Statistical Analysis and Recommendations for Further Research

Regarding 2018 State of Washington Office of Insurance Commissioner Study of Alleged Insurer Steering

I. Background:

This analysis is in response to the letter from Candice Myrum of the State of Washington Office of Insurance Commissioner dated June 22nd, 2018 and its accompanying survey results regarding an inquiry into the practice of insurance "steering" of claimants to the insurers' preferred or "direct repair program" (DRP) body shops.

I am providing this analysis to the Insurance Commissioner because I share the concerns of Rep. Kirby and his constituent, and because my professional background puts me in a unique position to understand the survey results. I spent several years working in the auto body industry, and my experience in that industry has given me an intimate knowledge of both how insurers communicate, and how claimants experience that communication. I also studied quantitative science at the University of Washington, which gave me an understanding of – and a passion for – statistics. It is especially my statistical knowledge that prompted me to dig deeper into the findings provided by the Insurance Commissioner's study.

In matters of great importance, such as upholding consumer protection, I strongly advocate a rigorous scientific approach to research and problem solving. I am sure this is also the intent of the Insurance Commissioner; however, my analysis of the currently available findings concludes that the 2018 study was not designed in such a way that statistically significant results could truly be obtained. In short, based on my knowledge of statistics, **I do not believe the current study is sufficient, and for the sake of consumer protection, better science must be utilized to properly investigate the matter of alleged steering.**

For the remainder of this analysis, I will provide constructive criticism of the survey design used, and I will provide suggestions for further and more rigorous research based on my statistical expertise. **I ask that the Office of Insurance Commissioner consider these suggestions, as well as those of other qualified statisticians, and conduct further research in support of this investigation.**

II. Basics of Statistical Analysis

Possible outcomes and the null hypothesis:

For many statistical tests (including the ones I will propose here) there are typically two and only two mutually exclusive outcomes: reject the null hypothesis, or fail to reject the null hypothesis. The null hypothesis typically represents the "status quo" or a finding that the factor of interest has no significant bearing on the situation being studied. The null hypothesis is typically denoted as "H_0", and the opposite possibility of "alternate hypothesis" as "H_a". If the null hypothesis is rejected, then by process of elimination, the alternate hypothesis should be adopted in its place.

In this study, for example, the broad overall outcomes are:

H_0: Auto insurers in Washington State are not engaging in steering

H_a: Auto insurers in Washington State are engaging in steering

(Note that this overall hypothesis cannot be tested with a single metric. This example shows only the broad question the study is seeking to answer. Proper analysis will include several statistical tests with their own, much narrower hypotheses.)

Confidence levels and sample size:

It is important to note a key difference between *evidence* and *proof.* There are varying degrees of "confidence levels" in statistical outcomes, but 100% certainty is all but impossible to obtain. It is critical to know what confidence level a statistician is using when presenting results. There is a great deal of difference between being 99% sure of rejecting a null hypothesis and being 80% sure of rejecting a null hypothesis!

Naturally, when researching important questions, higher confidence is better. Higher confidence is closely related to the *sample size* of the study. If you ask 10 people their favorite color and 8 of them say orange, would you conclude that 80% of the general population's favorite color is orange? No. You would suspect that something was biased in your study, and you would ask more people in hopes of getting closer to the true proportion. Suppose instead you asked 100,000 people their favorite color, and determined that 600 of them selected orange. Assuming your survey was unbiased and statistically sound, you would be much more confident in suggesting that about 6% of people's favorite color is orange. That is a massive difference from the 80% found in the flawed, too-small study of 10 people.

Study design:

Statistical results are only as valid as the study design that produced them. Good statistics rely on good data. When sampling a population, a basic tenet of statistics demands that the sample be random and unbiased. Biases can occur for many reasons, both intentional and accidental. Consider some examples:

- Suppose you want to calculate the average height of students at your child's school. You ask your child to ask 25 people their height. Your child happens to be on the basketball team, and out of convenience, they ask all their teammates. This sample is *highly* biased, because above-average height is common in basketball players! You might incorrectly conclude based on this study that the average height of an 8^{th} grader is 6'2".

- If your study involves complex questions, the way the questions are written may lead a respondent to answer in one way or another. Suppose you want to gauge public opinion on a political candidate. You might ask respondents to select from "very favorable," "somewhat favorable," "neutral," "somewhat unfavorable," or "very unfavorable." If your question is simply "What is your opinion on John Doe?" you may get different results than if you were to ask a leading question like "What is your opinion on local menace and master of tax evasion John Doe?" or "What is your opinion on family man and charitable donor John Doe?"

Correlation vs. Causation:

Another key distinction in statistics is the difference between correlation (factors appearing to be related) and causation (factors having an actual cause-and-effect relationship). To illustrate, I will use a classic example shared by a favorite teacher of mine: suppose you compare the number of ice cream cones purchased per day in a small town to the number of sunburns reported per day in that same small town. After months of data collections, you find that the two are closely *correlated* – days with high ice cream sales are also days with high rates of sunburn. If you assumed that correlation and causation are the same, you might conclude that eating ice cream causes sunburn! This is obviously flawed logic, and any rational person might suggest that *hot, sunny days* are the real factor driving both ice cream sales and sunburn rates. You might pull weather data from your study period and find that indeed, the hotter

the weather, the more ice cream is sold, and the more people report sunburns. This makes more sense as a *causation* relationship.

While this is a simplified example, and it doesn't take a seasoned statistician to point out the absurdity of ice cream consumption causing a sunburn, many studies of less understood subject matters might not be so obvious. It is important, therefore, to resist the temptation to conclude that A causes B just because the two are related. Further research into the relationship between factors is required before drawing such a conclusion.

Transparency and Replicability:

As with any science, statistics must be transparently explained so that any other statistician could follow the logic from start to finish, and repeat the study themselves and obtain the same or reasonably similar results. This means that any report of results should include a detailed description of the methods used to obtain them, and preferably the raw data used to make the analysis. For example, the following would add value to the rigor of a statistical report:

- Copies of the survey given to respondents along with any introductory text
- A description of how the sample was selected (e.g. how it was decided who received a survey)
- A discussion of why one analysis method was selected over another
- "Showing your work", or including steps showing how any end statistics were calculated
- Tables of raw data collected
- Disclosure of individuals or organizations financing the study

The above introduction is just the tip of the iceberg of what makes good statistics, but should provide the necessary context for my analysis of the 2018 study. More information can be found in any introductory statistics textbook.

III. Analysis of Insurance Commissioner Study and Recommendations for Improvement

A. Consumer Repair Shop Survey Results

In the cover letter, Ms. Myrum states that "a clear majority" of claimants were repaired by non-DRP repair shops, and that they "infer" consumers are aware of their rights. To support this conclusion, a proper statistical test would need to be performed. First, what is a "clear majority"? 51%? 60%? 75%? This term is too subjective and should be replaced with hard numbers.

To statistically analyze this matter, the question at hand is essentially "how many claimants choose a DRP shop, how many choose a non-DRP shop, and is there a significant difference?" First, for simplicity, limit the data to claimants who chose to have their vehicles repaired. The choice *not* to repair is worth its own study and is beyond the scope of this question.

Next, formulate the null and alternate hypothesis being tested:

H_0: The percentage of claimants choosing a DRP is equal to the percentage of claimants choosing a non-DRP shop.

H_a: The percentage of claimants choosing a DRP is NOT equal to the percentage of claimants choosing a non-DRP shop.

A t-test for significance could be used to test this hypothesis. (For more information about a t-test, consult an introductory statistics textbook.) The result will tell whether any observed differences in percentage are significant. For example (assuming a sufficient sample size), if the percentages are 52% and 48% respectively, the t-test will show high odds of obtaining such a result by chance (i.e. there is no

significant difference, and you would fail to reject the null hypothesis). On the other hand, if the percentages were 80% and 20% of a sufficiently large sample, the t-test will show infinitesimally small odds of obtaining such a result by chance. In this case, you would reject the null hypothesis and conclude that the two percentages are truly different; the percentage of claimants choosing a DRP is not equal to the percentage of claimants choosing a non-DRP shop.

Recall that sample size is very important in obtaining rigorous data. In the case of the Insurance Commissioner study (specifically Question 2 in the "Consumer Repair Shop Survey Results" section), there were 28 respondents. The question posed also did not truly ask the question "did an insurer influence your decision?" OR "did you choose an insurance network shop or a non-insurance network shop?" It does ask the respondents *how* they chose, but the responses still leave other questions unanswered. The fact that the most common answer was "none of the above/other" makes it difficult to draw any conclusions without more information. Fifteen respondents – over half – chose this answer. Analyzing differences in percentage among three remaining answers with only a total of thirteen respondents is almost meaningless. **A much larger sample size is required to reliably analyze the state of the industry.**

Furthermore, the "Question 2" as written will not address any incidences of steering that were ultimately unsuccessful. If steering includes *attempts* by an insurer to persuade a claimant to use a DRP, then only the point-blank question "did your insurer attempt to persuade you to use their DRP?" will uncover these incidents. If a claimant, for whatever reason, had a shop of their own in mind and was unpersuadable in choosing another, their answer of "I chose my own shop" will hide the fact that *steering still did occur* in their claim. Regardless of a claimant's ultimate choice, if the goal is to discern whether or not steering is occurring, all instances of *attempted* persuasion must be included in this analysis.

Considering the insufficient sample size and the nature of the questions that fail to capture all incidents of steering, it is premature to infer that consumers are aware of their rights. Though some consumers may ultimately choose their own shop, they may do so independently of an insurer informing them of their rights, or even in spite of attempted steering. This is a classic example of confusing correlation and causation. **The data collected do not adequately prove that steering is not occurring or that consumers are aware of their rights; the data show only (and weakly at that) that some consumers ultimately choose a non-DRP shop. A better designed survey with more thoughtfully worded questions *and a much larger sample size* is required.**

Regarding the analysis of labor rates, the rate is only one of many economic factors that make DRPs beneficial to insurance carriers. Others include contractual obligations to use junkyard-salvaged, aftermarket, or otherwise inferior parts; pressure to unrealistically reduce repair times, and more. Insurers also dictate labor rates to their DRPs. Furthermore, with respect to rates, many non-DRP shops will voluntarily match their labor rates to those of an insurer's shop in order to compete or avoid conflict. In my personal experience as an estimator at a non-DRP shop, insurance adjusters would routinely pressure me to reduce our rates to match theirs, threatening to refuse to pay the claim in full and suggest that my customer would have to pay the remainder out of pocket – or break their repair contract with me or move their vehicle to another shop. The discussion and analysis of labor rates is a complex issue that deserves an essay all its own, and for brevity I will not get further into it here, but I will say that **in my time at a non-DRP shop, <u>I often lost customers</u> – even some who had already signed a contract with me – due to insurers refusing to pay our rates and forcing my customers to choose between out-of-pocket costs and moving to another shop. <u>If that is not steering, I don't know what is.</u>**

The question of satisfaction with the repairs runs into the sample size problem as well. First, the OIC analysis assumes that 14 consumers chose a shop "without influence from the insurer," which (1) cannot be proven, and (2) does not specify whether or not that shop was a DRP. Therefore, a proper

analysis of whether or not DRPs perform satisfactory repairs cannot be performed with the current data. To do this, one would need a much larger sample size. Then, the percentages of satisfactory repairs should be compared using a t-test as above.

Lastly, the whole survey may be biased due to a convenience sample. To analyze the steering issue statewide, the survey should be given to **truly random claimants.** The use of a subset of claimants who had already filed a complaint makes the results highly at risk for bias. It is unknown from the results what any of these claimants had complained about, or what other dissatisfaction – whether with a shop or an insurer – may have influenced their responses. **It is unknown whether this complaint-filing subset of people is more or less likely to report steering, and with no way to correct for this unknown factor, none of the results are valid. A large, random sample must be used instead.**

B. Private Passenger Claims Processing Data Survey

First, it seems immediately obvious that asking only the insurers – i.e. the group allegedly engaging in an unlawful practice – for their side of the story seems incredibly biased. Although one should assume the responses are truthful based on a presumption of innocence, it cannot be overlooked that these organizations will be highly motivated to paint themselves in the best light, and it is unlikely they will admit to any practice that could be considered unlawful steering. To get the whole story, claimants must be surveyed about their experiences as well. Naturally, they will not have the inside information that insurers could provide, but they *will* be able to share their experience and show a clear picture of how communication was received. Unlike insurers, they will have nothing to lose by recounting incidences of steering.

This survey does use a better sample size (174 respondents), but the limited analysis of the data does not support the conclusion that steering is not occurring.

First, the results of how consumers are informed of their right to choose their own shop is very concerning. Over half of insurers give only a verbal notification (nearly impossible to prove) and almost a quarter do not inform consumers at all. In a situation where many consumers have a limited understanding of insurance, not informing them can be considering lying by omission, especially if communication is worded in a way that is misleading. For example, an insurer might say, "Well, XYZ shop isn't in our network," and without being informed of their rights, a consumer might assume that they cannot take their vehicle to that shop. Other insurers might say, "We have DRP shops at location A and location B, which is more convenient for you?" and without knowing their rights, a consumer might assume those are their only options.

Only 16.15% of insurers responded that consumers were informed in writing. In my opinion, this is the "gold standard" of providing information, and should be a legal requirement. Consistent language should be used. Verbal or non-existent notifications are impossible to trace and are too vulnerable to being abused or used in a misleading way.

The analysis of impacts on consumers for choosing a non-preferred repair facility are also concerning. Yes, the data show a large percentage of insurers responding "no impact", but when the other possible impacts are combined, that is over 30%. Even though the graph may make these results look small, the influence of these impacts on a consumer may be *huge*. It is interesting that one of the most commonly-reported impacts I have observed is not on this list at all: out-of-pocket costs for the repair. When I was an estimator, this and timeliness were the two biggest reasons I lost customers to DRP shops. Under the stress of a car accident, the last thing a consumer wants is more headaches, and many will gladly chose the path of least resistance. In this case, that's often a DRP, where the threats of these impacts are not an issue.

To properly acknowledge the actual *influence* of these impacts, the consumer side must be surveyed. Below is one way the question could be posed:

- If you selected a DRP, what factors influenced your decision?

- Location/convenience
- I didn't know I could choose a non-DRP
- Insurer suggested I might have out-of-pocket costs, lengthy repair time, etc.
- Good reviews of the shop (online, word-of-mouth, etc.)

To include analysis of unsuccessful steering, consumers should also be asked a series of questions regarding what they were or were not told. This data could then be compared to the customer's ultimate choice. The ultimate choice should also be analyzed with respect to how the customer was (or wasn't) informed of their rights.

Lastly, the final page of the OIC analysis states, "The results of this data survey do not support the assertion that insurance carriers are steering consumers to repair facility that are part of the carriers' preferred service center networks or that they are limiting or denying payment on claims when consumers choose a non-preferred repair facility." This is not exactly false, but not all the right questions were asked to truly support *or* disprove allegations of steering. Many more questions need to be asked, of both insurers and consumers, and the results need to be interpreted in context. Assuming that 60% of insurers reporting "no impact" on consumers if they choose a non-DRP shop *does not* prove that steering does not occur. Assuming that a marginal difference in labor rates is the only incentive an insurer has to steer consumers to their shops fails to consider the full picture and the nature of a typical DRP contract. In sum, too many assumptions are made and too many questions are not asked to confidently state whether steering is or is not occurring, and what impact that may have on consumers.

IV: Summary

The surveys conducted by the OIC in 2018 are not statistically rigorous enough to confidently support the conclusions reached. In order to properly investigate the matter of steering, additional research is necessary. A better survey would include the following features:

- Much larger sample size – preferably well over 100 considering the volume of claims in Washington State
- Sample that is randomly selected using an approved statistical method, and not selected out of convenience
- Statistical tests for significance of results (such as a t-test, chi-square test, or other, depending on the type of data being analyzed)
- Careful peer review of survey questions being asked in order to ensure:
  - Answers will provide data that will help support or invalidate allegations of steering
  - Questions are not "leading", or likely to influence the respondent with their wording
- Careful consideration of correlation vs. causation when interpreting results. **Avoid making assumptions about the results that have not themselves been statistically tested.**

I am pleased that the Washington State Office of the Insurance Commissioner has taken steps to begin investigating this important consumer protection matter, and I look forward to future, rigorous research that can be statistically tested. Thank you in advance for your diligent efforts. Should you have any questions, my contact information is included at the end of this analysis.

V: Acknowledgements and References

The following materials were consulted in the preparation of this analysis:

- Biostatistical Analysis, 4th ed., by Jerrold H. Zar
- Applied Linear Regression Models, 4th ed., by Michael H. Kutner, Christopher J. Nachtsheim, and John Neter
- www.quantitativeskills.com
- www.spss-tutorials.com

Though not consulted for this analysis, thanks are due to the following people:

- Amy Crum, my high school trigonometry and statistics teacher at Cascade High School in Everett who inspired my life-long love of statistics
- Prof. Loveday Conquest at the University of Washington, who taught me great depths of applications of statistics in the real world and introduced me to the term "going stats happy," as we are apt to do.

VI. Contact Information & Credentials

Lauren Wolfs
[redacted]

[redacted]

B.S. in Environmental Science & Resource Management
Minor in Quantitative Science
University of Washington, 2011

4 years' customer-facing experience in auto body industry, 2014-2018