ANCHORAGE CHAMBER OF COMMERCE

STATE & LOCAL GOVERNMENT AFFAIRS COMMITTEE

ISSUE PAPER & RECOMMENDATIONS

USE OF PERSONAL CREDIT HISTORIES IN SETTING INSURANCE RATES

April 2003

TABLE OF CONTENTS

THE USE OF PERSONAL CREDIT HISTORIES
IN SETTING INSURANCE RATES

PART 1. SOME ABCS ABOUT INSURANCE.

How Insurance Works. Insurance is about risk. In particular, it is about the risk of something bad happening. What that "something" might be depends on the type of insurance involved. For life insurance the risk is the chance that the insured person will die during the term of the policy. For health insurance it's the risk that s/he will become ill or develop some other condition requiring medical treatment. For liability insurance it's the risk that s/he might accidentally cause some kind of injury or harm to the person or property of another.

Bad things like this happen somewhere, to somebody, every day. They just don't happen every day to particular individuals. In fact, for individuals these are usually uncommon or even rare occurrences. It is only when we look at a large number of people together, that these events become regular or common occurrences for the group as a whole.

Insurance is a way for individuals to act collectively to protect themselves against events that are unlikely to happen to any particular one of them individually, but which would have severe or ruinous consequences for a person if they do occur. It is a way to avoid potential catastrophe as an individual by paying a small share of the total cost of the consequences suffered by individual members of the group. By dividing up among its members the costs for the group as a whole with respect to a particular risk, each individual in the group can be protected against that risk at an affordable cost. This, in a nutshell, is how insurance works — it is a group of people who are sharing the total cost to them collectively that is incurred as a result of a particular danger or harm experienced by individuals within that group.

In a sense insurance is like a lottery against misfortune in life. You buy a lottery ticket, and it pays off only if a particular misfortune befalls you. If you know what the odds are of winning this lottery by having that misfortune, you can figure out what price the lottery tickets should be in order to provide lottery payoffs that match the harm suffered the lottery winners who experience that misfortune.

Accounting for differences among individuals in the risk that each represents. Although the basic idea of insurance is quite simple, it is complicated by the fact that, for any given danger or harm to be insured against, individual people may have very different chances of suffering that harm. All of us are going to die eventually, for instance, but the chances of dying sometime during the next 12 months are generally much higher for a 90-year-old than for someone only 30. Similarly people living in the flood plain of a large river are more likely to experience water damage to their homes than those living on higher ground, while water damage from frozen water pipes that burst is more likely in Alaskan homes than it is for Hawaii.

When a group's total cost for a particular danger or harm is being divided up among its members, how should such differences among individuals within a group be accounted for? Basically there are only two possible approaches toward answering this question — either to have everyone in the group pay equally, or to have each one pay in proportion to the cost risk[1] that s/he individually represents for the group. The first approach works well when there are only slight or statistically insignificant differences in the cost risk that each individual represents for the group. There is little point, after all, in trying to adjust for a mere fraction of a cent when a person's share might well be in the hundred of dollars. But the greater the differences become among individuals, the more serious the issues of fairness and equal treatment become under such a one-size-fits-all approach. And the greater the fairness issues become, the more reason there is to have people pay in proportion to the risk or cost they represent.

While having each person pay in exact proportion to the cost risk that s/he represents may well solve the fairness issues,[2] the difficulty with this latter approach lies in the details of implementing it. Quite simply, for most kinds of harm and danger that we insure ourselves against (such as major illness, for example), the science does not yet exist to allow each person's individual chances of experiencing that harm or danger to be measured with precise accuracy. The best that can be done is, first, to find associations between a particular danger or harm and various traits or behaviors by individuals that appear before or in conjunction with the occurrence of that danger or harm, and second, to use statistical analysis to measure mathematically how close that association is.[3] The risk that a particular individual represents can then be inferred from how many traits or behaviors s/he has that have been found to be associated with the danger or harm, and from how close those associations are. Traits and behaviors that have relatively close association with a particular danger or harm are often called its "risk factors."

Complementing this analysis of an individual's chances of experiencing a danger or harm from his/her risk factors is another analysis of the range of costs to be incurred when that danger or harm is experienced. When the danger or harm in question is relatively common, there will be lots of empirical data from which a very accurate statistical analysis can be made about the likely cost per occurrence of that particular danger or

---

[1] This cost risk is a function of two variables: 1) the individual person's statistical chance of experiencing the danger or harm at all; and 2) the statistically expected cost incurred for an occurrence of that danger or harm. The statistically expected cost is itself dependent upon, first, the range of costs that could reasonably be incurred when someone experiences that danger or harm, and second, the statistical probabilities that the cost actually incurred for any given experience will fall at or above various levels within that range. Rather than being fixed dollar amounts, those levels are usually derived from the cost range itself; e.g., first, second and third quartiles (the levels at or below which 25%, 50% or 75% of the cost outcomes, respectively, statistically occur) or deciles (levels at or below which 10%, 20% etc. of the cost outcomes statistically occur).

[2] It fails to solve all the fairness issues if people within a group cannot afford to pay their true shares of the group's total cost for the particular danger or harm. For them the new system would not seem fair and — depending on one's social philosophies — might not in truth be fair as well.

[3] The closeness of an association is a function of the frequency with which the occurrence of the trait or behavior is followed or accompanied by an occurrence of the danger or harm. The more often the former is followed or accompanied by the latter, the closer the association between them.

harm. Conversely, costs per occurrence of very rare events may be very difficult to analyze statistically and have to be estimated by other methods. Once the expected cost per occurrence is analyzed, it can be combined with the statistical chances of experiencing such an occurrence for someone with this set of risk factors to determine the cost risk for the particular individual in question and anyone else with the same risk factors.

So when group-wide costs for a given danger or harm are divided up among the group members, what goes on is a compromise between theory and practical reality. The theory calls for a precise measuring of the actual cost risk posed by each individual, and then each person's share of the group's cost is set in proportion to the actual cost risk which that person represents. Practical reality, however, does not yet permit this to be done. Instead, individuals are first aggregated into subgroups that are each defined by a particular set of risk factors that its members all have in common.[4] The whole group's total cost for the danger or risk in question is then allocated among these various subgroups on the basis of *a*) the relative degree of cost risk statistically represented by each subgroup's set of risk factors, and *b*) the number of people in each subgroup. Finally, the costs allocated to a particular subgroup are shared equally among the individuals within that subgroup.

<u>*Making money as an insurer without overcharging*</u>. After reading or hearing about the profits that some large insurance companies report to their shareholders, many people believe the only way those companies could be making so much money is that they must be overcharging the consumer for their insurance. While one may not be able to guarantee that overcharging never happens, the fact is that the insurance business by its very nature can allow a large insurer with only average luck to make fairly sizeable profits without overcharging. Here's a simplified example that illustrates how.

EXAMPLE

Suppose you are setting up an insurance organization to cover the costs of homeowners to repair their homes if there is a flood, and you want to set it up so that statistically over time the total that your subscribers pay into the pool for this coverage will exactly equal the total amount paid out from the pool for their flood repair costs — in other words, you're setting it up to be strictly a break-even operation over time statistically. Now to avoid getting distracted by the arithmetic, let's make the following simplifying assumptions: One, you have 100,000 subscribers, and they all make their payments into your insurance pool at the beginning of each year. Two, it costs $10,000 to repair flood

---

[4] For those kinds of danger or harm where the costs incurred can vary widely from one person to the next, one would also identify risk factors that are associated either with high-cost outcomes or low-cost ones, and would measure the closeness of each association. If not already included among the risk-factors for experiencing the danger or harm in the first place, these additional factors relating to amount of cost incurred would probably be added to the risk factors being assigned to the various subgroups. At the same time, however, in setting up the various subgroups one should be prudent to avoid making unnecessary combinations of all the various risk factors just for the sake of making combinations. The mere fact that some combinations are analytically possible does not necessarily mean that their associated subgroups should be established for cost-division purposes — especially if those subgroups would have very few or no people in them.

damage each time it occurs,[5] no matter whose house is damaged or where it may be located. Three, there cannot be more than one flood in any given year. And four, every flood that does occur is always one of three types, and each type of flood is exactly the same whenever it occurs — 5-year floods occur once every five years on average[6] and 5% of your subscribers (5,000 homes) have flood damage, 20-year floods occur once every 20 years on average and 20% of your subscribers (20,000 homes) have flood damage, and 100-year floods occur once a century on average and half your subscribers (50,000 homes) have flood damage.

Your first task is to figure out how much to charge your subscribers so the total paid into the insurance pool will exactly equal the amount paid out from it over the long run. The amount you should collect each year equals the chances of each type of flood occurring that year times the damage it will cause if it does occur. Here is that calculation:

Type of Flood	Chance It Occurs	Payout per Occurrence	Payout times Probability	
5-year	$^1/_5$ or 20%	\$10,000×5,000 = \$50,000,000	$^1/_5$ × \$50,000,000 =	\$10,000,000
20-year	$^1/_{20}$ or 5%	\$10,000×20,000 = \$200,000,000	$^1/_{20}$ × \$200,000,000 =	\$10,000,000
100-year	$^1/_{100}$ or 1%	\$10,000×50,000 = \$500,000,000	$^1/_{100}$ × \$500,000,000 =	\$5,000,000
	Total you should collect each year:			\$25,000,000

Dividing this annual amount among your 100,000 subscribers, you find that you should charge them \$2,500 a year apiece in order to break-even over the long term.

Now, then, how can you make money in a system like this which has been deliberately designed just to break even over the long term? The answer is, you can invest the money in the pool from the time it is paid in until the time you have to pay it out. Okay, but how long are you likely to be able to invest that money before you have to pay it out? After all, it is possible that you'll have a 5-year flood in year 1, and you'll have to make a \$50 million payout when you will have collected only \$10,000,000 to cover that particular contingency (even for all three contingencies, you will have collected only \$25,000,000 — half of what you'd need to pay out). And after that happens, it will still be possible to have a 100-year flood the very next year. So how long, really, are you likely to be able to hold each year's payment into the pool until you have to pay it out?

To answer this, you first need to know what the chances are each year of not hav-

[5] In other words, it doesn't cost \$10,000 per instance on average, but \$10,000 each and every time. We could just as well set up a more realistic range of flood-repair costs and establish probabilities for each cost level so that they average \$10,000 per instance. It would get us to the same conclusion, but the arithmetic would be more complicated. The point to be made is not in the number crunching, but in the results of those calculations, so it's simplest just to assume that the average repair cost is the actual cost incurred in each individual instance.

[6] It is important to note that we are not assuming 5-year floods actually will occur once every five years, but only that statistically that is how often they occur on average. It would be quite possible to go for 20 years, say, without a single 5-year flood, just as it is possible to have 5-year floods each year for 4 years in a row. We are not making an assumption about when 5-year floods will actually occur, but only about how often they occur on average over a long period.

ing to make a payout that year. For 5-year floods the chance of having one is one in five by definition, so the chance of not having one is four out of five, or 80 percent. In this 80% of the time when a 5-year flood doesn't occur in a given year, there is still a possibility of having either a 20-year flood or a 100-year flood that year. The chance of a 20-year flood is one in 20, so the chance of not having one is 19 out of 20, or 95 percent. This means that within the 80% when there is no 5-year flood that year, there is a 95% chance of also not having a 20-year flood. So the chance of not having either kind of flood in a that year is 80% × 95%, or 76 percent. That still leaves the possibility of a 100-year flood. Of the 76% when neither a 5-year flood nor a 20-year one occurs in a year, there is one chance in 100 of having a 100-year flood and a 99% chance that there won't be. So the chance of not having any flood at all during any given year is 76% × 99% or 75.24 percent.[7]

Now, then, if there is a 75.24% chance each year of not having to make any payout at all, how long can you expect on average to be able to invest the money in the pool until you do have to make a payout? Statistically the average length of this investment period is the time it takes until the odds of making a payout or not are 50-50.

In the first year the odds of a payout are 75-25 against making one.[8] To figure how long it takes for the odds to drop from 75-25 to 50-50, we have to go beyond arithmetic to algebra (sorry). If we let n represent the number of years it takes for the chances of making a payout to get to 50%, then 75.24% to the n-th power (or 0.7524^n) equals 50% (or 0.50). When we solve this equation for n, we find that n equals 2.43648 years.[9] In other words, the average length of time you can expect to hold a given year's payments in the pool is 2.43648 years before there is a 50% that you will pay something out.

---

[7] Perhaps you are thinking that, by first considering the chance of not having a 5-year flood and then the chances of a 20-year flood and finally a 100-year one, we are omitting analyses where we would start first by considering the odds of not having one of the other two kinds of flood. However, it does not matter which order we use in considering the chances for the different kinds of floods. For instance, let's reverse the order: for a 100-year flood, there is a 99% chance that it won't occur during a given year. Of that 99% chance, there is a one-in-20 chance that a 20-year flood occurs, so the chance of neither one occurring is 99% × 95%, or 94.05 percent. Within this 94.05% chance, there is an 80% chance of not having a 5-year flood. So the chance of not having any flood is 94.05% × 80% = 75.24%, which is the same as what we got before as the chance of not having any flood in a given year. The reason this is the same result as the first one — and the reason why it doesn't matter what order we use in considering the chances of not having these floods — is that we always end up multiplying the same three numbers together: the 99% chance of not having a 100-year flood, the 95% chance of not having a 20-year one, and the 80% chance of not having a 5-year one. All that changes from one analysis to another is the order in which we take up these percentages. However, 99% × 95% × 80% always yields the same final result no matter what order we use in doing the multiplication.

[8] Strictly speaking these odds are 75.24–24.76 against making a payout.

[9] This solution is found as follows: We start with $0.7524^n = 0.50$ (Equation 1). Taking the logarithm of both sides yields a new equation, $\log(0.7524^n) = \log(0.50)$ (Equation 2). Since the logarithm of a number raised to the n-th power equals n times the logarithm of the number itself, we know that $\log(0.7524^n) = n\log(0.7524)$ (Equation 3). This means we can substitute $n\log(0.7524)$ for the $\log(0.7524^n)$ that appears in Equation 2, which yields a new equation, $n\log(0.7524) = \log(0.50)$ (Equation 4). If we divide both sides of Equation 4 by log(0.7524), we get $n = {}^{\log(0.50)}/_{\log(0.7524)}$ (Equation 5). Since log(0.50) equals –0.30103 and log(0.7524) equals –0.12355, then n equals –0.30103 divided by –0.12355, or 2.43648. *Q.E.D.*

Each year $25,00,000 is paid into your insurance pool, and on average you can invest it for 2.43648 years. How much will you earn while you are holding that money? It depends, of course, on how well your investments do, but let's assume that you can earn the same 8% rate that the IRS assumes private foundations and charitable trusts can earn.[10] Of course you earn this not only on the principal in the pool, but also on its earnings as they are received. So over 2.43648 years, the principal plus interest for the first year's deposits into the pool will be $\$25,000,000 \times (1.08)^{2.43648}$, or $30,156,176. In other words, the $25,000,000 earns you $5,156,176 on average during the 2.43648 years that you hold it before paying it out.

This example obviously does not accurately represent any particular insurance pool in real life, because of the assumptions we made to keep the mathematics relatively simple and straightforward to follow. However, it is adequate to illustrate the key point. The investment of deposits in an insurance pool during the time they remain in the pool can yield income amounting to a significant percentage of those deposits, even though the pool is deliberately designed only to break even statistically. In this example, the investment income from each year's deposits represents slightly more than one-fifth of the total deposited.

This demonstrates that, if you want to make money as an insurer without overcharging your policyholders,[11] you really can.[12, 13]

---

[10] In order for a private foundation or charitable trust to qualify for tax-exempt status, IRS regulations require it to pay out annually 5% of the market value of its assets at the beginning of each year. Then in order for such a foundation or trust to be permanent, its real earnings and capital appreciation per year have to average at least 5% above inflation so that the next year's 5% withdrawal won't be from principal. For a 5% return after inflation at 2.86%, the return before inflation is $(1.05 \times 1.0286) - 1 = 0.0800$ or 8 percent.

[11] Some might claim that you as the insurer under this analysis <u>would</u> be overcharging your policyholders, because your earnings on the money held in pool are not factored into the calculation of what your policy-holders pay — in other words, they would say that making $5,156,176 on average over 2.43648 years on the $25 million paid each year into the insurance pool is excessive. However, we started with the premise that a "lottery against misfortune in life" (*see* p. 1, *supra*) is fair if the total that "lottery ticket" purchasers pay into the "lottery" is equal to what it is expected to pay back out statistically over the long run. Overcharging under that understanding only occurs if the purchasers are being charged something more than this long-term break-even price for the "lottery tickets." So to say now that your earnings from investing the ticket-sales proceeds must be included in figuring the "ticket" price is a fundamental change in the concept of "overcharge" from what we started with. We started by setting the "ticket" price solely on the basis of the risk and rewards inherent in the "lottery" itself so that it breaks even statistically. But these critics' new approach would set that price, not just on the characteristics of the "lottery" itself, but also on the basis of an additional philosophical opinion about what an insurer's "fair" profit should be, which is more properly associated with utilities operating as regulated monopolies than with competitive free enterprise.

[12] As another simplifying assumption, the example does not explicitly include the costs of operating your flood insurance pool. That cost would also have to be covered by the premiums you would charge in order for it to be breaking even statistically.

[13] The foregoing analysis implicitly assumes that you start out with the means to be able to stay in the business until enough years go by for the odds to even out statistically. This is why it repeatedly refers to "statistical" results and to occurrences "on average." But in the real world the fact that statistics are on your side over the long term is no guarantee that in the short term you won't suffer a run of back luck at the outset. Even though statistically an opening run of bad luck should eventually be offset, if you don't have

PART 2. RISK ANALYSIS — SOME ABCS ABOUT STATISTICS.

The crucial importance of knowing the risks as accurately as possible. The Introduction to this report goes on at such length describing the fundamental principles of how insurance works, in order to make it clear why the heart of the insurance business lies in knowing as precisely as possible what the risks are that you're dealing with. An accurate assessment of the risks is essential if the payments into the system are going to match the payments out of it over the long run. The more accurate the assessment of the risk is, the more accurately the policyholders' payments can be tailored to keep the system balanced without overcharging them.

The best way to determine what the risks are for a given kind of danger or harm is to see what actually happens in the real world. Scientific or empirical evidence about the how often the danger or harm occurs, and about how severe it is when it does occur, is the best foundation for evaluating the risks for that particular danger or harm. Otherwise the risks can only be estimated. Given the central importance that risk plays in insurance, the penalties for underestimating the risks can be severe for an insurer. Thus, when only estimates or "gut feelings" are available about a risk, the evaluation of that risk is more likely to be conservatively pessimistic rather than optimistic. To the extent the evaluation of that risk makes it out to be worse than it actually is, the evaluation will lead to more money being paid into the insurance pool for that risk that is actually necessary. This is because the pessimistic assessment, by definition, overstates the frequency and/or severity of that risk and hence also overstates the expected statistical payout to cover that risk. As a result the size of the insurance pool collected from policyholders is larger than it actually needs to be. The degree to which it is too large depends on how much the insurer has to rely on estimates and "gut feelings" to assess the risk, instead of solid empirical evidence from the real world.

The amount that people pay for insurance decreases as the accuracy increases in an insurer's knowledge and understanding of the risk being covered.

Cause and effect, and the statistical concept of "correlation". We tend naturally to view things as happening in the physical world as the result of other things happening which cause them. We see an apple fall from a tree, for example, and know that gravity (the cause) attracts the apple toward the earth and makes it fall down (the effect).

There are two conditions that must be met before two things have a true cause-and-effect relationship. One is that when one thing happens, the second always follows. If the second thing can happen without the first, then when that occurs the first thing ob-

---

the means to stay in business long enough for that offset to occur, such a run of back luck would break you. This is why, historically, Lloyd's of London and similar insurers started out as a syndicate of wealthy individuals who each undertook to cover a share of the risk to be insured — by taking small enough individual shares, their fortunes would be large enough for them to stay in business long enough to get the odds on their side, while at the same time the whole risk could be insured if enough individuals signed on for a share of it (literally they "subscribed" from the Latin words *sub* meaning "under" and *scribere* "to write" — hence the term "underwrite"). Lloyd's remains such a syndicate of underwriters to this day. (For more information about Lloyd's, see their web page at www.lloyds.com and click on "Go to About Lloyd's".)

viously cannot be causing the second. The other condition is that if the first thing does not happen, then second one never happens either. When this condition is met, it eliminates the possibility the reason why the two things are always observed together is that a third thing is causing them to happen.

Unfortunately, the complexity of the real world often makes it impossible to discover or prove a direct cause-and-effect relationships between specific things. Take smoking, for example. We all know now that smoking causes lung cancer, yet not everyone who smokes gets lung cancer. This appears to violate the first condition for a cause-and-effect relationship: that when the first thing happens (smoking), the second (lung cancer) is always supposed to follow. So how do we know that smoking "causes" lung cancer?

The answer lies in statistics. If smoking and lung cancer are not causally related, then statistically the percentage of smokers who get lung cancer should be the same as the percentage of non-smokers who get it. In addition, the percentage of lung cancer patients who are or were smokers should be the same statistically as the percentage of lung cancer patients who never smoked. These two percentages — the percentage of smokers getting lung cancer and the percentage of cancer patients who smoke(d) — do not have to be identical with those of the non-smokers in order to for the statistics to show there is no causal relationship between smoking and lung cancer. This is because random variations in any sample of empirical data will cause some deviation from the norm, and so some degree of difference between smokers' and non-smokers' percentages should be expected. But the larger that sample becomes (i.e., the more cases in the study), the more powerfully and accurately the statistical analysis will draw the line between simply random variations between the two groups' percentages and those that imply some causal relationship.

We know that smoking causes lung cancer in at least some sense, because the percentage of smokers getting lung cancer is so much higher than the percentage for non-smokers that the chance of this difference being due simply to random variation is most unlikely. Similarly, the percentage of lung cancer patients who are or were smokers is so much greater than that for non-smokers that the difference almost certainly is not due to random variation.

The number of cases upon which these statistical results are based is now quite huge, which makes these statistical conclusions very reliable. This is because, besides measuring the degree of correlation between two variables such as smoking and lung cancer, the mathematics of statistical analysis can also rigorously examine its own results about a correlation, to determine how likely it is that the correlation might only be a statistical oddity instead of being evidence of a true causal relationship. In the case of smoking and lung cancer, the chance that the link between them is just a fluke is extremely small.

The key point of this discussion about smoking and lung cancer is this: From the huge amount of observational data gathered to date about the incidence of lung cancer and who develops it, we know that there has to be a causal relationship between smoking and lung cancer because the differences between smokers versus non-smokers are simply too great to be plausibly explained otherwise. A causal relationship between smoking and lung cancer has been established by overwhelming statistical evidence even though we may not yet be able to identify the precise cause-and-effect mechanism(s) leading step by step from smoking to the development of lung cancer for any given individual.

The statistical concept that scientists have used to demonstrate the causal linkage between smoking and lung cancer is called "correlation." Generally, using the terms we've been using so far, correlation is the degree to which the occurrence or absence of events (i.e., the data about those events) in one category is, or is not, observed to occur in conjunction with the occurrence or absence of events in another category (i.e., data about the second category).[14] There is no requirement beforehand that there be any kind of causal linkage between the two categories of event or their sets of data. In other words, in order to apply this analysis you do not have to have any kind of theory about how two kinds of observed data might somehow be related, nor even a suspicion that they are. Rather, the utility of this statistical concept is that it can be applied to any two kinds of empirical data, and so its use can lead to discoveries of unexpected or even surprising associations between different phenomena.

The degree or strength of a correlation between two variables[15] is most commonly measured by the "correlation coefficient" for them. Each scientific experiment or observation records the values of both variables at the time when that experiment or observation is made. This establishes a natural pairing of values for the two variables through empirical observation. For example, if X_t is the value of variable X that is observed at time t and Y_t is the value of variable Y observed at the same time, then (X_t, Y_t) would be the pairing of the values for these two variables that were empirically observed at time t. The correlation coefficient measures the degree to which changes in the magnitude of one variable are associated with changes in the magnitude of the other. A correlation coefficient of 1.00 (its maximum possible value) means there is a perfect and direct linear relationship between changes in one variable and the changes in the other; i.e., if X_2 is 20% larger than X_1, then Y_2 is 20% larger than Y_1. A correlation coefficient of –1.00 (its minimum possible value) means there is a perfect but inverse linear relationship between the two variables; i.e., if X_2 is 20% larger than X_1, then Y_2 is 20% smaller than Y_1. A correlation coefficient of 0.00 means there is no relationship between the value of one

14 The mathematical analysis underlying this concept is so versatile that its applicability is not limited to specific events— it can be applied to any pair of mathematical variables whether they relate to real phenomena or something else.

15 The term "variable" more precisely describes the mathematical concept here than the term "event" does. Data for "events" will often reflect only two outcomes — the event occurs or it doesn't. The data for such an event will have only two values; e.g., one if the event occurs and zero if it doesn't. But unlike this simplistic event with only two values for its data, a variable is more generalized and can be a set of values over any range of numbers instead of being just one or zero or the numbers in between. Thus the statistical analysis not only applies to simple yes-no cases, but can be applied to phenomena where the observed data may take on any degree of shading or nuance that is experimentally feasible to show.

variable and the observed values of the other.[16]

Statisticians also use the square of the correlation coefficient, or "R^2," [17] which is known as the coefficient of determination. It measures the proportion of variance in *Y* that is contained in *X*. Whereas the correlation coefficient can be positive or negative, R^2 is always a positive number because it is a square. This often makes R^2 a more useful measure because it indicates how much of the change in *Y* seems to be due to an increase in *X*, regardless of whether the change in *Y* is positive or negative.

PART 3. SCIENTIFIC EVIDENCE OF A CORRELATION BETWEEN CREDIT HISTORY AND THE INSURANCE RISK THAT A PERSON REPRESENTS.

The SALGA Committee has had available for its review and study four recent papers and presentations about a possible correlation between a person's credit history and the insurance risk which that person represents. These are:

1. James E. Monaghan, ACAS, MAAA, *The Impact of Personal Credit History on Loss Performance in Personal Lines* (2000) (the "Monaghan Study").
2. Robert P. Hartwig, Ph.D., *The Use of Credit Information as an Underwriting Tool in Personal Lines Insurance: Analysis of Evidence and Benefits* (presented to the Texas Department of Insurance, Austin, TX August 26, 2002) (the "Hartwig Presentation"). At the time of his presentation Mr. Hartwig was Senior Vice President & Chief Economist for the Insurance Information Institute, located in New York City.
3. American Academy of Actuaries, Risk Classification Subcommittee of the Property/Casualty Products, Pricing, and Market Committee, *The Use of Credit History for Personal Lines of Insurance; Report to the National Association of Insurance Commissioners* (November 15, 2002) (the "Actuaries-Academy Report"). This paper includes a review and critique of four other papers on the use of credit history for rating and underwriting personal lines of insurance, namely:
    - James E. Monaghan, *supra*;

---

[16] If there are *n* observations of the values of the variables *X* and *Y* and thus *n* pairs of the form (X_i, Y_i) where (X_i, Y_i) is the pair observed in *i*-th observation, then the correlation coefficient for *X* and *Y* is:

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) \Big/ \sqrt{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] * \left[\sum\left(Y_i - \overline{Y}\right)^2\right]},$$

where $\overline{X}$ is the arithmetical mean of the *n* observed values of *X* — i.e., $\overline{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)$ — and $\overline{Y}$ is the arithmetical mean of the *n* observed values of *Y*.

[17] Perversely it seems, the symbol for the correlation coefficient is not R, but the Greek letter rho (ρ), even though the correlation coefficient is the square root of R^2.

- Conning & Company, *Insurance Scoring in Personal Automobile Insurance – Breaking the Silence* (2001);
- Fair, Isaac, *Predictiveness of Credit History for Insurance Loss Ratio Relativities* (1999); and
- Commonwealth of Virginia, State Corporation Commission, Bureau of Insurance, *Use of Credit Reports in Underwriting* (1999).

4. University of Texas at Austin, McCombs School of Business, Bureau of Business Research, *A Statistical Analysis of the Relationship Between Credit History and Insurance Losses* (March 2003) (the "University of Texas Study").

This report will not attempt to recapitulate the details of these four papers/presentations. For that the reader is advised to contact the Anchorage Chamber of Commerce to make arrangements about hard copies or electronic versions of these papers, subject to any legal limitations and requirements under applicable copyright law. However, this paper will summarize the analysis and conclusions of these studies.

The Monaghan Study. This study involved approximately 270,000 new policies of automobile and homeowners insurance written by one insurance company in 1993 and cross-checked them against an archive of personal credit histories kept by an independent national credit vendor for 1993. About 170,000 of the 270,000 named insureds were found in the credit-history archive,[18] so that is the size of the statistical sample. Because of the nature of the data available, the analysis had to be conducted on the basis of each policy instead of the individuals insured under it. Accordingly, the analysis first determined the "loss ratio relativity" for under each policy — that is, the ratio of claims paid out under a given policy during the period 1993-95 versus the premiums earned on that policy during that same period,[19] divided by the mean of all such ratios for the 170,000 policies.[20] This loss ratio relativity was then compared against seven elements (listed below) appearing in the named insured's credit history when the insurance policy was issued. Each person's credit data for each of these elements was paired with his/her loss ratio relativity, and each set of pairs was analyzed for the degree of correlation between the loss ratio relativity and the particular credit element. Here are the results for each credit element:

- amounts past due (i.e., delinquent amounts that are uncollected as of the credit report date): the correlation coefficient of the loss ratio relativity versus the logarithm of the amounts past due was 0.83; the chance that this degree of correlation

---

[18] To comply with the federal Fair Credit Reporting Act, the national credit vendor stripped each person's personal identification information from his/her credit-history record before providing the credit history data for analysis.

[19] This ratio of insured loss incurred versus earned premiums for a policy is called the policy's "loss ratio." It is a direct measurement of how well the insurance company's existing risk-evaluation criteria are working. The higher the ratio is, the worse those criteria are working in predicting how much insurance risk a given individual represents.

[20] This "loss ratio relativity" directly measures how well the insurance company's risk-evaluation criteria worked for individuals versus how well those criteria worked on average in evaluating the 170,000 new policyholders as a whole.

was not due to a fluke in the data sample was over 99.5 percent (Monaghan Study, p. 82).

- derogatory public comments (i.e., liens, bankruptcies, foreclosures): the coefficient of determination between the loss ratio relativity and the fact that there were derogatory public comments and the number of such comments was 0.95, implying a correlation coefficient of 0.975 (*id.*, pp. 82-83).[21]
- collection records (i.e., the fact that a delinquent account has been turned over to a collection agency): the coefficient of determination between the loss ratio relativity and the fact that an account had been turned over to a collection agency was 0.96, implying a correlation coefficient of 0.980 (*id*, p. 83).
- status of trade lines (how long it was after the payment's due date — i.e., less than 30, 30-59, 60-89, 90-119, over 120 days, or "no information" — when a person's most recent payment was made, for each item of credit extended to that person): a correlation coefficient was not reported, but loss ratio relativity was shown to be 21.5% less if no trade line's last payment was 60 days late or more than if one or more lines had such late payments, and loss ratio relativity is shown to be 27.5% less if no trade line is being paid under a wage-earner plan or is in repossession or is written off as a bad debt, than if one or more trade lines are (*id.*)
- age of oldest trade line (i.e., how long a person has been borrowing money from, or been extended credit by, the party who has done so the longest for that person): the coefficient of determination between the loss ratio relativity and the age of the oldest trade-line relationship was 0.86, implying a correlation coefficient of 0.927; the chance that this degree of correlation was not due to a fluke in the data sample was over 99.5 percent (*id.*, p. 84).
- number of non-promotional inquiries (i.e., how many times there were inquiries into the person's credit history other than those associated with a promotional activity such as direct mail marketing campaigns): the correlation coefficient of the loss ratio relativity versus the number of non-promotional credit inquiries was 0.94; the chance that this degree of correlation was not due to a fluke in the data sample was over 99.5 percent (*id.*, pp. 84-85).
- Leverage ratio on revolving-type accounts (i.e., the sum of all outstanding balances for credits cards and similar revolving-type accounts, divided by the sum of the credit limits for those accounts): the coefficient of determination between the loss ratio relativity and the leverage ratio was 0.996, implying a correlation coefficient of 0.998; the chance that this degree of correlation was not due to a fluke in the data sample was over 99.5 percent (*id.*, p. 85).

The Monaghan Study found high to extremely high correlation between 6 of these 7 elements in a person's credit history and his/her loss ratio relativity — in other words, each

---

[21] The correlation coefficient implied from a given R^2 value could be either the positive or negative square root of R^2; however, an inspection of the tabulated data presented in the Monaghan study shows the actual correlation to be positive, not negative. The same is true for the other implied correlation coefficients for the Monaghan study.

element was directly correlated with an increase in the loss ratio as that credit element got worse, always in a fairly linear fashion and almost perfectly so for three of these elements (derogatory comments, having been turned over to a collection agency, and the leverage ratio on revolving credit accounts).[22] The seventh element, status of trade lines, was presented using a different statistical analysis from correlation coefficients and coefficients of determination, but it too indicated that insured losses were higher when the evidence from this credit element was negative. These correlations indicate that the insurance company's risk-evaluation criteria being used in 1993 failed to fully reflect some elements of credit risk that are statistically associated with these seven elements in a person's credit history.

The Hartwig Presentation. Mr. Hartwig appeared before the Texas Department of Insurance as an advocate for the use of credit-history data as an insurance underwriting tool. His presentation compiled research results by others, rather than providing original research results. Given its intended audience, his presentation was much less technical in its statistical analysis than it presumably would have been if it had been made to a group of professional actuaries. However, even in layman's terms and despite the advocacy in his presentation, the statistical data indicate a strong correlation between the individual insurance performance of a named insured and various elements in his/her credit history, including so-called "credit scoring."

The Hartwig Presentation quoted automobile insurance data developed by the Casualty Actuarial Society showing that the "loss ratio relativity" (used in the same sense as in the Monaghan Study) for drivers with "Unacceptable" credit ratings was 133 (i.e., 133% of the average loss ratio), was 103 for drivers with no credit history (or not falling in any of the other three categories), was 91 for drivers with "Good" credit ratings, and only 57.4 for drivers with "Excellent" credit ratings. It also cited similar data from a study by Fair, Isaac[23] showing how the loss ratio relativity for homeowners insurance correlates to certain specific elements in an insured homeowner's credit history.

- For the 4% of all insured homeowners with one or more adverse public records in their credit history, the loss ratio was 1.54 times that for those with no adverse public records.
- For homeowners with one trade line (charge account) that became delinquent 60 days or more during the last 24 months, the loss ratio was 1.293 times that of the 89% of insured homeowners with no 60-day delinquencies. For homeowners with two or more 60-day delinquencies in the past 24 months, the loss ratio shot up to 1.804 times that for homeowners with no 60-day delinquencies.
- For the 40% of insured homeowners opening new trade lines in the past 12 months, the loss ratio relative to that of insured homeowners not opening any new trade lines was —

---

[22] Remember, the more nearly linear a correlation is, the clearer it becomes that there is a relationship between the two variables. In a perfect linear correlation, all the data pairs fall in a completely straight line.

[23] This appears to be the same study as the one reviewed in the American Academy of Actuaries paper that the SALGA Committee has received.

- 1.147 times greater for those opening one new trade line,
- 1.220 times greater for those opening two,
- 1.503 times greater for those opening three, and
- 1.658 times greater for those opening four or more.

The Hartwig Study also presented data gathered by Tillinghast Towers-Perrin about nine samples of data from eight automobile and homeowners insurance companies.[24] The data for eight out of the nine samples indicated that the probability was over 99% that the apparent correlation between credit scores and insurance performance was not a mere statistical fluke, and for the ninth the probability was about 92 percent.

To demonstrate that the correlation between credit history and insurance performance is not a reflection of how much someone's income is,[25] the Hartwig Presentation offered data from the American Insurance Association based on a sample of nearly half a million policyholders, which showed that the mean and median credit scores for people in a particular income bracket remain nearly flat as the bracket increases.

The Actuaries-Academy Report. For SALGA Committee purposes the value of this report lies in its summaries of the findings and conclusions in three additional reports that the Committee does not have, as well as providing the Academy Subcommittee's impartial opinion about the technical soundness of the methods and conclusions of all four reports.[26]

Regarding the Monaghan Study, the Actuaries-Academy Report noted as one of its strengths, "The study uses loss ratio and multivariate analysis to demonstrate that the credit characteristics are adding predictive power, above and beyond the existing variables" that the insurance company had used in deciding what premiums to charge when it agreed to provide coverage to its new customers. Actuaries-Academy Report, p. 9. It notes further that the Monaghan Study "demonstrates that a large number of credit characteristics are adding predictive power, *independent* of one another" (original emphasis). *Id*. The Academy Subcommittee's criticisms of the Monaghan Study were mainly that it could have been broader in design, could have gone into greater technical detail than it did, and stated some conclusions without providing the results of the analysis underlying

---

[24] Each company provided Tillinghast Towers-Perring with data either about homeowners insurance or automobile insurance, except for one that provided sets of data about both kinds of insurance.

[25] State insurance regulators generally do not allow insurance companies to discriminate against people with lower incomes in setting their insurance rates. Since it is conceivable that some elements in a person's credit history could be caused or influenced by how much income s/he has, it was necessary for Mr. Hartwig's case to show that "credit scoring" — which is the particular form of credit-history data that he was advocating — is not income-based discrimination in disguise.

[26] The recommendations of the Academy's report, which were addressed to the National Associate of Insurance Commissioners, addressed 1) how the NAIC should design and conduct an empirical study of its own regarding the statistical relationships between credit history and insurance claims experience and the possible causal factors responsible for those relationships (Actuaries-Academy Report, pp. 28 – 35) , and 2) what the best practices are for state insurance regulators in reviewing and regulating insurance companies' uses of credit information in setting insurance rates to their policyholders (*id*., pp. 36 – 38).

those conclusions. *Id.*, pp. 9 – 10.

Regarding the study of automobile insurance by Conning & Company (the "Conning Study"), the Actuaries-Academy Report noted the following points and conclusions in the Conning Study that are of greatest use for the present report:

1. "Credit data can be used to create scores that in fact provide additional predictive information about future [insured] losses." Actuaries-Academy Report, p. 11. "The Conning study concludes that the use of credit information has merit because it appears to have a correlation to loss ratio performance and does not appear to overlap other variables used by insurers." *Id.*, pp. 15 – 16.
2. Use of credit data has led to "rapid grown in automated underwriting systems that minimize subjective judgment by relying on more objective, rigorous, data-driven decision processes." *Id.*, p. 11.
3. Insurance companies appear to be focusing their use of credit data and insurance risk scoring on four strategic goals: *A*) better risk classifications to minimize subsidies between classes due to some classes paying lower premiums than the should; *B*) targeting their marketing towards potential customers indicated to be better insurance risks; *C*) more accurately pricing their policies to match the risk each policyholder represents; and *D*) keeping customers who are indicated as better insurance risks, by offering them discounts to renew their insurance. *Id.*, pp. 12 – 13.
4. Use of credit data in assessing an individual's likely insurance performance has become controversial for several reasons. One is that using credit history is often perceived to be in conflict with what society considers fair, particularly if the individual's score is affected by catastrophic events such as divorce, medical problems or loss of a job." *Id.*, p. 11. Another is that credit data might differ along ethnic or socioeconomic lines, leading in effect to discrimination on legally and socially unacceptable grounds — however, the empirical data in the five studies examined by Conning were mixed and inconclusive about this. *Id.*, p. 13. A third concern is errors in the credit histories themselves; Conning noted that the studies to date (as of the time of his report in 2001) did show significant data errors in credit reports, but added that those studies all failed to assess what the effects of those errors might be, if any. *Id.*, p. 14.

Noting that the Conning Study was a review of existing literature and not new original research, the Actuaries-Academy Report said, "In our opinion, the authors' findings are reasonable and provide a good overview of the issues." *Id.*, p. 15. The Academy Subcommittee called the Conning group "unbiased observers" who "conducted a thorough analysis of the available literature[.]" *Id.* The Academy Subcommittee criticized the Conning Report for "not distinguish[ing] between score-based or rule-based [underwriting] models in application."[27] *Id.*, p. 16. The Academy Subcommittee also suggested

[27] Explaining the difference between these two kinds of models, the Academy wrote: "The rule based model gives a set of conditions that result in either a credit or surcharge for each condition that is present or absent. A score-based model will provide an aggregate score resulting from all of the risk parameters but

additional issues that could have been addressed in the Conning Report. *Id.*

Regarding the 1999 study by Fair, Isaac ("Fair-Isaac") — "a prominent provider of insurance scoring models" (Actuaries-Academy Report, p. 21) — which presented results of using credit-history data and credit scoring models that Fair, Isaac developed, here are the conclusions in Fair-Isaac together with the Academy Subcommittee's commentary on them:

1. Fair-Isaac's Point: "The accuracy of the credit data should not be a matter of concern. … The fact that insurance scores are so predictive of insurance loss performance testifies to the overall accuracy of the credit information. Several studies … show very low error rates for credit data … much lower error rates than motor vehicle reports (MVRs), which are readily accepted and routinely used for auto insurance." Actuaries-Academy Report, p. 17.

   Academy Subcommittee's Commentary: "Fair, Isaacs states 'In view of the error rate of MVRs, the credit report error rate should not be an issue,' but this seems to bee too strong a conclusion." *Id.*, p. 19

2. Fair-Isaac's Point: "The Fair, Isaac study gives examples of five specific credit variables and how they are related to personal property and automobile insurance loss ratios. … The actual model scores also are very effective at predicting loss ratio relativities. … The general statistical techniques are well known but the exact models are proprietary." *Id.*, p. 18.

   Academy Subcommittee's Commentary: "The Fair, Issac [*sic*] study provides many results (statistical relationships), showing both individual credit characteristics and insurance scores are closely related to loss ratios. However, little or no in-depth data analysis is directly included in the paper, and therefore it is not possible to comment on the validity of these results." *Id.*, p. 20.

3. Fair-Isaac's Point: "Statistical models do not determine [i.e., demonstrate] causality." *Id.*, p. 18.

   Academy Subcommittee's Commentary: "It should not be necessary to demonstrate causality. Actuarial Standard of Practice No. 12 states that causality cannot be required for risk classification systems. … Risk classes should be neither obscure nor irrelevant, but they need not exhibit a cause-and-effect relationship." *Id.*, p. 20

4. Fair-Isaac's Point: "The Fair, Isaac scoring models are not unfairly discriminatory [and] avoid the use of … income, location, nationality, net worth, race, color, religion, and disability" as underwriting factors. *Id.*, p. 18

   Academy Subcommittee's Commentary: The "Subcommittee accepts Fair, Isaac's statement [but] there is no way for the subcommittee to verify this statement without reviewing Fair, Isaac's models. However, this statement

---

does not permit the recipient of the score to understand which items were the drivers of the score." Actuaries-Academy Report, p. 16.

[about not being unable to verify non-discrimination in Fair, Isaac's models] cannot be generalized to other models that are in use. … [T]he paper does not address the question of whether or not any of the credit variables used, or the overall insurance score, might be a surrogate or a proxy for any prohibited factor or factors." *Id.*, p. 21

5. Fair-Isaac's Point: "The use of Insurance Bureau scores (based on Fair, Isaac models) enables insurers to improve the speed, objectivity, and consistency of their underwriting." *Id.*, p. 18

   Academy Subcommittee's Commentary: "[N]o evidence is presented to indicate that insurers use the Insurance Bureau scores in an objective and consistent manner." *Id.*, p. 21

6. Fair-Isaac's Point: "Credit scores, unlike Insurance Bureau scores [based on credit-history data], were developed to predict credit risk and are not appropriate for the purpose of predicting insurance risk." *Id.*, p. 19

   Academy Subcommittee's Commentary: "Although this was not a major point in the Fair, Isaac study, the distinction between credit (lending) scores and insurance scores is important. The study does not present any information about the relationship between credit scores and insurance scores." *Id.*, p. 21

The Academy Subcommittee criticized Fair-Isaac for providing "little description of the underlying data analysis" and for not providing "any multivariate analysis, to determine if credit history might be essentially replacing another variable." *Id.*, p. 22.

Finally, regarding the 1999 study by the Virginia Bureau of Insurance the "Virginia Study"), here are Virginia's relevant conclusions as summarized by the Academy Subcommittee, together with the Subcommittee's comments on them:

1. Virginia Study's point: Some form of credit scoring was being done in conjunction with approximately 36% of new automobile insurance being written in Virginia and some 49% of new homeowners insurance. Actuaries-Academy Report, p. 23.

   Academy Subcommittee's Commentary: "The conclusion is probably a reasonable estimate of what the market is doing." *Id.*, p. 24.

2. Virginia Study's point: Roughly 30% of Virginia insurers using credit history data may decline new business solely on the basis credit history, and 1% may refuse to renew solely on that basis. *Id.*, p. 23.

   Academy Subcommittee's Commentary: "[T]here may be a bias in responding [to Virginia's survey]. For example, companies using credit scoring as a potentially sole criterion for the acceptance or rejection of a potential policyholder may have tended to decline to respond." *Id.*, p. 24.

3. Virginia Study's point: There is a statistical correlation between credit score and policy loss performance. *Id.*, p. 23.

   Academy Subcommittee's Commentary: "The conclusion was based on company filings [with Virginia's Bureau of Insurance] in which there was a proposal to use credit score as a factor in rating. The study includes no actual data, so it is not possible to comment on the quality of the supporting evidence. The fact that there were at least 50 survey respondents using credit history, who apparently submitted filings with appropriate support for the use of credit history, indicates that there is a correlation." *Id*., p. 25.

4. Virginia Study's point: Credit scoring is unlikely to lead to impermissible redlining "because income and race alone are not reliable predictors of credit score." *Id*., p. 23.

   Academy Subcommittee's Commentary: "The data is [*sic*] reviewed on an aggregate basis, by ZIP codes, and there is no attempt to match the credit scores of individual consumers with their income as race. As with item 3 the paper does not include any of the supporting data, so it is not possible to comment directly on the conclusion.." *Id*., p. 25.

5. Virginia Study's point: Consumer complaints in Virginia about using credit reports were very low in 1999, but could rise as more insurers use credit data and more would-be insureds are turned down or not renewed. *Id*., p. 23.

6. Virginia Study's point: 63% of the insurance agents responding to the Virginia survey favored a law to prohibit insurers from refusing to issue or renew policies due to adverse credit reports. *Id*., p. 23.

   Academy Subcommittee's Commentary: "It is not clear that the 63 percent is representative of agents in total. Because of this there may be a greater tendency for the strongly opinioned to respond to the survey. … [I]t is not known to what degree there was follow-up with the non-responding agents." *Id*., p. 26.

7. Virginia Study's point: "None of the credit variables used in the Fair, Isaac models appear to be unfairly discriminatory." *Id*., p. 24.

   Academy Subcommittee's Commentary: "The basis for this conclusion is not clear. There was at least one interview with representative of Fair, Isaac, and the study seems to contain the suggestion that the Bureau was allowed to see the actual list of credit variables used by Fair, Isaac. It is not possible to verify this conclusion, because its basis is unclear." *Id*., p. 26.

The Academy Subcommittee criticized the Virginia Study for "includ[ing] only a limited amount" of data, which made it "difficult for readers to assess the validity of the conclusions. The inclusion of some summarized data displaying the correlation between credit score and loss performance, and data supporting the ZIP code analysis, would have strengthened the study." *Id*., p. 27.

<u>*The University of Texas Study*</u>. This study, published just last month (March 2003), examined the statistical relationship between credit history and insurance losses for the automobile insurance industry in Texas. The University constructed a database of 175,647 automobile insurance policies written in the first quarter of 1998, including premium and loss data during the following 12 months. University of Texas Study, p. 2. The database was transferred to a third-party, Choicepoint, to find the credit histories for

the policies' named insureds. *Id*. Some 22,284 policyholders did not have sufficient or matchable information or credit history to allow a credit score to be made for them (*id*., p. 8), and 214 policies were deleted from the database[28] — leaving 153,149 policies that had been matched with credit histories of the policies' named insureds. *Id*. Choicepoint's credit data included a total of 445 credit variables as well as a "credit score" created by Choicepoint to reflect the strength of a person's credit history. *Id*., p. 3. Typically the credit scores under Choicepoint's system ranged between 200 and 1000. *Id*. The database was sorted by credit score and then grouped into 10 groups of equal size, or deciles, with the first decile being comprised of the named insureds with the worst credit scores.[29] *Id*. The average loss ratio relativity was then calculated for each decile.[30] *Id*.

The University of Texas Study made the following empirical findings:

1. There was a strong correlation between a low credit rating (i.e., low decile number) and high loss ratio relativity. The first decile had an average loss ratio 1.53 times the average ratio for all 10 deciles, and the second and third deciles' averages were 1.35 times and 1.14 times, respectively, the average for all 10 deciles. The correlation coefficient between a decile's average loss ratio relativity and its decile number (average credit score) was 0.95.
2. There was a strong correlation between a low credit score (decile number) and the likelihood that an insurance claim was made during the 12-month period.
3. There was a strong correlation between a low credit score (decile number) and the size of the claims being made, with larger claims being made on average as the decile number (and credit score) got lower.
4. For each of the first three correlations, a statistical analysis was performed to see how likely it was that the observed correlation was merely the result of a fluke in the data sampled. For each one, the analysis indicated less than one chance in 10,000 that the correlation was a statistical fluke.

Id., pp. 9 – 10.

The University of Texas Study acknowledges several limitations in its design and scope. First, it was not designed to, and did not examine the accuracy of the credit data that Choicepoint used to develop its credit scores. This leaves open the possibility that that some of the indicated correlations might be erroneous due to the number of errors in the data records. *Id*., p. 12. Second, the study used only Choicepoint's credit-scoring method, which may or may not reflect the credit-scoring used by individual insurance companies. Thus, the correlations observed in the study might not be as strong, or might

---

[28] Of these 157 were deleted for having impossible data — earned premiums of zero or less, incurred loss of less than zero, or zero or a negative number of automobiles. Another 57 were deleted because they had loss ratios of 100 or more, absurdly high although not quite an impossibility theoretically. *Id*., p. 8.

[29] The 22,284 policies without matched credit scores for their named insured became an 11th group, which was analyzed for loss ratio relativity the same as the other 10 groups of policies with matched credit scores. *Id*., p. 9, footnote 9.

[30] This "loss ratio relativity" has the same meaning as before — it is the average of the loss ratios for the decile, divided by the average of the loss ratios for all the policies.

not even exist, under an individual insurer's particular credit-scoring method. *Id*. Third, there was no attempt to determine whether more than one driver was insured under any given policy, and so it is possible that a named insured (e.g., the father of one or more teenage sons) may have a good credit score but a high loss ratio because of other drivers covered by the policy. However, this factor would tend to loosen the correlations being observed, rather than strengthen them — and so it is quite possible that these correlations are even stronger than the study indicates. *Id*. Finally, the database did not contain information about the income, ethnicity or physical residence of the named insureds, and so the study was not designed to, and did not examine whether Choicepoint's credit-scoring system might be unintentionally discriminating against individuals along one or more of these dubious or forbidden lines. *Id*.

PART 4. CONCLUSIONS AND RECOMMENDATION.

<u>*Conclusions*</u>. There are three major issues relating to insurance companies' use of credit-history data to evaluate the insurance risk posed by an individual person. One is the relevance of that data to the insurance risk which that person actually poses. This includes both the statistical correlation between the two as well as the reliability of the credit data upon which the statistical correlations are made. The second is the fairness of using credit data when a person has recently experienced a major financial crisis beyond the person's ordinary control, such as a major illness in the family, the loss of a breadwinner's job, the collapse of the Nasdaq, a divorce and costly property settlement, or some similar extraordinary event. The third is whether the particular credit data that an insurance company uses in evaluating a person's insurance risk amounts to a form of forbidden discrimination along ethnic or socioeconomic lines.

This paper began with a review of the principles of how insurance works, and of the crucial importance in the insurance business of being able to assess accurately the risks to be insured against. The reports and studies discussed in Part 3 demonstrate that there is a very strong statistical correlation between an insured's credit data (or the particular credit-scoring methods that were studies) and the insurance risk that s/he represents. The chance that the strong correlations being observed are due to statistical flukes is too tiny to be plausible. The necessary conclusion to be drawn from the empirical evidence is that poor quality credit data <u>*are*</u> somehow linked to a person's insurance risk, whether as a homeowner or the driver of an automobile.

With respect to errors in the credit data, it is not possible from the studies available to determine how much error there is in the credit records of the three major credit bureaus (TransUnion, Experian and Equifax). However, it can be noted that the very strong correlations being observed between credit factors and insurance risk make it unlikely that the credit data errors are very significant statistically. This is because the data errors should be random, improving someone's credit record as often as they make it worse. In other words, unless some kind of systematic bias or skewedness is observed in the credit data errors, they should cancel each other out for the population as a whole.

This large-scale canceling-out may not be of particular comfort to the person whose credit rating has suffered as a result of errors in her/his own data. However, there are two responses to this concern. One is that the benefits from more accurate assessment of people's insurance risks will flow a large majority of the population with good or excellent credit ratings, and hence these benefits far outweigh the must less frequent injustice that would flow from credit data errors in individual cases. The other is that there is a better way to fix the problem of errors in individuals' credit histories. It can be fully taken care of by making insurers (if they don't already do so on their own) inform the person about the factors in her/his credit history that are affecting her/his insurance rates, so s/he has a reasonable opportunity with the insurer to correct the error and be charged the correct rate for that insurance.[31] It is, as explained above, as much in the insurers' interest as the insured's to make sure that the actual insurance risk posed by that insured is as accurately assessed as possible.

The second issue raised by insurers' use of credit data is the fairness of using that data when someone has recently experienced a major setback of some kind that was beyond their reasonable control. Once again, there is a better alternative to solve this problem instead of forbidding insurers from using credit data as an underwriting tool. This alternative is to require insurers to place less weight, or no weight at all, on the credit data for these people until some reasonable time has passed for them to get back on their feet.[32]

The third issue is possibility that the use of credit data might discriminate against people on ethnic and/or socioeconomic grounds or in some other socially unacceptable fashion. Here, too, this possibility can be addressed and remedied as necessary by a means other than forbidding insurers from using such a statistically important tool. State insurance regulators simply must monitor how insurance companies are using the credit data, and must require insurers to develop and maintain the necessary additional data that will show whether or not any such forbidden or insidious discrimination is occurring. The Actuaries-Academy Report, pp. 28 – 38, has already outlined how this should be done,[33] and it is simply a matter of following that advice.

The final conclusion to be drawn from all of this is that it would be a mistake to pass legislation forbidding insurance companies in Alaska from using such an important, useful and valuable tool as individual credit-history data in assessing the insurance risks posed by individuals. Most Alaskans have good credit histories, and they will stand to

---

[31] Having insurers conduct this discussions with their insureds will have the further benefit of helping the latter identify what defects in their credit histories they should cure, so they can start correcting their situation on their own.

[32] The details about how little weight, if any, to give to credit data during such a recovery period, and how long such a recovery period should be, are beyond the scope of this report, and in any event, would be best left to those who regulate the insurance industry because of their knowledge and experience in these matters.

[33] The Actuaries-Academy Report is addressed to the National Association of Insurance Commissioners, but its recommendations could certainly be acted upon by any individual state insurance commissioner/regulator. Whether Alaska's own Division of Insurance should act on its own or in concert with its counterparts in other states, is a matter beyond the scope of this report and one which the SALGA Committee would leave, at least in the first instance, to that Division's best discretion.

benefit if insurers can use this underwriting tool. Any problems that may result from using this tool can be better taken care of by other means, instead of banning the use of credit data altogether.

Recommendation. The SALGA Committee recommends that the Board of Directors of the Anchorage Chamber of Commerce, opposing any legislation to ban entirely the use of credit data by insurers in evaluating an individual's insurance risk:

Board of Directors, Anchorage Chamber of Commerce
Resolution 2002/03-__
Opposing Legislation to Ban the Use of Credit-History Data
By Insurers in Assessing the Insurance Risks Posed by Their Insureds

WHEREAS, legislation has been introduced in the state legislature that would ban the use of an individual's credit-history data by an insurer in assessing the insurance risk posed by that person, and

WHEREAS, the essence of insurance is the accurate quantification and management of the risks to be insured, and if this cannot be done as accurately as possible, some insured policyholders will necessarily end up paying more than they properly should while others will pay less than they should; and

WHEREAS, empirical studies have repeatedly found extremely strong statistical correlations between an individual's credit-history data and the insurance risk which s/he represents; these correlations are independent of and in addition to the other factors that insurers use to evaluate people insurance risk; and the chances that the correlations being observed are the result of statistical flukes in the data are overwhelmingly slim, such as 1 in 10,000 or less; and

WHEREAS, most Alaskans have good or excellent credit histories and thus would stand to benefit from lower insurance rates if those with bad credit histories and associated greater insurance risks would begin to pay what they properly should be paying to cover the insurance risks that they actually represent; and

WHEREAS, the Division of Insurance already can and surely will monitor the insurance industry's use of credit-history data to prevent it from somehow discriminating against groups of Alaskans along ethnic, socioeconomic or other unacceptable or inappropriate lines; and

WHEREAS, the Division of Insurance can and should require insurers in Alaska to inform their policyholders and applicants for insurance of the particular factors in their credit histories that adversely affect their insurance rates, so that those people will have a reasonable opportunity to correct any errors in their credit-history data and have their insurance rates adjusted to reflect such corrections;

NOW, THEREFORE, BE IT RESOLVED by the Board of Directors of the Anchorage Chamber of Commerce that the Board opposes the enactment any legislation that would ban insurers in Alaska from using an individual's credit-history data in evaluating the insurance risk posed by that person;

AND FURTHER RESOLVED that the Anchorage Chamber will inform its members of this resolution and will post it, together with the report and recommendations of the State and Local Government Affairs Committee on these matters, on the Anchorage Chamber's website;

AND FURTHER RESOLVED that the Anchorage Chamber will issue a press release and public announcement of this Resolution, and send copies of it to each state Senator and Representative, to the Governor, to other chambers of commerce in Alaska, and to the press.

DATED __________________, 2003