that data quality is a leading cause of problems when implementing CRM (Customer Relationship Management) systems (46% of survey respondents in a 2000 survey selected it as a challenge).

According to Wand and Wang, 60% of executives from 500 medium sized surveyed firms reported data quality problems.

Poor data quality can also have credibility consequences and motivate regulatory intervention to curb the use of some information deemed important by corporations. In property and casualty insurance in the United States, the use of credit information in underwriting and pricing insurance is a very controversial practice. A key argument of consumer groups opposed to the use of credit is the poor data quality of credit data. Among actuaries who price and reserve small (self insured or alternative market) accounts, there is a general belief that the quality of data from Third Party Administrators is perhaps worse that that of insurance companies. Popelyukhin (1999) reviewed the loss runs of 40 TPAs and

---

[3] Olson, p9.
[4] This citation is apparently from a study done by The Data Warehouse Institute

concluded that no TPA provided data has satisfied his data quality definition (similar to that in the CAS-IDMA white paper above).

In 2004 PricewaterhouseCoopers (PricewaterhouseCoopers, 2004) distributed a data management survey to executives at 450 companies in the US, UK and Australia. The following results were cited by PricewaterhouseCoopers:

- almost half of all respondents do not believe that senior management places enough importance on data quality.
- board level concern about data quality declined since the previous survey (in 2001). That is, the frequency of discussing data quality at board meetings declined between 2001 and 2004.
- only 18% of respondents whose organizations share data with third parties are very confident in the quality of that data
- organizations' confidence in their own data has actually fallen since 2001
    - This was a subjective estimate, as only 45% of companies actually measured the quality of their data
- on average respondents thought data represented 37% of the value of their company
    - only 15% actually measured the value of data to their company
- the survey indicated that when data improvement initiatives were undertaken and when their value was measured, significant returns on investment were realized

II Data Quality Horror Stories

1. Non-insurance Industry Stories

As the anecdotes below illustrate, data errors can result in very serious consequences. In some cases the result is serious embarrassment. In other cases, the result is a large financial loss. In yet other cases, loss of life results, demonstrating that data quality can be a matter of life and death. Many of the most highly publicized data quality horror stories are from non-insurance industries. It should be noted that non-insurance industry errors sometimes have implications for insurance as they may result in errors and omissions or medical malpractice claims as in the first example below.

- A 17-year old Mexican girl received a heart-lung transplant at Duke University

Hospital in South Carolina. She soon fell into a coma as it was discovered that the organs she received were of the wrong blood type (Archibald, 2003). Apparently none of the medical personnel at the hospital performing the transplant requested or verified that proper documentation of a match in blood types was provided. A subsequent transplant with organs of the correct blood type failed and the girl died.

- During the conflict in Bosnia, American pilots accidentally bombed the Chinese embassy in Belgrade as a result of faulty information. "It was the result of neither pilot nor mechanical error," Cohen and Tenet stated. "Clearly, faulty information led to a mistake in the initial targeting of this facility. In addition, the extensive process in place used to select and validate targets did not correct this original error" (CNN, 1999a)
- NASA lost a $125 million Mars orbiter when it came too close to Mars and its engines weren't able to function properly. This was due to one of the project teams using the imperial measurement system rather than metric. (CNN, 1999b)
- Fidelity, a large mutual fund company, had to withdraw a dividend to the shareholders of its flagship Magellan mutual fund (at the time the largest fund in the world) when it was discovered that a capital loss of $1.3 billion had inadvertently been recorded as a gain (NY Times, 1995).
- In Porter County, Illinois, a house worth a little over $100,000 was accidentally valued at $400 million. This caused the county to bill the owner $8 million for what should have been a $1,500 real estate tax bill. Due to the glitch, the county significantly overestimated its tax revenue and experienced significant budget shortfalls
- Approximately 4,000 students received the wrong test scores on a SAT college entrance examination in October, 2006 (Arenson, 2006). The scores were too low by as much as 100 points (in one case a discrepancy of 300 points was noted). On the SAT exams, a student can score up to 800 points in each of the math and verbal sections, and entrance exam scores are believe to follow a normal distribution. Thus a discrepancy of 100 points can drop the student into a significantly lower percentile of the population of students applying to colleges. For some students, admissions decisions were made on the basis of the faulty test scores.

2. Insurance Industry Stories

Although we contacted a number of insurance regulators, we are not at this time aware of any insolvency that resulted primarily from data quality errors. On the other hand, there is a lot of sentiment that data quality often deteriorates badly after insolvency occurs and that it significantly impairs the quality of post-insolvency estimates of liabilities. It is possible that the role of data quality issues in insolvencies is obscured by other management issues.

a. Reserving stories

- In June 2001, The Independent went into liquidation and became the UK's largest general insurance failure. A year earlier, it'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.
- The National Association of Insurance Commissioners[5] stated that it often cannot rely on typical domiciliary country data when reviewing the condition of alien (non - US) insurers. However, they indicated that when they request data from the companies themselves, it is usually supplied (Otis, 1977)
- The Canadian federal regulator (the Office of the Superintendent of Financial Institutions, or OSFI for short) has uncovered instances of:
    - Inaccurate accident year allocation of losses and double-counted IBNR loss estimates (i.e., the actuary calculated IBNR from triangles that already included IBNR)
    - Sometimes claims were reported after a company is insolvent and it is discovered that the original notices (sometimes from years before) were not properly recorded in the company's systems
- In the US, actuaries providing statements of actuarial opinion to insurance regulators concerning the adequacy of reserves for an insurance company are required to supply an exhibit balancing totals from data used in their actuarial analysis to totals in the statutory financial statement. A former regulator indicated this requirement is motivated by disclaimers in opinions letters (i.e., the data was supplied by the company and responsibility for its accuracy was deemed to be theirs) and concerns that invalid data would be used in the actuary's reserve analyses. Beginning with 2004 annual statements, the auditors were required to obtain an understanding of the data and data elements that the Appointed Actuary would rely on in forming their

[5] An association of state insurance regulators in the United States

opinion. The auditor has the responsibility for considering testing such data in the statutory financial statement audit. (COPLFR, 2004). Because "write-ins" and aggregate reserves are often not included in the normal data management systems used in reserving, and because of other systems and accounting idiosyncrasies, significant effort is sometimes expended on the data reconciliation requirement. In addition, the level of data quality checking in the opining actuary's report is somewhat cursory and falls short of that suggested by the CAS/IDMA white paper and the AAA standards of practice.

b. Ratemaking Data Quality Anecdotal Stories

We uncovered a number of instances where data quality issues impacted ratemaking data. Advisory organizations in the United States such as the National Council on Compensation Insurance (NCCI) for workers compensation and the Insurance Services Office, Inc. (ISO) for most of the remaining property/casualty lines of insurance devote significant resources to finding and correcting errors in data.

The stories below are a just a few examples of data anomalies that have been faced by ISO over the years in its role as an advisory organization along with other examples drawn from the consulting community. These are cases where the anomaly was found during the rate level experience review and caused extra expense to either correct the error or remove the company's experience from the rate level experience review. It is not a complete list but rather gives a flavor of the data quality glitches that typically occur.

- A company reported its homeowners exposure (the amount of insurance on the dwelling) in units of $10,000 instead of units of $1,000. Since the exposure was understated by a factor of 10, applying current manual base loss costs (or manual rates) and rating factors to the exposure would have resulted in greatly understated aggregate loss costs at current manual level (or aggregate premium at present rates). Therefore the experience loss ratio (= incurred losses / aggregate loss costs at manual level) and the statewide rate level indication would have been overstated.
- One of the ten largest insurers in a state reported all of its personal auto data under a miscellaneous coverage code. Since miscellaneous coverage code data are excluded from the rate level review for the core coverages, this would have had a significant effect on ratemaking results if it had not been detected.
- A company reported all its homeowners losses as fire in the state of Florida. It is evident what this error can do for any homeowners rate level review especially when the experience period included the hurricane-heavy accident years of 2004 and 2005.

- Another common error occurs when the premium and loss records for the same policy are not coded identically for the common fields. For example, a company may code all their general liability premium records to the composite rated subline, but the corresponding general liability loss records are coded to another subline. This is commonly known as a premium-loss mismatch error. A recent occurrence of this type of anomaly in homeowners affected about 25% of a company's book of business.

In workers compensation, a multi-state examination conducted in 2001 found that about 11% of unit reports were submitted late, impacting the timeliness of data. A sample of unit reports noted discrepancies between data submitted to NCCI and insurers' internal data in 42% of sampled cases. In addition, at a more detailed level (i.e., by class) a 16% error rate was observed. Although the severity of these errors was not mentioned, they can affect the accuracy of rates as discussed above in the ISO examples.

- At NCCI about one half of the companies require some form of follow-up after initial data screening.
- If an error is deemed to be significant and cannot be corrected, the data is excluded from ratemaking calculations, which can impact the credibility of the sample and cause the results to be biased.
- The NCCI data used in ratemaking is also probably biased because it excludes data from large deductible programs (which is a significant proportion of the workers compensation exposure).

c. Hurricane Katrina Data Disasters

Data glitches affecting both insurers and policyholders impacted by Hurricane Katrina are among the most highly publicized data disasters of 2005.

- In October of 2005, one publication (Westfall, 2005) estimated that the models that are used widely in the US to price weather related catastrophes may have underestimated the cost of Katrina by 50%. The quality of exposure information was believed to play a key role in model underestimates. A 2004 study by RMS indicated that exposure data was often out of date. In addition, exposure data was often incomplete or miscoded (RMS, 2005).
- The impact of exposure quality problems has been complicated by the rapidly rising real estate values in much of the United States, causing many properties to be undervalued. This can affect the estimated cost of any loss event, not just

catastrophes.

- Though the storm and the breaking of the levees inflicted significant damage on their houses, many residents of New Orleans surprisingly did not have flood coverage and their homeowners policies covered only wind and not flood losses (Cornrjo, 2006). A number of these policyholders had been told by their agents that they did not need to buy flood coverage, as they were not in a flood zone. Apparently this faulty information was based on maps that had not been updated in decades.

d. Credit Scores

One of the most controversial issues in general insurance in recent years is the use of credit scores in underwriting and pricing insurance policies. In the US, many state regulators are under pressure to ban the practice. Thus quality problems in data gathered and distributed by the providers of credit information provides a powerful argument against the use of credit data. Research performed by a consumer organization (Consumer Federation of America) found:

- Frequent discrepancies between different credit rating companies' scores for a given consumer
- A high rate (33% of a sample) of missing information relating to positive credit data (accounts paid off)
- A high error rate (43% of a sample) concerning accounts late by at least 30 days

e. Conclusion

The anecdotes in this section indicate that data quality is a significant problem, both within the insurance industry and outside of it. Moreover these stories indicate that severe consequences are often associated with data quality failures.

III Survey

The working party members believe that data quality issues have a significant impact on the work undertaken by general insurance actuaries. However, we decided to subject this presumption to a more formal examination. Consequently, we conducted a brief survey of actuaries[6] in order to assess the impact of data quality issues on their work.

There were two issues that we wanted to understand:

[6] In some case, other quantitative analysts and systems people who work with and support actuaries were included in the survey.

- What percentage of their time do general insurance actuaries spend on data quality issues?
- What proportion of projects undertaken by general insurance actuaries is adversely affected by data quality issues?

As a result, our survey was very brief, consisting of only two questions. The precise wording of the survey questions was as follows:

- Based on the time spent by both you and your actuarial staff, what percentage of this effort is spent investigating and rectifying data quality issues?
- What percentage of the project results are adversely affected by data quality issues? Adversely affected includes re-working calculations after data is corrected; or stating results/opinions/conclusions but allowing for greater uncertainty in results; or finding of adverse runoff over time due to initial work based on faulty data; etc.

We were aware that surveys recently undertaken by other GIRO working parties had often received disappointing response rates. These tended to have been distributed to all general insurance actuaries, either directly via email, or via the UK actuarial profession's eNews bulletins. Consequently, we decided to adopt a more targeted and personal approach. Copies of the survey were sent to the following groups:

- All original members of the Data Quality working party, including those who had subsequently chosen not to take part in our work;
- Members of the CAS Committee on Management Data and Information; and
- Members of the CAS Data Management and Information Educational Materials Working Party
- In addition, each member of the Data Quality working party contacted a handful of people to ask them to answer the survey questions. This survey was carried out by phone.

As a result of these efforts, we received 38 responses to the survey.

The tables below summarise the results of the survey. We have split the results between those actuaries who work for insurers or reinsurers, those who work as consultants, and the remainder. The latter category includes statistical agents, reinsurance brokers and rating agencies, as well as those respondents who we were unable to categorize. In each case, we show both the mean and median responses. In addition, we show the highest and lowest responses to give an indication of the range of the responses.

Question 1: Percentage of Time Spent on Data Quality Issues

Employer	Number of Responses	Mean	Median	Minimum	Maximum
Insurer/Reinsurer	17	26.4%	25.0%	5.0%	50.0%
Consultancy	13	27.1%	25.0%	7.5%	60.0%
Other	8	23.4%	12.5%	2.0%	75.0%
All	38	26.0%	25.0%	2.0%	75.0%

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

Employer	Number of Responses	Mean	Median	Minimum	Maximum
Insurer/Reinsurer	15	27.9%	20.0%	5.0%	66.0%
Consultancy	13	43.3%	35.0%	10.0%	100.0%
Other	8	22.6%	20.0%	1.0%	50.0%
All	36	32.3%	30.0%	1.0%	100.0%

The discrepancy between the numbers of responses to the two questions arises due to two respondents who provided quantitative answers to question 1 but not to question 2.

The first point to make about these results is that they support the hypothesis that data issues have a significant impact on the work undertaken by general insurance actuaries. The mean response to question 1 implies that actuarial staff spend about a quarter of their time on issues of data quality. There was not much variation between the employer groupings here with all three means covered by a span of less than four percentage points.

For each employer grouping, the median result was below the mean, indicating that the means had been pulled up by a small number of relatively high responses. However, the difference was only significant in respect of the "other" category, where the mean is heavily influenced by one response of 75%. Without this one respondent, the "other" category mean would reduce from 23.4% to 16.0%.

The responses to the second question also indicate that data quality is a major issue for general insurance actuaries since they show that almost a third of projects are adversely affected by data issues. There is more variation between the employer groupings here with the mean and median for consultants being significantly higher than for other actuaries. This could have something to do with the fact that consultants will be less familiar with the data they are using than actuaries working for insurers and reinsurers. It could also be related to the fact that consultants may be more likely to caveat their work than company actuaries for two reasons. Firstly, consultants will be addressing audiences that they do not deal with on a day-to-day basis and so may feel the need for more explanation. Also, consultants need to be conscious of their legal liability to their clients. A third possible explanation is that consultants work with data supplied by external vendors, such as third party administrators more often than company actuaries. Such data is widely viewed as being significantly lower in quality than that of large insurance companies. That is, third party data is frequently sparser and less complete than insurance company data and in addition, all too frequently also contains coding, allocation and other data errors.

It is worth noting that consultants also had the highest mean and equal highest median of the three groups in response to question 1, although most of the dfferences there were much smaller.

As with the first question, the median responses to question 2 were below the means for all three employer groupings. Again, this indicated that the means are being increased by a small number of high responses. For example, in the case of consultants, the mean is significantly affected by two responses of 100%. Without these, the mean would reduce from 43.3% to 33.0%.

It is clear from the above tables that we received a wide range of responses, with answers to question 1 varying between 2% and 75%, and those to question 2 varying between 1% and 100%. The range of responses was wide everywhere - of the two questions and three employer groupings, the narrowest range of responses was 45 percentage points. This implies that, even within our employer groupings, the significance of data issues varies substantially. Presumably, this has to do with such factors as:

- The quality of the data systems at the actuary's employer or client companies;
    - One respondent clearly indicated that his company experiences almost no data quality issues or adverse consequences after initial processing by a systems and statistical team. However, at this company significant effort is applied to screening and cleansing the initial data resulting in a very

"clean" database. The respondent's company perhaps serves as a model for other companies.

- The types of work the actuary is involved in, and the extent to which they are reliant on external and internal data.

However, despite the wide variation in responses, data quality issues appear to be significant for most general insurance actuaries. Only 11% of the responses to question 1 were below 10%, and only 34% were below 20%. Similarly, on question 2, only 11% of the responses were below 10% and only 39% were below 20%. Only one respondent provided answers that were below 10% to both questions, and only 28% answered both questions with figures that were below 20%.

On average, the answers to question 2 were higher than those to question 1 - in other words, the proportion of projects affected by data quality issues is higher than the percentage of time spent dealing with such issues. However, this was by no means the case universally - the mean responses to question 2 were higher for insurers / reinsurers and consultants, but lower for the "other" category. The median responses to question 2 were higher than those to question 1 for consultants and "other", but lower for insurers / reinsurers. The overall median was higher for question 2. Of all respondents who provided quantitative answers to both questions, 47% gave higher answers to question 2, while 36% gave higher answers to question 1, and 17% gave identical answers to both questions. It follows that the only conclusion that can reasonably be drawn about the relationship between the percentage of products impacted and the percentage of time spent is that there is no definitive pattern.

In summary, the conclusions we draw from our survey are that:

- The quality of data has a significant impact on the work undertaken by general insurance actuaries;
- On average, about a quarter of the effort expended by general insurance actuarial teams is spent on data quality issues;
- On average, about a third of projects undertaken by general insurance actuaries are adversely affected by data quality issues;
- There is some limited evidence to suggest that data quality issues have a more significant impact on the work of consultants than on the work of other general insurance actuaries;
- The impact of data quality issues varies widely between different general

insurance actuaries, even those working in similar types of organisations;

- There is no conclusive evidence to suggest any particular pattern in the relationship between the percentage of time that an actuarial team spends on data quality issues, and the percentage of their projects that are adversely affected by such issues.

These survey results support the working party’s initial assumption that data quality problems impose a significant cost on the insurance industry.

IV. Data Quality Experiment

1. Purpose of the Experiment

In order to examine the impact of data quality problems on critical financial quantities, the working party decided to conduct a data quality experiment with data used for an actuarial application. This experiment is designed to examine the impact of incomplete and/or erroneous data on loss reserves. The working party decided it would be more efficient and demonstrative to use real data and its ultimate losses rather than to develop a dataset using Monte Carlo simulation. (Note that a working party of the Casualty Actuarial Society is working on developing a database to be used in reserving and other actuarial research using Monte Carlo methods). Data of sufficient maturity was obtained -- all years are fully developed and the true ultimate losses are known -- and various methods were employed to estimate ultimate losses. Subsets of the data that varied with respect to completeness were then produced. The various methods were used to estimate the ultimate losses on the subsets. Modifications were introduced into duplicates of the data subsets to simulate data errors and data quality problems. The various estimates of ultimate losses, based both on modified and unmodified datasets, are compared to the "true" ultimates to measure the accuracy of the estimates.

We restricted methodologies to mechanical approaches in order to filter out the impact of different actuaries making different subjective judgments. The selected approaches for estimating ultimate losses are: (1) incurred chain ladder, (2) paid chain ladder, and (3) claim count x severity (where each estimate is based on incurred data and the chain ladder method).

We begin with a brief summary of the data in subsection 2. Next we examine the impact of varying the size of the dataset by methodology. The following subsection discusses the simulation of errors in the datasets and examines their impact on the estimates. The subsection also discusses a simple bootstrap analysis of the unmodified and modified data. Finally we compare the results from the different estimates of the ultimate losses and we provide our observations and conclusions.

2. The Data

a. Original Data Set

A database with 18 accident years of data, including incurred and paid losses, reported counts, closed counts and exposures, was located and distributed to the working party members for this project. The triangles contain an accident year in each row with annual evaluations of the statistic in each column (e.g., the second column is the cumulative value of the statistic at two years or 24 months of development). The data are for accident years 1974 to 1991.

The data are from primary, private passenger automobile bodily injury liability business from a single no-fault American state. The data are direct with respect to reinsurance and limited to policy limits written. Policy limits distributions remained somewhat constant during the experience period. Although the data were slightly distorted to guard against identification, they are reflective of an actual situation.

The "ultimate losses" have been supplied by the provider of the triangles. However, because the original data was altered somewhat to hide the identity of its source, the "actual" ultimate losses do not exactly track the true actual numbers. These data are in Appendix A and the projections based on it are included in Appendix B.

3. Experiment 1: Quantity of Data

a. Samples of Varying Sizes Were Used in the Analysis

In order to evaluate the impact of the completeness, i.e., the quantity of data, subsets of the data were created with varying number of accident years. Subsets were created with (1) all years, (2) a subset that only looks at the data for accident years 1986 to 1991[7], and (3) the latest three diagonals of information. The loss development pattern selected for each dataset is the volume-weighted average of (1) all years, (2) latest years available up to five, and (3) latest four years, respectively.

b. Results

The results of the experiment are summarised in Figures 1 and 2. In each graph, the columns represent the actual answer known with the benefit of hindsight, whilst the lines show the results from the various approaches. The model legend is as follows:

- All points identified with triangles are from paid chain ladder models,

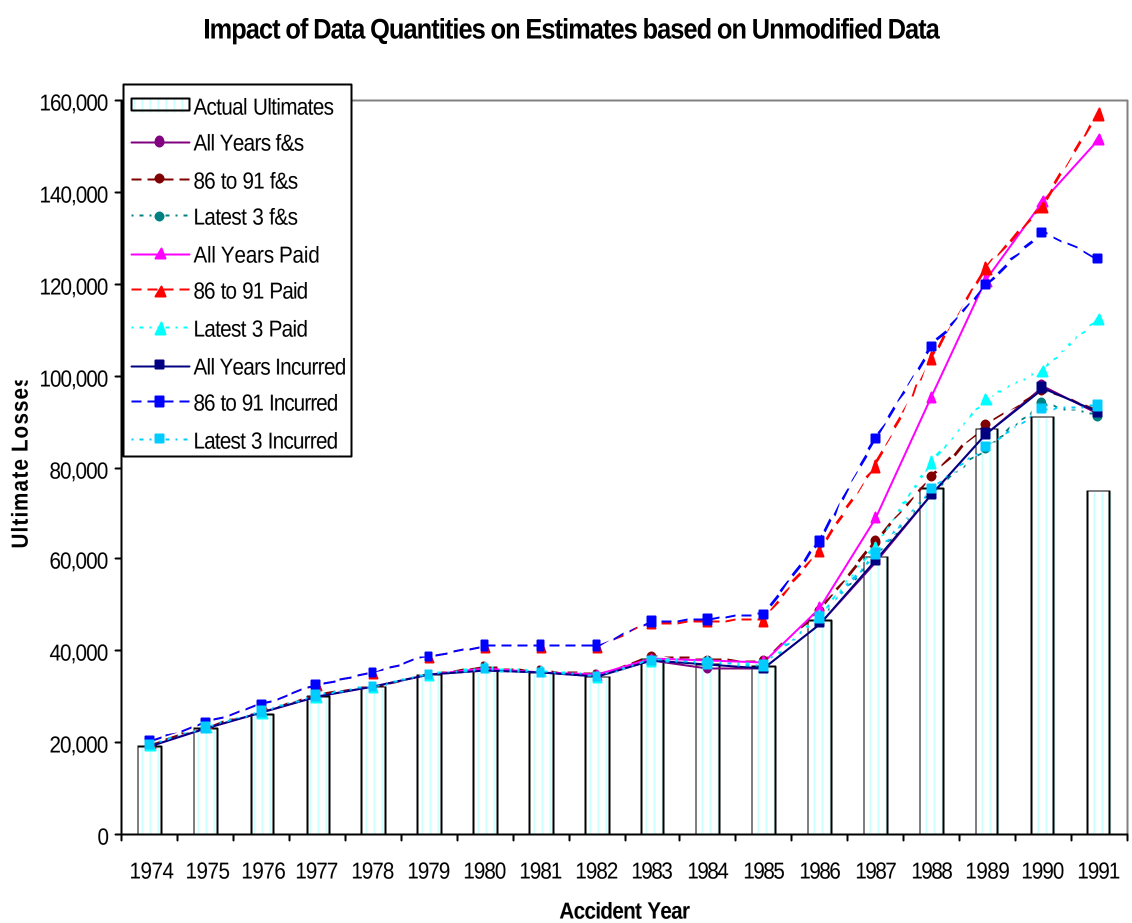
[7] An inverse power curve (Sherman, 1984) was used to fit the tail factors on the 1986 to 1991 data.

- All points identified with squares are from incurred chain ladder models, and
- All points identified with circles are from models that multiply claim counts by severity.

As for datasets,

- Solid lines represent all years datasets,
- Dotted lines represent three year datasets, and
- Long dashed lines represent data from 1986 to 1991 only.

Figure 1



A brief inspection of the estimated ultimates arising from paid and incurred chain ladder models indicates that the incurred ultimates are near the frequency/severity ultimates but the paid ultimates are very different from the incurred ultimates and are also much higher than the actual ultimate losses. This is largely due to the impact of the 12-to-ultimate factor and to a lesser extent the 24-to-ultimate factor. A more stable approach (such as a Bornhuetter-Ferguson model) might be more appropriate in this situation, but in the interest of keeping the analysis simple and mechanical, no modifications to ultimates resulting from the chain ladder methods were computed. We believe it is likely that practicing actuaries would use other methods in this situation.

Observations:

- All methods produce reasonable estimates for all but the most immature points,
- The paid method, which is based on less data, produces worse estimates than the models based on incurred data,
- Datasets with more historical years of experience produce better estimates than datasets with fewer years of experience, and
- There is some correlation between the number of accident years in the dataset and the accuracy of the estimate.
- Finally, all the methods are high by a significant amount on the latest year. This is a result of changing patterns and the challenge of estimating ultimates with this data (there were inherent challenges with the original dataset). For instance, inspection of the closing rate triangle indicates that closing rates on the diagonal are higher than for most of the history supporting the loss development calculations, suggesting later "operational time" (Wright, 1992) for the diagonal than for the prior years. A reserving actuary might make an adjustment for this. However, since we are trying to keep the analysis simple and mechanical, no attempt has been made to deal with unusual patterns in the data.

In summary, with accurate information, more data generally reduces the volatility of estimation errors. This is a reassuring result given actuaries' reliance on the law of large numbers. However, it implies that the widespread use of very small and sometimes very sparse datasets for pricing large deductible, self-insurance and other alternative market accounts as well as reinsurance contributes to the significant overall uncertainty in the estimates.

4. Experiment 2: Modified data

a. Data Modifications to Simulate Data Quality Problems Encountered in Real Life

Based on the actual experience of members of the working party, we postulated various events inducing data glitches such as systemic misclassification of claims to the wrong accident year and erroneous entries escaping systems edits. The datasets were then modified to reflect the effects of such issues. The working party decided to introduce more than one error at a time to improve the realism of the scenario and to explore how the interaction of errors can affect estimates. The errors simulated were judgmentally selected based on working party members experience with data problems.

The modified triangles simulate the following data quality issues:

1. Losses from the 1983 and 1984 accident years have been misclassified as 1982 and 1983 respectively.
2. Data prior to the 1982 calendar year is not available.
3. Approximately half of the financial movements from 1987 were processed late in 1988.
4. The paid losses in the latest diagonal are crude estimates rather than actual losses.
5. The incremental paid for accident year 1988 development period 12-24 has been overstated by a multiple of 10. This was corrected in the following development period. Similarly, the outstanding reserve for accident year 1985 at the end of development month 60 was overstated by a multiple of 100 and was corrected in the following period.
6. From 1988 onwards, the definition of "Reported claims" was changed to exclude claims closed without payment.

The projections based on the modified data appear in Appendix C. It should be noted that all of the modifications affected all of the datasets except the 1986 to 1991 dataset, which was not affected by modifications 1, 2 or the reserve modification of 5.

The same subsets and development methodologies used for the unmodified data were applied to the modified data. Again, all of the methods used to project the claims are mechanical: there is no judgment involved. This means that in places where there is missing data, the selected development factors on the approaches using volume-weighted averages will be wrong because there is a mismatch between the numbers of years containing claims figures in the numerator and the denominator. In practice, an actuary may well spot this and correct it, but we wanted to demonstrate the more extreme distortion caused by a failure to do so.

b. Results

Figure 2

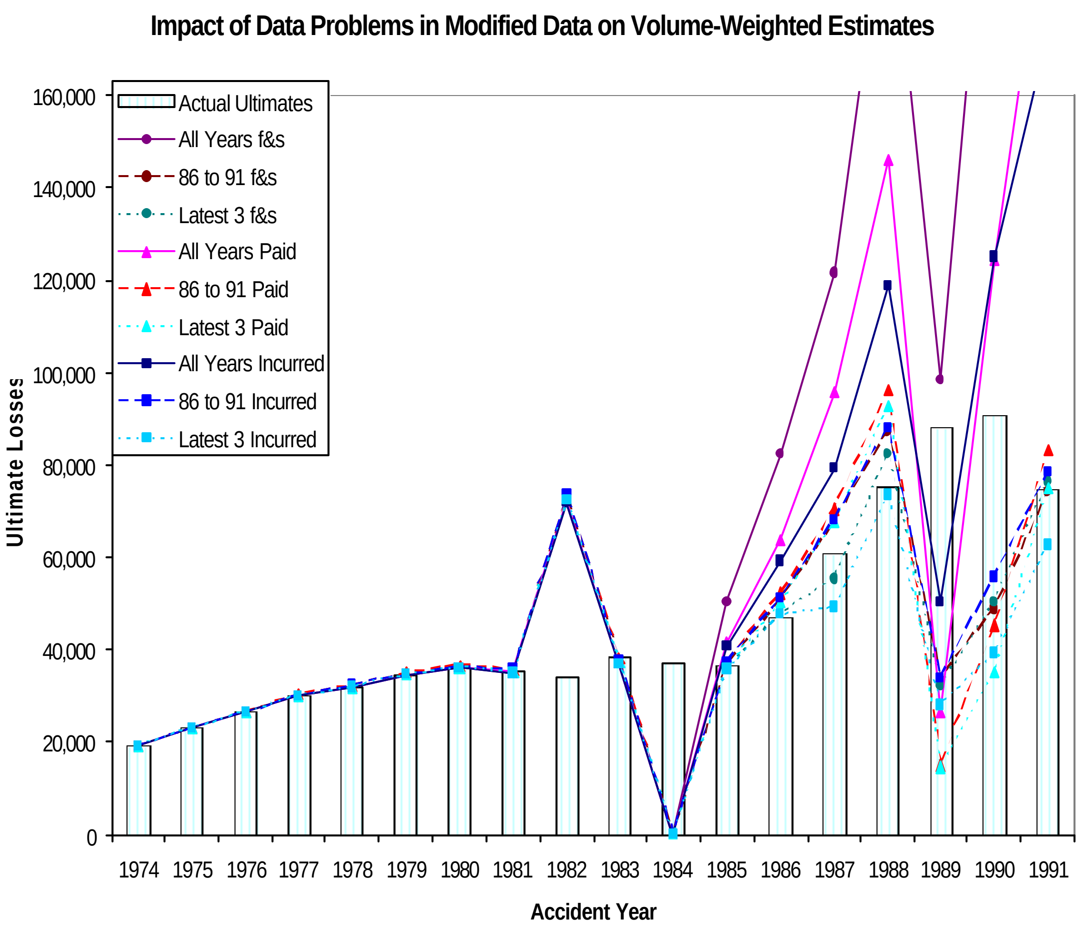
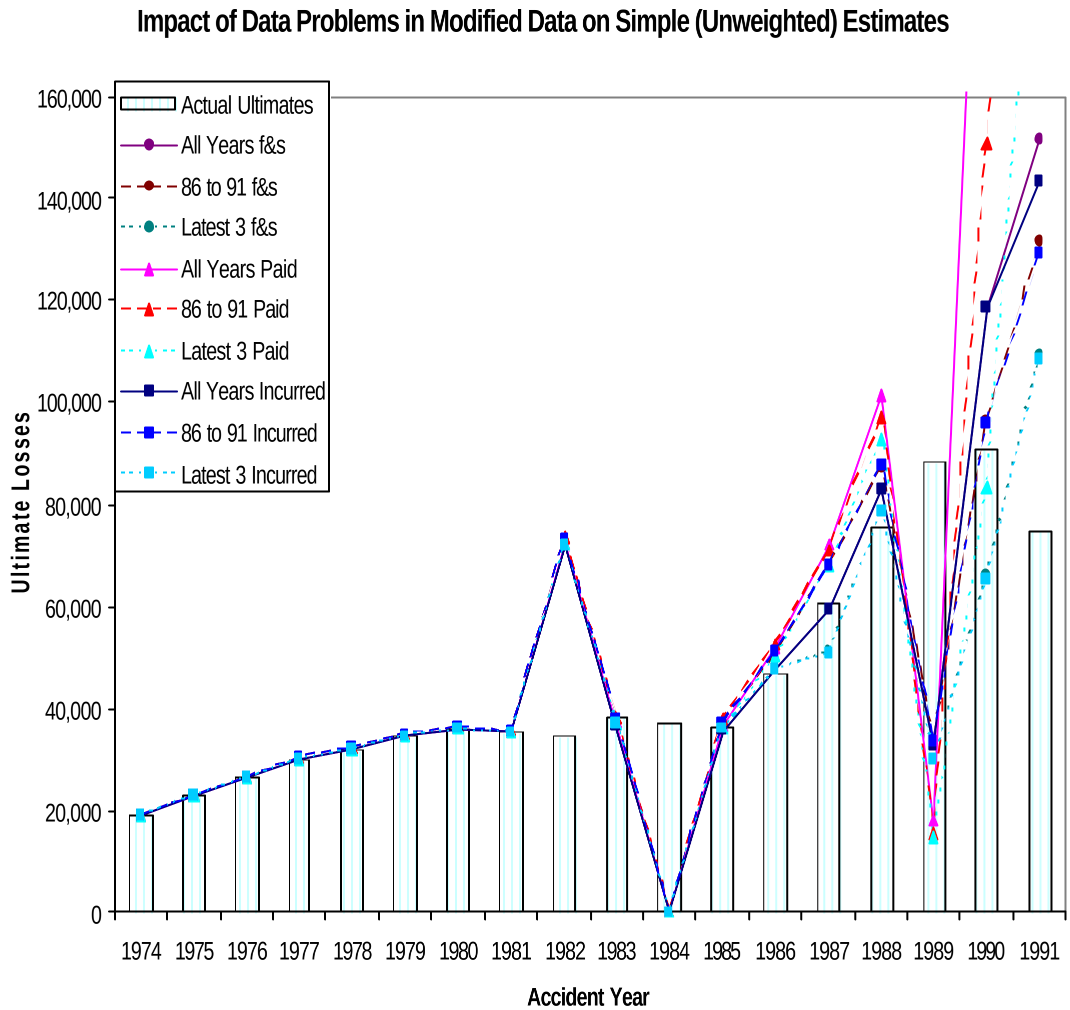


Figure 3



Figures 2 and 3 show the projected ultimate losses for the modified data using each of the projection methods. The legend is the same as in the graph of the previous subsection; the columns represent the actual answer known with the benefit of hindsight, whilst the lines show the results from the various approaches. Ultimate losses based on both weighted and unweighted average link ratios are shown, as the weighted average results are highly variable. Our comments are focused on results displayed in Figure 3.

The graph indicates there is some extreme volatility for some of the projections based on modified data, particularly for ultimate values based on paid losses. When compared to the unmodified data in Figure 1, the results for the modified data show a large amount of both additional volatility and error. In practice an actuary will undoubtedly spot many of the errors and try to correct for them. Nevertheless the actuary will often be unable to get back to the "unmodified" data, so some of the additional volatility and error will almost certainly remain. Indeed, in some cases, an attempt to correct the data may introduce additional volatility, if not error.

An unexpected observation is that the datasets with more accident years produce worse estimates. In fact there seems to be an inverse relationship between the number of accident years used and the accuracy of the resulting estimates. The working party hypothesizes that this occurs because estimation errors due to undiscovered quality issues in older accident years can be perpetuated and compounded in later accident years with progressively larger calendar-year impacts when using particular models and assumptions for projection.

5. Bootstrapping

a. Description of Bootstrapping

A limitation of the analyses we have performed is that they are based on single realizations of claim counts, paid losses and incurred losses from a distribution of potential outcomes. Other realizations would have resulted in different development factors and different ultimate loss estimates using the same estimation methods and based on the same underlying stochastic processes generating the data. In order to augment our analysis with information about a distribution of realizations for the development factors, the technique of bootstrapping was used. Bootstrapping is a computationally simple way of obtaining prediction errors and probability distributions of the predictions. In its simplest form, bootstrapping assumes that the empirical data supply a probability distribution that can be sampled to derive uncertainty measures of functions (such as means, sums and projected ultimates) based on the data. A description of how to apply bootstrapping to loss development models to obtain information about reserve uncertainty is provided by England and Verrall (England and Verrall, 1999, 2002). Our approach is based on using link ratios to estimate the “expected” amounts in each cell of the loss development triangle. The paid and incurred link ratio methods were used (but not the claim severity and claim count method). The original and modified data for both the paid and incurred claims were passed through a mechanical bootstrapping process. The process used a commercial reserving package. However, the following broad steps were followed in the calculation:

- A link ratio model was fitted to derive the best estimate of the development pattern underlying the data. Link ratio selections were based on a simple average of the last 5 diagonals. [8]

- An “expected triangle” of data was derived by applying this pattern backwards so that the each origin year arrived at the current latest point by following the derived pattern precisely.
- A triangle of residuals was calculated by comparing the actual data with the expected data triangle.
- The triangle of residuals was adjusted to allow for the different exposure volumes in the original years being modeled and also for the expected differences between residuals at varying development ages.
- 10,000 simulations were run on each set of data. During each simulation, the adjusted residuals were sampled and added to the expected triangle to generate a new data triangle. The link ratio projection method was then applied to each of the generated data triangles to produce an estimate of the ultimate claims.

Bootstrap results were generated based on each of the complete unmodified and modified data for the total ultimate claims cost (i.e. ultimate losses for all accident years combined). Percentiles were calculated from the bootstrapped samples in steps of 0.5%

[8] An unweighted rather than a weighted average was used, due to a large number of very extreme realizations from the weighted average bootstrap. These results were deemed unreasonable by the working party.

b. Results

Table 1 presents summary statistics from the bootstrap analysis. Descriptive statistics from the bootstrap are presented at the top of the table followed by a display of the results of each scenario at various percentiles. As one would expect, bootstraps based only on paid data have a lot more variation than those based on incurred data and, of course, the bootstraps on modified data have much higher variation than those based on the original clean data.

Table 1

Bootstrap Results

	Original (Clean) Data		Modified Data	
	Paid Triangle	Incurred Triangle	Paid Triangle	Incurred Triangle
Mean	847,331.33	801,467.56	871,826.86	813,972.56
Median	842,540.70	800,366.59	839,299.99	796,824.57
Standard Error	41,922.67	13,118.45	173,196.85	137,163.59
Co-Efficient of Variation	0.049	0.016	0.199	0.169

Percentiles	Clean Paid	Clean Incurred	Modified Paid	Modified Incurred
0.5	774,535.27	772,851.16	614,432.25	631,059.77
5.0	790,896.47	781,897.55	677,675.80	689,848.42
10.0	797,565.47	785,128.49	705,994.02	708,280.32
15.0	802,992.60	787,631.74	727,536.80	721,691.11
2+0.0	808,072.71	789,793.62	746,119.93	734,320.99
25.0	813,521.73	791,871.54	763,197.22	745,449.71
30.0	818,857.25	793,589.32	779,209.55	755,894.28
35.0	824,872.70	795,178.48	793,858.17	765,841.98
40.0	830,650.39	796,705.71	808,940.71	775,747.19
45.0	836,770.47	798,313.44	823,775.66	785,876.67
45.5	837,261.97	798,485.07	825,060.70	786,923.13
50.0	842,469.55	799,968.66	839,061.61	796,318.14
55.0	848,316.87	801,536.22	854,439.44	806,401.76
60.0	854,053.84	803,296.53	872,574.73	817,747.42
65.0	860,678.34	805,255.49	888,340.88	829,337.46
65.5	861,340.82	805,437.92	889,933.14	830,582.11
70.0	866,853.86	807,018.37	907,697.78	842,458.84
75.0	874,646.71	809,107.00	932,673.24	856,696.96
80.0	882,432.59	811,431.14	962,019.76	874,277.28
85.0	891,604.91	814,462.96	1,005,571.24	896,074.78
90.0	904,833.60	818,030.38	1,065,473.50	931,223.53
95.0	923,521.78	823,790.41	1,173,371.82	991,517.62
99.5	972,926.41	841,729.87	1,712,821.06	1,258,773.23

Note that this table displays estimates of the ultimate losses (in thousands of dollars), for all years combines, not IBNR or total reserves, for all years combined. It follows that an estimation error of 80,000 could well be an entire year's premium.

Figure 4 displays the cumulative distribution of ultimate losses from the bootstrap experiment. This graph shows that the clean incurred ultimates and the clean paid ultimates reach a cumulative probability of 100% at much lower values than do the modified incurred and modified paid ultimate losses. Figures 5 (paid data) and 6 (incurred data) display the probability density function for the bootstrapped results for the clean and modified data. These graphs indicate that there is less dispersion in the estimates based on error free (i.e., unmodified) data. For both the incurred method and the paid method, the modified data displays a much heavier tail and much higher variability. From Table 1, it is also apparent that the standard deviation of estimated ultimates is much greater (by an order of magnitude) for the modified data. The distributions and statistics from the bootstrap analysis confirm our original hypothesis, that the uncertainty of estimates based on poor quality data is significantly higher than that of good data. While actuarial estimates usually contain uncertainty, when reserving using data not processed through a rigorous quality review process, the uncertainty is likely to be much greater, and therefore the magnitude of any under or over-estimates is likely much higher than for data that have been screened.

Figure 4

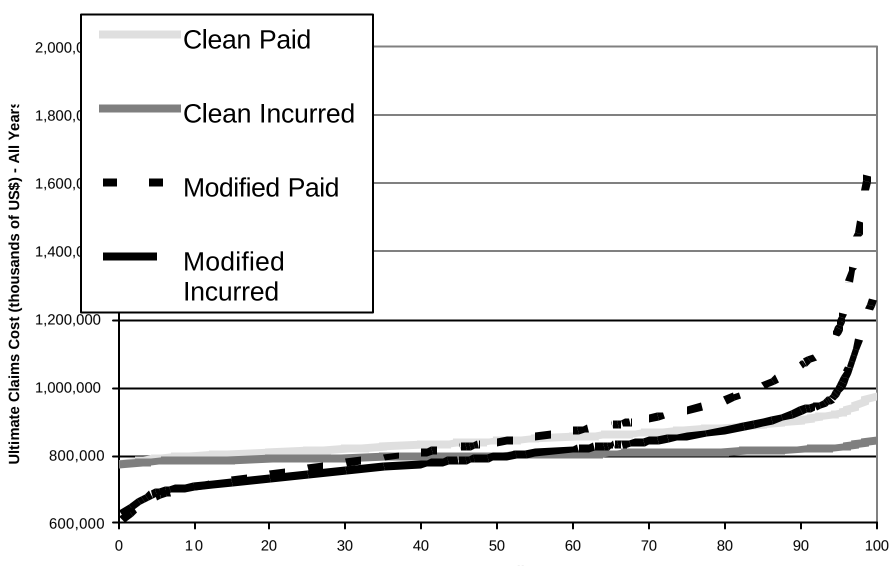
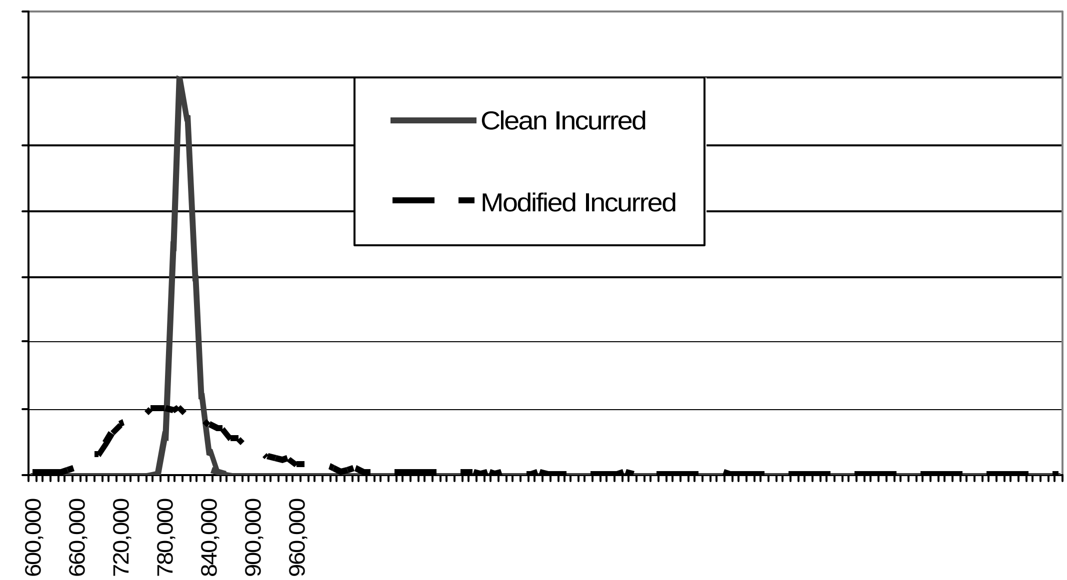


Figure 5

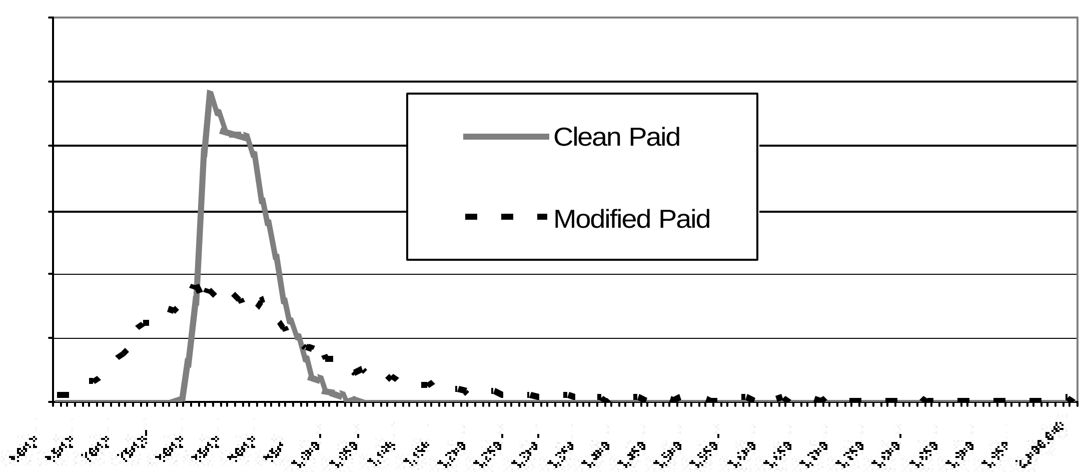
Density Function - Bootstrap based on Incurred Claims



Ultimate Claims Cost (thousands of US$) - All Years

Figure 6

Density Function - Bootstrap based on Paid Claims



Ultimate Claims Cost (thousands of US$) - All Years

6. Conclusions from the Data Quality Experiment

Figure 7

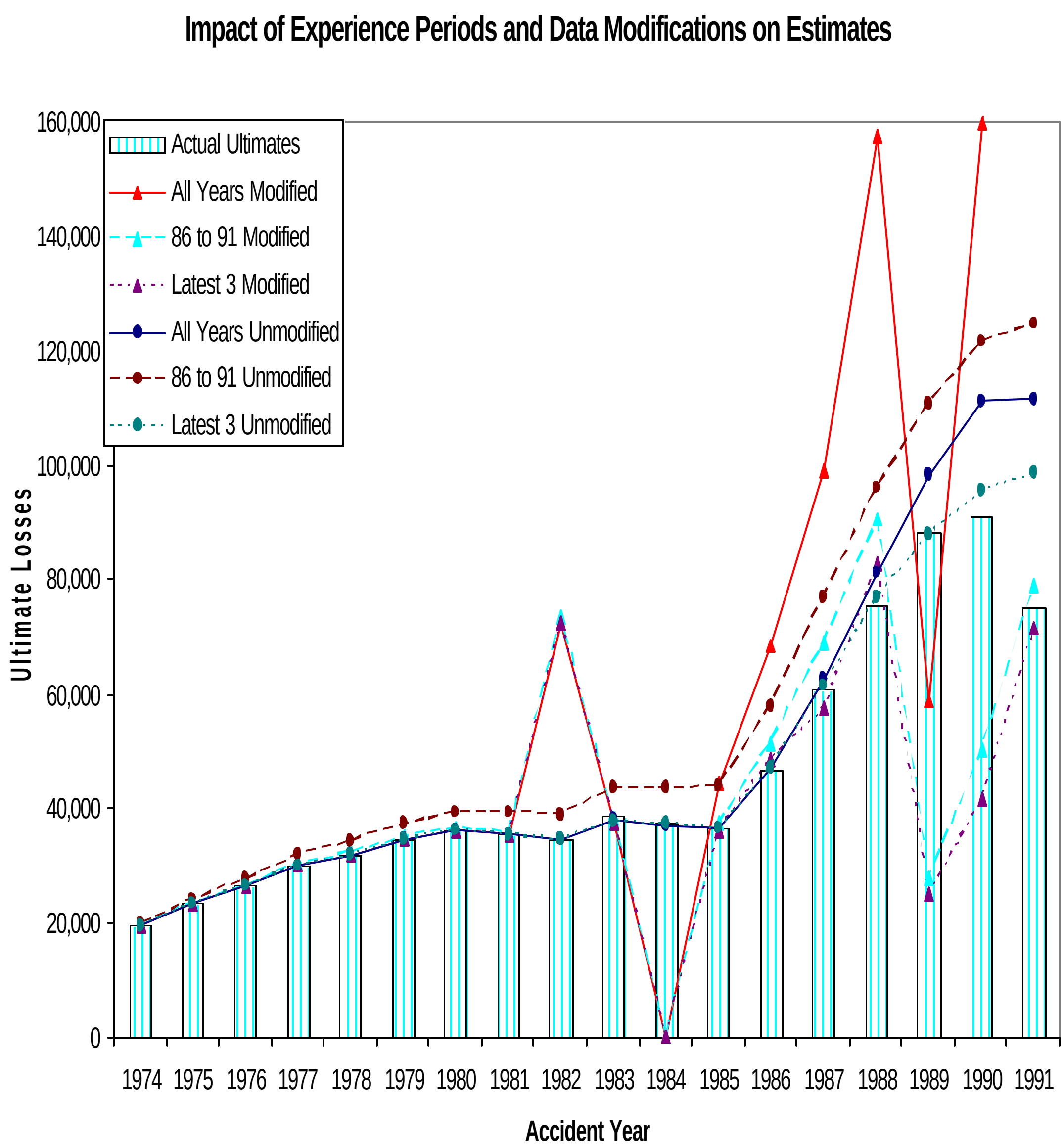


Figure 7 presents a summary of results from the data quality experiment. Summary statistics were produced for the subsets of modified and unmodified data by averaging the ultimates from each of the three chain ladder estimation methods. Figure 7 indicates that there is greater accuracy and less variability in the estimates based on error free (i.e., unmodified) data as well as estimates based on greater numbers of accident years. The interesting finding of this example is that the advantage of more complete data assumes that the data is relatively clean. The example suggests that data quality issues can erode or even reverse the gains of increased volumes of data. This effect seems to be a function of the calendar year impact of the errors, but it may also be exacerbated by the methods and assumptions selected. Further research might identify the precise relationships among these factors.

Our research indicates the there is a significant increase in the uncertainty of results when data quality problems arising from incompleteness of data or data errors occur. The magnitude of these errors can seriously degrade the reliability of actuarial analyses. Our research is only a beginning in examining the consequences to insurance companies of data quality problems. It was limited to one relatively small dataset. A variety of datasets from a variety of lines of business would provide a more complete picture of the impact of data quality problems. In addition, predictive modelers might be interested in the impact on their work of errors in large corporate databases.

However, we believe our research indicates that the most efficient way to mitigate the consequences is to minimize errors in the data by ensuring that quality data enters systems, that errors are corrected promptly, and that the systems and processes handling the data are error free.

V. Conclusions and Actions

1. Conclusions

As discussed in Section I, the Data Quality Working Party was formed because of the view that

- data quality issues significantly impacted the work of general insurance actuaries; and
- such issues could have a material impact on the results of general insurance companies.

The Working Party wanted to encourage the insurance industry and the actuarial profession to improve practices for collecting and handling data and, in order to do so; much of our work was designed to test the accuracy of the statements in the two bullet points above.

In Section II, we highlighted a number of anecdotal incidents in which data errors had very serious repercussions. This section included anecdotes both from the insurance industry and from elsewhere. They demonstrate that the adverse consequences of poor data quality can be very significant. Within the insurance industry, such incidents tend to negatively impact the company's profits, potentially to a very material extent. Outside the insurance sector, the consequences can sometimes be even more severe - in some of the examples we discussed, data errors resulted in people losing their lives.

In Section III, we discussed the results of a survey of general insurance actuaries (and other quantitative analysts and support people who work with such actuaries) that demonstrated that data quality issues have a significant impact on the work they undertake. The survey indicated that, on average, about a quarter of the effort expended by actuarial teams is spent on data quality issues, and about a third of the projects they undertake are adversely affected by such issues. There was some limited evidence to suggest that the impact might be more severe on consultants than on other actuaries.

In Section IV, we discussed an experiment that we conducted in order to examine the impact of data issues on critical financial information - in this case, the level of an insurer's required claims reserves. We obtained a dataset that was around 15 years old, meaning that the actual ultimate outcomes were available, and used various mechanical actuarial projection methods to estimate the ultimate claims. In order to test the impact of only having access to restricted information, we then created various subsets of the data that varied in their level of completeness. The actuarial projection methods were then used to estimate the ultimate claims based on each of these restricted datasets. In addition, in order to test the impact of errors in the data, the dataset was modified to reflect the effect of various hypothetical data errors and the various projections were repeated using the modified data. From the results of this analysis, we drew the following conclusions:

- there was some correlation between the size of the dataset and the accuracy of the estimates;
- estimates based on paid claims produced worse estimates than those based on incurred claims, presumably because they utilize less data (that is the case reserve information is not used which particularly impacts immature years);
- when data errors were introduced, the accuracy of the estimates deteriorated significantly;
- in addition, when data errors were introduced, the volatility of the estimates increased.

The outcome of the data experiment indicated that there is a significant increase in the uncertainty of results when data quality problems arising from incompleteness of data and data errors occur. The size of these errors can significantly reduce the reliability of actuarial analyses, and this could have a direct impact on an insurer's financial statements.

The conclusions from our survey, data experiment, and research into the past impact of data issues, support the working party's initial hypotheses that were stated at the start of this section, namely that

- data quality issues significantly impacted the work of general insurance actuaries; and
- such issues could have a material impact on the results of general insurance companies.

It follows that, if insurers improved the quality of their data, it could have a number of highly beneficial effects:

- it could increase their profitability;
- it could improve the accuracy and reliability of their financial statements; and
- it could free up actuarial resources (as well as resources in other areas such as finance and IT) to concentrate on other work that could add more value to the organization.

Because actuaries are typically heavy users of data and must frequently contend with poor quality data, we believe actuaries should become data quality advocates. In the next section, we describe some actions that can be taken by actuaries and insurance company managements to improve data quality.

2. Actions

a. Data Quality Advocacy

Two organisations in the United States are working to increase the profile of data quality issues in the industry:

- The Casualty Actuarial Society:
  - The Data Management and Information Education Materials Working Party is reviewing the data quality literature in order to make recommendations for the examination syllabus and for continuing education on the subject of data quality. This Working Party is also planning presentations at CAS seminars and conferences to raise awareness and generate interest in the topic of data quality.
  - The CAS Committee on Management Data and Information also regularly sponsors presentations at conferences and seminars.
- The Insurance Data Management Association is an excellent source of information on insurance data quality.
  - The IDMA web site contains material on the "value proposition" that describes the value of data quality from the perspective of various insurance specialists. This contains sections on the value to senior management, the value to claims, the value to marketing, etc. as well as the value to actuaries.
  - The IDMA also sponsors an annual conference where data quality is typically a topic on the schedule and its web site contains suggested readings on data quality.
  - The CAS and IDMA jointly sponsor a Data Quality/Data Technology call paper program every two years. Data quality is one of the issues that authors submit and have published papers on.

These are examples of data quality advocacy which can be undertaken by professional actuarial and industry organizations. More specific actions that can be taken to improve data quality within organizations are discussed next.

b. Data Quality Measurement

As a tool for promoting data quality improvement, a number of authors recommend regular measurement of an organisation's data quality (Dasu and Johnson 2003, Redman, 2001). Among the advantages of measurement noted by Redman[9] are that measurement replaces anecdotal information with factual data, quantifies the severity of the problem and identifies where the problems are (so they can be acted upon).

Some of the measures recommended by Dasu and Johnson quantify traditional aspects of quality data such as accuracy, consistency, uniqueness, timeliness and completeness. Some capture systems related aspects of data quality such as extent of automation (sample some transactions, follow them through the database creation processes and tabulate the number of manual interventions), successful completion of end-to-end processes (count the number of instances in a sample that, when followed through the entire process, have the desired outcome). Yet others are intended to measure the consequences of data quality problems (measure the number of times in a sample that data quality errors cause errors in analyses, and the severity of those errors). Dasu and Johnson recommend that the different metrics be weighted together into an overall data quality index using business considerations and the analysts' goals to develop weights.

Redman points out that the most appropriate measure depends on the organisation. An organisation that is just beginning its data quality initiative probably only needs simple measures, while a more advanced organisation might employ more sophisticated measures. Redman offers the following algorithm for implementing a simple data quality measure[10]:

- determine who will take the action
- select a business operation
- select needed data fields
- draw a small sample
- inspect sampled records
- estimate impact on business operation
- summarise and present results

[9] Redman, p107
[10] Redman, p108

- follow up

c. Advocating Data Quality – Management Issues

In this section we briefly summarise some of the recommendations in the data quality literature for implementing data quality programs.

For data originating within one's company, Redman suggests managing the information chain. Redman notes that most information is distributed horizontally. For instance, an information technology department programs and maintains a claims system that collects and stores claims data, and performs edits on data as they are entered. Claim adjusters record information into the claims system. Actuaries use the claims data, perhaps after aggregation by yet another department. The flow of this data is from department to department, not hierarchically. Redman notes that departments often do not communicate effectively with each other and this exacerbates data quality problems. He suggests that once departments understand the needs of the users of the data, they will be more motivated to satisfy those needs. Redman describes a formal program for information chain management including[11]

- establish management responsibilities
- describe information chart
- understand customer needs
- establish measurement system
- establish control and check performance
- identify improvement opportunities
- make improvements

Redman suggests that some middle managers will resist data quality initiatives, thinking their jobs may be eliminated (because as data processes become more efficient fewer people are needed) and that managers should be assured that this will not occur.

Redman advocates supplier management for data originating outside the company. "The most difficult aspect of supplier management for most organisations is coming to the realization that they have contributed to the inadequate data quality they currently receive. They believe that these suppliers are simply incompetent, don't care, don't have enough good people or use old technology."[12] On the contrary, Redman suggests that organisations do not provide adequate communication and feedback to their data suppliers. Thus Redman

[11] Redman, p.162
[12] Redman, p. 154

suggests[13]

- customers define for the supplier the quality of the data they need
- the supplier measures baseline performance as to how well the requirements are met
- the supplier and user agree on improvements
- the supplier regularly remeasures performance

d. Screening Data

Even when data quality initiatives have been undertaken, actuaries and other analysts will need to screen their data. A fairly extensive literature that is relevant to data quality exists in statistical journals and publications. This includes the tools of exploratory data analysis, pioneered by Tukey (Hartwig and Dearing, 1979 discuss Tukey's contribution), and graphical analysis of data, popularized by Chambers and Cleveland (Chambers et al., 1983, Cleveland, 1993). Exploratory data analysis techniques are particularly useful for detecting outliers. While outliers, or extreme values, may represent legitimate data, they are often the result of data glitches and coding errors. Francis (Francis, 2005) describes a number of exploratory techniques useful for screening data and illustrates their application to a personal automobile database. Some of the methods recommended include:

- produce and examine descriptive statistics such as mean, median, minimum, maximum and standard deviation
- for categorical variables, tabulate the frequency of records in the data containing each value of the categorical variable
- tabulate the percentage of records with missing values for each variable
- produce histograms (possibly on a log scale)
- produce box and whisker plots (possibly on a log scale)
- apply multivariate techniques that screen multiple variables for outliers at once. Francis uses the Mahalanobis depth as a multivariate technique measuring how far a given record is from the centre of the data.

13 Redman, p.155

3. Concluding Remarks

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.

The Working Party also believes that actuaries are well suited to be data quality advocates. In order to fulfill such a role, actuaries will need to familiarize themselves with the data quality literature, perhaps by reading one of the books recommended by the CAS Educational Materials Working Party or the IDMA. They will need to participate in data quality initiatives that manage data quality both from within their company and from external suppliers. Finally, even in the best of scenarios where both their internal and external suppliers initiate a data quality program, they will need to screen data for problems. Vigilance is never ending!

References

Actuarial Standards Board, *Actuarial Standard of Practice 23, Data Quality, Revised Edition*, December 2004, www.actuaries.org

Archibald, R, “Girl in Heart Transplant Dies After Two Weeks”, NY Times, Feb 23, 2003

Arenson, K, “Technical Problems Cause Problems in SAT scores”, New York Times, March 8, 2006

CAS Committee on Management Data and Information, “White Paper on Data Quality”, *Casualty Actuarial Society Winter Forum*, 1997

Chambers, J, Cleveland, W., Kleiner, B. and Tukey, P., *Graphical Methods for Data Analysis*, Wadsworth International Group, 1983

Cleveland, W., *Visualizing Data*, Hobart Press, 1993

COPLFR, Property & Casualty Practice Note, December 2004

CNN, “Amid Protest, US says Faulty Data Led to Chinese Embassy Bombing”, cnn.com, May 9, 1999a

CNN, “NASA’s Metric Confusion Causes Mars Orbiter Loss”, September 30, 1999b

Consumer Federation of America, “Millions of Americans Jeopardized by Inaccurate Credit Scores”, 2002

Cornejo, R, “Searching for a Cause”, *Bests Review*, February, 2006

Copeman, P., Gibson, L, Jones, T., Line, N, Lowe, J, Martin, P., Mathews, P., Powell, D., “A Chang Agenda for Reserving: A Report of the General Insurance Reserving Issues Task Force” 2006, www.actuaries.org.uk

Dasu, T and Johnson, T, *Exploratory Data Mining and Data Cleaning*, Wiley 2003

Eckerson, W, “Data Warehousing Special Report: Data Quality and the Bottom Line”, http://www.adtmag.com/article.aspx?id=6321&page=

England, P and Verrall, R. “Analytic and bootstrap estimates of prediction errors in claims reserving”, *Insurance Mathematics and Economics* 25 (1999) pp. 281-293.

England, P. and Verrall, R., “Stochastic Claims Reserving in General Insurance”, Presented to the Institute of Actuaries, January, 2002, pp443-518.

Francis, L, “Dancing With Dirty Data”, *Casualty Actuarial Society Winter Forum*, 2005

Hartwig, F. and B. Dearing, *Exploratory Data Analysis*, Sage Publications, 1979

Insurance Data Management Association, “Value Proposition”, www.idma.org

Moore, L, “Data Business - No data quality control? Expect to count the cost.”, 2006, Business Media Europe

Olson, J., *Data Quality: The Accuracy Dimension*, Morgan Kauffman Publishers, 2003

New York Times, January 4, 1995, “Magellan Error is Explained”

Popelyukhin, A, “Watch Your TPA”, *Casualty Actuarial Society Winter Forum*, 1999. pp. 239 - 254.

PriceWaterHouseCoopers “Global Data Management Survey 2004”, 2004

Redman, T, *Data Quality: The Field Guide,* Digital Press, 2001

RMS, “Hurricane Katrina: Profiles of a Supercat: Lessons and Implications for Catastrophe Risk Management, October 2005

Ruethling, G., *New York Times*, February 15, 2006

Schneier, Bruce, http://www.schneier.com/blog/archives/2006/02/database_error.html, “Database Error Causes Unbalanced Budget”

Schnacker, B and Chute, E, Colleges Adjust to SAT Errors”, Pittsburgh Post Gazette

Sherman, R., “Extrapolating, Smoothing and Interpolating Development Factors”, Proceedings of the Casualty Actuarial Society, 1984, pp. 122 - 155

Wand Y and Wang R, “Anchoring Data Quality Dimensions in Ontological Foundations”, Communications of the ACM, November 1996

Westfall, C, “How Did Catastrophe Models Weather Katrina?”, Insurance Journal, October 3, 2005

Williams D, “NATO Missiles Hit Chinese Embassy”, Washington Post, May 8, 1999

Wright, T, “Stochastic Reserving When the Past Claims Numbers are Known”, Proceedings of the Casualty Actuarial Society, 1992, pp. 255 - 361

Appendix
Sample Data and Estimated Ultimate Losses

Appendix A - Loss Development Triangles Data

Cumulative Paid Losses
Cumulative Closed With Payment Claims
Reported Claims
Outstanding Claims
Outstanding Losses
Average Incurred Severity
Cumulative Closing Rate
Exposures

Appendix B – Estimate of Ultimate Claims Losses Based on Unmodified Data

1. Ultimate Paid Losses
2. Ultimate Incurred Losses
3. Ultimate Losses Using Ultimate Frequency * Ultimate Severity

Appendix C – Estimate of Ultimate Claims Losses Based on Modified Data

1. Ultimate Paid Losses
2. Ultimate Incurred Losses
3. Ultimate Losses Using Ultimate Frequency * Ultimate Severity

Cumulative Paid Losses

Accident	Months of Development																	
Year	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	$267	$1,975	$4,587	$7,375	$10,661	$15,232	$17,888	$18,541	$18,937	$19,130	$19,189	$19,209	$19,234	$19,234	$19,246	$19,246	$19,246	$19,246
1975	310	2,809	5,686	9,386	14,884	20,654	22,017	22,529	22,772	22,821	23,042	23,060	23,127	23,127	23,127	23,127	23,159	
1976	370	2,744	7,281	13,287	19,773	23,888	25,174	25,819	26,049	26,180	26,268	26,364	26,371	26,379	26,397	26,397		
1977	577	3,877	9,612	16,962	23,764	26,712	28,393	29,656	29,839	29,944	29,997	29,999	29,999	30,049	30,049			
1978	509	4,518	12,067	21,218	27,194	29,617	30,854	31,240	31,598	31,889	32,002	31,947	31,965	31,986				
1979	630	5,763	16,372	24,105	29,091	32,531	33,878	34,185	34,290	34,420	34,479	34,498	34,524					
1980	1,078	8,066	17,518	26,091	31,807	33,883	34,820	35,482	35,607	35,937	35,957	35,962						
1981	1,646	9,378	18,034	26,652	31,253	33,376	34,287	34,985	35,122	35,161	35,172							
1982	1,754	11,256	20,624	27,857	31,360	33,331	34,061	34,227	34,317	34,378								
1983	1,997	10,628	21,015	29,014	33,788	36,329	37,446	37,571	37,681									
1984	2,164	11,538	21,549	29,167	34,440	36,528	36,950	37,099										
1985	1,922	10,939	21,357	28,488	32,982	35,330	36,059											
1986	1,962	13,053	27,869	38,560	44,461	45,988												
1987	2,329	18,086	38,099	51,953	58,029													
1988	3,343	24,806	52,054	66,203														
1989	3,847	34,171	59,232															
1990	6,090	33,392																
1991	5,451																	

Claims Closed with Payment

Accident Year	Months of Development 12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	268	607	858	1,090	1,333	1,743	2,000	2,076	2,113	2,129	2,137	2,141	2,143	2,143	2,145	2,145	2,145	2,145
1975	294	691	913	1,195	1,620	2,076	2,234	2,293	2,320	2,331	2,339	2,341	2,343	2,343	2,343	2,343	2,344	
1976	283	642	961	1,407	1,994	2,375	2,504	2,549	2,580	2,590	2,596	2,600	2,602	2,603	2,603	2,603		
1977	274	707	1,176	1,688	2,295	2,545	2,689	2,777	2,809	2,817	2,824	2,825	2,825	2,826	2,826			
1978	269	658	1,228	1,819	2,217	2,475	2,613	2,671	2,691	2,706	2,710	2,711	2,714	2,717				
1979	249	771	1,581	2,101	2,528	2,816	2,930	2,961	2,973	2,979	2,986	2,988	2,992					
1980	305	1,107	1,713	2,316	2,748	2,942	3,025	3,049	3,063	3,077	3,079	3,080						
1981	343	1,042	1,608	2,260	2,596	2,734	2,801	2,835	2,854	2,859	2,860							
1982	350	1,242	1,922	2,407	2,661	2,834	2,887	2,902	2,911	2,915								
1983	428	1,257	1,841	2,345	2,683	2,853	2,908	2,920	2,925									
1984	291	1,004	1,577	2,054	2,406	2,583	2,622	2,636										
1985	303	1,001	1,575	2,080	2,444	2,586	2,617											
1986	318	1,055	1,906	2,524	2,874	2,958												
1987	343	1,438	2,384	3,172	3,559													
1988	391	1,671	3,082	3,771														
1989	433	1,941	3,241															
1990	533	1,923																
1991	339																	

Cumulative Reported Claims

Accident	Months of Development																	
Year	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	1,912	2,854	3,350	3,945	4,057	4,104	4,149	4,155	4,164	4,167	4,169	4,169	4,169	4,170	4,170	4,170	4,170	4,170
1975	2,219	3,302	3,915	4,462	4,618	4,673	4,696	4,704	4,708	4,711	4,712	4,716	4,716	4,716	4,716	4,716	4,717	
1976	2,347	3,702	4,278	4,768	4,915	4,983	5,003	5,007	5,012	5,012	5,013	5,014	5,015	5,015	5,015	5,015		
1977	2,983	4,346	5,055	5,696	5,818	5,861	5,884	5,892	5,896	5,897	5,900	5,900	5,900	5,900	5,900			
1978	2,538	3,906	4,633	5,123	5,242	5,275	5,286	5,292	5,298	5,302	5,304	5,304	5,306	5,306				
1979	3,548	5,190	5,779	6,206	6,313	6,329	6,339	6,343	6,347	6,347	6,348	6,348	6,348					
1980	4,583	6,106	6,656	7,032	7,128	7,139	7,147	7,150	7,151	7,153	7,154	7,154						
1981	4,430	5,967	6,510	6,775	6,854	6,873	6,883	6,889	6,892	6,894	6,895							
1982	4,408	5,849	6,264	6,526	6,571	6,589	6,594	6,596	6,600	6,602								
1983	4,861	6,437	6,869	7,134	7,196	7,205	7,211	7,212	7,214									
1984	4,229	5,645	6,053	6,419	6,506	6,523	6,529	6,531										
1985	3,727	4,830	5,321	5,717	5,777	5,798	5,802											
1986	3,561	5,045	5,656	6,040	6,096	6,111												
1987	4,259	6,049	6,767	7,206	7,282													
1988	4,424	6,700	7,548	8,105														
1989	5,005	7,407	8,287															
1990	4,889	7,314																
1991	4,044																	

Outstanding Claims

Accident Year	Months of Development 12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	1,381	1,336	1,462	1,660	1,406	772	406	191	98	57	23	13	3	4	0	0	0	0
1975	1,289	1,727	1,730	1,913	1,310	649	358	167	73	30	9	6	4	2	2	1	1	
1976	1,605	1,977	1,947	1,709	1,006	540	268	166	79	48	32	18	14	10	10	7		
1977	2,101	2,159	2,050	1,735	988	582	332	139	66	38	27	21	21	8	3			
1978	1,955	1,943	1,817	1,384	830	460	193	93	56	31	15	9	7	2				
1979	2,259	2,025	1,548	1,273	752	340	150	68	36	24	18	13	4					
1980	2,815	1,991	1,558	1,107	540	228	88	55	28	14	8	6						
1981	2,408	1,973	1,605	954	480	228	115	52	27	15	11							
1982	2,388	1,835	1,280	819	354	163	67	44	21	10								
1983	2,641	1,765	1,082	663	335	134	62	34	18									
1984	2,417	1,654	896	677	284	90	42	15										
1985	1,924	1,202	941	610	268	98	55											
1986	1,810	1,591	956	648	202	94												
1987	2,273	1,792	1,059	626	242													
1988	2,403	1,966	1,166	693														
1989	2,471	2,009	1,142															
1990	2,642	2,007																
1991	2,366																	

Outstanding Losses

Accident	Months of Development																	
Year	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	$5,275	$8,867	$12,476	$11,919	$8,966	$5,367	$3,281	$1,524	$667	$348	$123	$82	$18	$40	$0	$0	$0	$0
1975	6,617	11,306	13,773	14,386	10,593	4,234	2,110	1,051	436	353	93	101	10	5	5	3	3	
1976	7,658	11,064	13,655	13,352	7,592	4,064	1,895	1,003	683	384	216	102	93	57	50	33		
1977	8,735	14,318	14,897	12,978	7,741	4,355	2,132	910	498	323	176	99	101	32	14			
1978	8,722	15,070	15,257	11,189	5,959	3,473	1,531	942	547	286	177	61	67	7				
1979	9,349	16,470	14,320	10,574	6,561	2,864	1,328	784	424	212	146	113	38					
1980	11,145	16,351	14,636	11,273	5,159	2,588	1,290	573	405	134	81	54						
1981	10,933	15,012	14,728	9,067	5,107	2,456	1,400	584	269	120	93							
1982	13,323	16,218	12,676	6,290	3,355	1,407	613	398	192	111								
1983	13,899	16,958	12,414	7,700	4,112	1,637	576	426	331									
1984	14,272	15,806	10,156	8,005	3,604	791	379	159										
1985	13,901	15,384	12,539	7,911	3,809	1,404	827											
1986	15,952	22,799	16,016	8,964	2,929	1,321												
1987	22,772	24,146	18,397	8,376	3,373													
1988	25,216	26,947	17,950	8,610														
1989	24,981	30,574	19,621															
1990	30,389	34,128																
1991	28,194																	

Cumulative Incurred Severity

Accident	Months of Development																	
Year	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	$2,899	$3,799	$5,093	$4,891	$4,838	$5,019	$5,102	$4,829	$4,708	$4,674	$4,632	$4,627	$4,618	$4,622	$4,615	$4,615	$4,615	$4,615
1975	$3,122	$4,275	$4,970	$5,328	$5,517	$5,326	$5,138	$5,013	$4,929	$4,919	$4,910	$4,911	$4,906	$4,905	$4,905	$4,905	$4,910	
1976	$3,421	$3,730	$4,894	$5,587	$5,568	$5,609	$5,411	$5,357	$5,334	$5,300	$5,283	$5,278	$5,277	$5,271	$5,274	$5,270		
1977	$3,122	$4,187	$4,848	$5,256	$5,415	$5,301	$5,188	$5,188	$5,145	$5,133	$5,114	$5,101	$5,102	$5,098	$5,095			
1978	$3,637	$5,015	$5,898	$6,326	$6,324	$6,273	$6,127	$6,081	$6,067	$6,068	$6,067	$6,035	$6,037	$6,030				
1979	$2,813	$4,284	$5,311	$5,588	$5,647	$5,593	$5,554	$5,513	$5,469	$5,456	$5,454	$5,452	$5,445					
1980	$2,667	$3,999	$4,831	$5,313	$5,186	$5,109	$5,052	$5,043	$5,036	$5,043	$5,037	$5,034						
1981	$2,840	$4,087	$5,033	$5,272	$5,305	$5,213	$5,185	$5,163	$5,135	$5,118	$5,115							
1982	$3,420	$4,697	$5,316	$5,232	$5,283	$5,272	$5,258	$5,249	$5,229	$5,224								
1983	$3,270	$4,286	$4,867	$5,146	$5,267	$5,269	$5,273	$5,269	$5,269									
1984	$3,886	$4,844	$5,238	$5,791	$5,848	$5,721	$5,717	$5,705										
1985	$4,246	$5,450	$6,370	$6,367	$6,369	$6,336	$6,357											
1986	$5,031	$7,106	$7,759	$7,868	$7,774	$7,742												
1987	$5,894	$6,982	$8,349	$8,372	$8,432													
1988	$6,455	$7,724	$9,275	$9,230														
1989	$5,760	$8,741	$9,515															
1990	$7,461	$9,232																
1991	$8,320																	

Cumulative Closing Rate

Accident	Months of Development																	
Year	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	192	204	216
1974	0.278	0.532	0.564	0.579	0.653	0.812	0.902	0.954	0.976	0.986	0.994	0.997	0.999	0.999	1.000	1.000	1.000	1.000
1975	0.419	0.477	0.558	0.571	0.716	0.861	0.924	0.964	0.984	0.994	0.998	0.999	0.999	1.000	1.000	1.000	1.000	
1976	0.316	0.466	0.545	0.642	0.795	0.892	0.946	0.967	0.984	0.990	0.994	0.996	0.997	0.998	0.998	0.999		
1977	0.296	0.503	0.594	0.695	0.830	0.901	0.944	0.976	0.989	0.994	0.995	0.996	0.996	0.999	0.999			
1978	0.230	0.503	0.608	0.730	0.842	0.913	0.963	0.982	0.989	0.994	0.997	0.998	0.999	1.000				
1979	0.363	0.610	0.732	0.795	0.881	0.946	0.976	0.989	0.994	0.996	0.997	0.998	0.999					
1980	0.386	0.674	0.766	0.843	0.924	0.968	0.988	0.992	0.996	0.998	0.999	0.999						
1981	0.456	0.669	0.753	0.859	0.930	0.967	0.983	0.992	0.996	0.998	0.998							
1982	0.458	0.686	0.796	0.875	0.946	0.975	0.990	0.993	0.997	0.998								
1983	0.457	0.726	0.842	0.907	0.953	0.981	0.991	0.995	0.998									
1984	0.428	0.707	0.852	0.895	0.956	0.986	0.994	0.998										
1985	0.484	0.751	0.823	0.893	0.954	0.983	0.991											
1986	0.492	0.685	0.831	0.893	0.967	0.985												
1987	0.466	0.704	0.844	0.913	0.967													
1988	0.457	0.707	0.846	0.914														
1989	0.506	0.729	0.862															
1990	0.460	0.726																
1991	0.415																	

Exposures

Accident Year	Earned Exposures
1974	11,000
1975	11,000
1976	11,000
1977	12,000
1978	12,000
1979	12,000
1980	12,000
1981	12,000
1982	11,000
1983	11,000
1984	11,000
1985	11,000
1986	12,000
1987	13,000
1988	14,000
1989	14,000
1990	14,000
1991	13,000

1. Estimated Ultimates Using Paid Chain Ladder Models for Unmodified Data

Comparison of Estimated Ultimate Losses

		Using Unweighted (i.e. Simple) Averages for			Using Volume-Weighted Averages for		
Accident Year	Actual Ultimates	All Years (1)	1986-1991 (2)	Latest 3 Diagonals (3)	All Years (4)	1986-1991 (5)	Latest 3 Diagonals (6)
1974	19,256	19,246	19,568	19,246	19,246	19,934	19,246
1975	23,161	23,159	23,695	23,159	23,159	24,302	23,159
1976	26,400	26,415	27,192	26,415	26,417	28,086	26,417
1977	30,049	30,070	31,182	30,070	30,072	32,448	30,072
1978	31,991	32,019	33,458	32,015	32,020	35,091	32,020
1979	34,529	34,577	36,431	34,586	34,581	38,527	34,601
1980	35,984	36,052	38,316	36,042	36,053	40,879	36,066
1981	35,207	35,283	37,878	35,239	35,279	40,794	35,285
1982	34,418	34,590	37,471	34,473	34,574	40,766	34,504
1983	38,354	38,107	41,636	37,938	38,084	45,793	37,874
1984	37,175	37,799	41,645	37,470	37,739	46,348	37,392
1985	36,446	37,418	41,235	36,568	37,289	46,490	36,478
1986	46,777	50,028	53,777	47,616	49,475	61,506	47,268
1987	60,676	71,835	70,188	63,411	68,911	80,276	62,628
1988	75,418	103,610	90,885	82,657	95,093	103,702	80,904
1989	88,115	138,951	108,937	99,075	120,591	123,201	94,869
1990	90,938	170,321	123,954	110,560	138,214	136,614	100,918
1991	74,807	197,512	146,515	131,064	151,661	156,758	112,010
All Years	819,701	1,116,993	1,003,961	917,604	1,008,460	1,101,516	881,712

2. Estimated Ultimates Using Incurred Chain Ladder Models for Unmodified Data

Comparison of Estimated Ultimate Losses

		Using Unweighted (i.e. Simple) Averages for			Using Volume-Weighted Averages for		
Accident Year	Actual Ultimates	All Years (1)	1986-1991 (2)	Latest 3 Diagonals (3)	All Years (4)	1986-1991 (5)	Latest 3 Diagonals (6)
1974	19,256	19,246	19,851	19,246	19,246	19,943	19,246
1975	23,161	23,162	24,166	23,162	23,162	24,321	23,162
1976	26,400	26,448	27,916	26,448	26,450	28,145	26,450
1977	30,049	30,076	32,172	30,076	30,077	32,496	30,074
1978	31,991	31,994	34,721	32,005	31,997	35,142	32,001
1979	34,529	34,550	38,082	34,542	34,548	38,624	34,538
1980	35,984	35,981	40,341	35,988	35,982	41,008	35,978
1981	35,207	35,183	40,215	35,163	35,181	40,979	35,210
1982	34,418	34,332	40,115	34,371	34,344	40,985	34,411
1983	38,354	37,755	45,196	37,856	37,780	46,307	37,856
1984	37,175	36,758	45,411	37,007	36,821	46,672	37,053
1985	36,446	36,045	46,251	36,589	36,183	47,700	36,637
1986	46,777	45,890	61,318	47,020	46,069	63,487	47,092
1987	60,676	59,455	82,816	60,573	59,577	86,128	61,020
1988	75,418	74,122	101,659	74,436	74,101	105,854	74,995
1989	88,115	88,563	114,987	84,195	87,227	119,612	84,445
1990	90,938	100,739	126,334	93,923	97,147	131,190	92,393
1991	74,807	95,766	120,777	92,183	91,612	125,174	93,242
All Years	819,701	846,066	1,042,326	834,785	837,504	1,073,766	835,801

3. Estimated Ultimates Using Claim Count * Average Severity Models for Unmodified Data

Comparison of Estimated Ultimate Losses

		Using Unweighted (i.e. Simple) Averages for			Using Volume-Weighted Averages for		
Accident Year	Actual Ultimates	All Years (1)	1986-1991 (2)	Latest 3 Diagonals (3)	All Years (4)	1986-1991 (5)	Latest 3 Diagonals (6)
1974	19,256	19,246	19,251	19,241	19,246	19,267	19,246
1975	23,161	23,162	23,171	23,156	23,162	23,198	23,162
1976	26,400	26,448	26,444	26,424	26,449	26,486	26,449
1977	30,049	30,076	30,099	30,056	30,077	30,146	30,077
1978	31,991	31,994	32,023	31,991	31,995	32,106	32,006
1979	34,529	34,550	34,604	34,529	34,549	34,716	34,542
1980	35,984	35,981	36,072	36,009	35,981	36,217	35,988
1981	35,207	35,183	35,336	35,191	35,179	35,512	35,158
1982	34,418	34,332	34,580	34,471	34,333	34,793	34,366
1983	38,354	37,755	38,145	37,986	37,758	38,437	37,850
1984	37,175	36,758	37,435	37,159	36,423	37,793	37,002
1985	36,446	36,046	37,130	36,837	36,059	37,580	36,583
1986	46,777	45,890	47,766	47,402	45,883	48,513	47,017
1987	60,676	59,454	61,889	60,945	59,390	63,764	60,612
1988	75,418	74,119	75,970	75,454	73,902	77,974	74,491
1989	88,115	88,560	85,931	84,623	87,224	89,048	84,232
1990	90,938	100,651	94,405	87,956	97,675	96,800	93,612
1991	74,807	95,812	90,291	66,330	91,218	92,108	91,139
All Years	819,701	846,017	840,542	805,760	836,503	854,455	833,531

1. Estimated Ultimates Using Paid Chain Ladder Models for Modified Data

Comparison of Estimated Ultimate Losses

		Using Unweighted (i.e. Simple) Averages for			Using Volume-Weighted Averages for		
Accident Year	Actual Ultimates	All Years (1)	1986-1991 (2)	Latest 3 Diagonals (3)	All Years (4)	1986-1991 (5)	Latest 3 Diagonals (6)
1974	19,256	19,246	19,291	19,246	19,246	19,287	19,246
1975	23,161	23,127	23,208	23,127	23,127	23,196	23,127
1976	26,400	26,397	26,525	26,397	26,397	26,503	26,397
1977	30,049	30,070	30,266	30,070	30,070	30,230	30,070
1978	31,991	32,034	32,296	32,033	32,035	32,243	32,034
1979	34,529	34,573	34,910	34,575	34,576	34,835	34,578
1980	35,984	36,064	36,513	36,050	36,065	36,411	36,054
1981	35,207	35,289	35,871	35,245	35,287	35,743	35,250
1982	34,418	72,437	73,846	72,168	72,411	73,513	72,177
1983	38,354	37,482	38,369	37,306	37,446	38,152	37,286
1984	37,175	0	0	0	0	0	0
1985	36,446	36,720	37,813	36,213	41,679	37,492	36,228
1986	46,777	51,611	52,694	50,296	63,723	52,173	50,298
1987	60,676	72,276	70,996	67,765	95,852	70,295	67,767
1988	75,418	101,427	97,000	92,713	146,063	96,042	92,680
1989	88,115	17,857	15,281	14,605	26,622	15,108	14,579
1990	90,938	252,378	150,636	83,398	124,043	45,375	35,161
1991	74,807	373,332	275,072	213,186	206,227	83,022	74,945
All Years	819,701	1,252,320	1,050,588	904,394	1,050,869	749,622	717,876

2. Estimated Ultimates Using Incurred Chain Ladder Models for Modified Data

Comparison of Estimated Ultimate Losses

		Using Unweighted (i.e. Simple) Averages for			Using Volume-Weighted Averages for		
Accident Year	Actual Ultimates	All Years (1)	1986-1991 (2)	Latest 3 Diagonals (3)	All Years (4)	1986-1991 (5)	Latest 3 Diagonals (6)
1974	19,256	19,246	19,269	19,246	19,246	19,270	19,246
1975	23,161	23,130	23,171	23,130	23,130	23,173	23,130
1976	26,400	26,430	26,494	26,430	26,430	26,497	26,430
1977	30,049	30,076	30,181	30,076	30,075	30,186	30,075
1978	31,991	32,014	32,161	32,023	32,015	32,167	32,022
1979	34,529	34,528	34,745	34,531	34,528	34,753	34,531
1980	35,984	35,980	36,285	35,996	35,982	36,296	35,994
1981	35,207	35,174	35,594	35,170	35,172	35,609	35,172
1982	34,418	72,060	73,262	72,165	72,087	73,303	72,169
1983	38,354	36,938	37,795	37,059	36,966	37,823	37,058
1984	37,175	0	0	0	0	0	0
1985	36,446	35,585	37,188	35,889	40,695	37,240	35,899
1986	46,777	47,671	50,952	47,740	59,299	51,055	47,752
1987	60,676	59,494	68,032	50,833	79,368	68,170	49,353
1988	75,418	82,976	87,544	78,614	118,947	87,984	73,352
1989	88,115	32,635	33,488	30,072	50,573	33,786	28,167
1990	90,938	118,751	95,935	65,340	124,885	55,935	39,216
1991	74,807	143,319	129,319	108,501	174,007	78,593	62,545
All Years	819,701	866,008	851,415	762,815	993,404	761,840	682,111

3. Estimated Ultimates Using Claim Count * Severity Models for Modified Data

Comparison of Estimated Ultimate Losses

		Using Unweighted (i.e. Simple) Averages for			Using Volume-Weighted Averages for		
Accident Year	Actual Ultimates	All Years (1)	1986-1991 (2)	Latest 3 Diagonals (3)	All Years (4)	1986-1991 (5)	Latest 3 Diagonals (6)
1974	19,256	19,246	19,268	19,246	19,246	19,268	19,246
1975	23,161	23,130	23,168	23,130	23,130	23,168	23,130
1976	26,400	26,430	26,490	26,430	26,430	26,490	26,430
1977	30,049	30,076	30,174	30,076	30,076	30,174	30,076
1978	31,991	32,014	32,150	32,023	32,014	32,151	32,014
1979	34,529	34,528	34,729	34,531	34,528	34,730	34,528
1980	35,984	35,980	36,264	35,996	35,981	36,265	35,990
1981	35,207	35,174	35,566	35,170	35,171	35,569	35,160
1982	34,418	72,060	73,186	72,165	72,062	73,197	72,137
1983	38,354	36,938	37,742	37,059	36,940	37,752	37,018
1984	37,175	0	0	0	0	0	0
1985	36,446	35,585	37,083	35,889	50,379	37,111	35,782
1986	46,777	47,671	50,733	47,740	82,246	50,803	48,046
1987	60,676	59,484	67,741	50,841	121,685	67,833	55,363
1988	75,418	83,006	87,168	78,620	207,203	87,359	82,310
1989	88,115	32,645	33,351	30,081	98,442	33,867	32,343
1990	90,938	118,391	96,136	65,868	229,538	48,991	50,504
1991	74,807	151,739	131,452	109,459	481,966	74,370	76,666
All Years	819,701	874,098	852,400	764,323	1,617,037	749,097	726,744