Appendix 3: Research Methodology

This appendix reviews the major elements of the research methodology underlying the findings, observations, and recommendations regarding the performance audit of the claims management function of the Washington workers' compensation system. It proceeds in the following order:

1. Stakeholder and staff interviews
2. Documentation research and review
3. Review of claim files
4. Customer opinion survey
5. Best practices survey of panel of claims management experts
6. Data analysis of L&I claims data
7. Comparative data analysis of data from other jurisdictions

As we will note, these research tasks were interconnected and supported each other.

1 Stakeholder and Staff Interviews

The overall purpose of stakeholder interviews was to gain insights about the workings of the L&I claims process. We did not go into the interviews looking for problems or with preconceived notions about a reform agenda. Rather, we were looking for a balanced and objective assessment about the performance of the Washington system in general and suggestions about where the system could be improved. Where concerns or successes were indicated, we sought specific examples. A final motive for these interviews was to prepare for the claim file review and the survey of employers and injured workers, and to be alert to trends and patterns in the electronic data.

By design, the targets for our interviews were those stakeholders who have contact frequently with L&I through various phases and conditions of the claim process. They have much valuable information about how the process is working to advance their particular constituency's needs. Not surprisingly, the stakeholders contacted had different views of L&I because their underlying vested interests and range of experiences are different. For example, a union representative is likely to hear about claims problems from members, rather than observe the vast majority of claims that are processed without friction. As another example, group Retro managers can be expected to defend the concept of Retro premium refunds against the criticism of organized labor.[1] These differing perspectives were why we interviewed a representative and balanced sample of experts, and remained aware of their potential biases.

In the process of documenting interviews, we generally included the following details:

- Date ranges for all the interviews
- Parties interviewed and titles and relevant job responsibilities
- Contact information for interviewees
- Where the interviews took place (phone or physical location)
- Approximate duration of the contact

[1] The Washington Labor Council has criticized several aspects of the retrospective rating program, particularly the uses of premium refunds by group managers; see their position on retrospective rating found March 14, 2014 at: http://www.wslc.org/legis/workcomp.htm.

In addition to this standard background we documented the responses to specific questions about the aspect of the claims process that are most familiar to the interviewed groups. For stakeholders outside of L&I, the questions for each group were premeditated to follow a pattern. However, some degree of customization of the list was necessary to follow the flow an interesting discussion, or to pick up on points heard in previous interviews.

1.1 L&I PERSONNEL

We completed extensive interviews with all the key staff in claims related functions of L&I. We also interviewed staff related to Human Resources and training. The process of contacting L&I staff was rather formal at first. All contacts were arranged through Rachel Aarts. During interviews, an additional staff member sat in the interview to record the conversation. Later in the process during subsequent contacts, the interviews become much less formal. By agreement, we simply copied Ms. Aarts in the question and answer process following initial contacts. Throughout the process Ms. Aarts was extremely attentive to our needs and consistently followed up on requests.

The process began with interviews of all the section managers within the Division of Insurance Services. The initial "kick off" meeting at L&I took place in December, with many high-level managers present, along with Joel Sacks, Director of the Department of Labor and Industries. JLARC staff attended the meeting as well. Additionally, we scheduled an interview with Vickie Kennedy early in the process. It was a general "get acquainted" meeting without significant substantive discussion. In late February, near the end of our interviews, we scheduled another meeting with Ms. Kennedy, which was much more substantive than the first. We covered the management initiatives launched in 2013, with a focus on the Return to Work Program. We also discussed the seemingly controversial topic of "side-bar" agreements to resolve claim issues. We had a follow up meeting with Director Sacks and L&I management in June 2014 to discuss progress and early discoveries. We did detailed interviews of staff in two "waves" of approximately 40 people total. We also conducted numerous follow up phone and in person interviews as needed with staff. We submitted written interview questions (approximately 100) seeking clarification and documentation of certain processes and procedures. The team received extensive training from L&I staff on the LINIIS and ORION claims systems and demos of the FileFast Early Claims Solution system and the SI SIEDRS and SICAM systems. We also worked with L&I Retro staff to conduct a scenario case study on rating and refund methods. In the course of our follow up with staff we obtained numerous reports and metrics used internally by L&I; these proved to be invaluable sources for the report. In November 2014 we interviewed Retro program staff to discuss detailed scenario modeling needed to research the premium and refund process. In March 2015, we conducted follow-up interviews with several staff members, including claim unit supervisors and several members of the management staff, concerning additional topics identified for follow up. We also conducted a follow-up discussion with Ms. Kennedy to provide an update concerning the audit.

1.2 BOARD OF INDUSTRIAL INSURANCE APPEALS (BIIA).

Interviews were conducted with the Board of Industrial Insurance Appeals (BIIA). Among other administrative law responsibilities, BIIA handles appeals to claim decisions by L&I and self-insured employers. We interviewed seven staff members, and all three Commissioners. We also interacted with staff numerous times on data questions.

1.3 RETROSPECTIVE ACCOUNT EMPLOYERS.

We interviewed public members of the Retro Advisory Committee, as well as the chair Tim Smolen from L&I. The interviews of the public members solicited both their individual perspectives on Retro, and also about the role of the Advisory Committee and the issues it has been addressing. All the public members of the Advisory Committee were associated with group plans, but the questions were mainly about how well the system serves retrospectively rated employers in general. We also interviewed three Retro employers.

1.4 ACCOUNT PLAN MANAGERS.

We interviewed three L&I policyholder service specialists who work with policyholders to answer questions about employer accounts, and sometimes help educate them on how they can lower their workers' compensation costs. The Account Managers are the primary contact for employers regarding their workers compensation account (including claim free discount questions).

1.5 ADMINISTRATORS FOR RETROSPECTIVE GROUP ACCOUNTS

We interviewed three group managers involved in administering Retro-rated groups. Group retro-rated insurance is in principle open to any employer in the state. The underwriting standards for group membership and plan design are left to the control of the group management. The interviewees were selected to include a range of groups by size and industry homogeneity. We also wanted to interview groups with both high-end service levels and groups with basic member services. In addition, two other group managers (one from a very large and one a small Retro plan) were interviewed. Their perspectives about Retro overall were similar to the members of the Advisory Committee. However, we found a diversity of organizational structures, rules, and management style among the Retro groups. The interview findings were reinforced by browsing the websites for most of the group programs.

1.6 RETROSPECTIVE GROUP EMPLOYERS

We interviewed a limited number of employers within the Retro groups. From other interviews with group managers, it appears that most group members give a great amount of deference to group managers on claims handling. The bottom line for the group members is cost. The group managers know that if they cannot consistently show premium refunds, and distribute them fairly, their group membership will decline.

1.7 NON-RETROSPECTIVE STATE FUND EMPLOYERS

We interviewed four state-fund insured, non-Retro employers. We interviewed both eastern and western Washington employers. This is probably the most diverse of the interest groups, and the most difficult to generalize about. They range in size and degree of injury hazard. We were told that approximately 80% of all fund employers had not had a LT claim in three years. Thus, the vast majority of fund employers have little knowledge of the claims process and little or no interaction with the claims staff. For this reason we need to be careful about over generalizing these three or even double that number. One SF employer hired an employer representative, to assist with workers' compensation issues. We discovered that some SF employers also hired Third Party Administrators to assist in managing their claims.

1.8 Self-Insured Employers.

We interviewed three individual self-insured employers. The self-insurance community in Washington employs about 30% of the Washington workforce. It is relatively diverse compared to other states. There are cases of small Washington only employers that probably would be considered too small for self-insurance in other states. There are a large number of public employers and health care organizations that are self-insured. In selecting the employers to interview we thought it desirable to begin with the Executive Director of the Washington Self-Insured Association (WSIA). In discussion with him and in consideration of the employers on the Board and Executive Committee of WSIA we conducted a formal interview of the WSIA president. We also had less formal conversations with other relevant individuals and recorded their feedback. We also attended a meeting of the Workers' Compensation Advisory Committee, which has self-insured members on its roster, and discussed workers' compensation issues with attendees and documented results. We had the opportunity to attend a WSIA meeting in Gig Harbor, WA. During that meeting we informally interviewed several self-insured employers and defense attorneys. There was turnover in the WSIA Executive Director position in 2014, and we conducted an interview with the new Director.

1.9 Educational Service Districts (ESDs)

ESDs are essentially "group self insurance," and operate like a self-insured employer. Hospitals are allowed to do the same. The audit team met with a group of ESD administrative personnel, and discussed the audit project and received general feedback. A second meeting with a focus group of ESD claims subject matter experts also was conducted, to receive more specific feedback on L&I claims management performance.

1.10 Union representatives

We spoke to a wide variety of labor leaders and conducted five interviews. These included the director of Project Help, an ombuds like service project, staffed through a bid process overseen by L&I, and currently managed by the WA State Labor Council. We also spoke to several staff and business managers at a Seattle union hall.

1.11 Workers' compensation bar

We formally interviewed three members of the bar, and attended a WSIA meeting and conducted several informal attorney interviews at the meeting. The interviews covered both worker and employer attorneys.

1.12 Third Party Administrators (TPA)

We interviewed five representative WA TPAs. It was clear to us that Third Party Administrators played a very important role in the claims process, not only for self-insured employers, but also for group and individual Retro employers. The people interviewed all had 12+ years of experience handling claims, most of this time in Washington, but they also offered some interesting comparisons with their experiences in Oregon. Their reaction to working with L&I had some common features, but a number of divergences as well.

1.13 Office of the Self-Insured Ombuds

We interviewed the long-serving ombuds appointed by the Governor to head the Office of the Ombuds for Injured Workers of Self-Insured Businesses. This interview helped identify documentation that provided insight into the SI claims function.

1.14 Non-Washington interviews

In the course of the performance audit, several state workers' compensation individuals not formally connected with Washington workers' compensation were interviewed, to gain insight into their respective systems. These included management from several states, including Ohio, Idaho, and Oregon. In addition, the audit team members themselves had in-depth working knowledge of the workers' compensation systems in several states, including British Columbia, Saskatchewan, Virginia, Wisconsin, California, and Tennessee, to name a few jurisdictions.

Our starting point in this learning process was to learn as much as necessary about the rules, procedures, and culture of L&I to complete this project. The L&I claims staff interviews were indispensable in the design of the file review methodology. In addition, these interviews cast light on some of the fundamental research questions in this engagement. We learned about operating procedures that showed differences in the consistency of treatment of various employers and injured workers. Additionally, we obtained valuable insights from stakeholders to the Washington system regarding the functions of L&I. We found a general level of harmony and respect of stakeholders toward L&I staff. We did discover concerns from stakeholders about certain L&I processes, e.g. some TPAs and employers expressed dis-satisfaction with the L&I Self-Insurance claim review process. Finally, we obtained comparative information needed to establish benchmarks and standards used in workers' compensation systems, to evaluate the Washington system.

2 Document Research and Review

A fundamental research methodology utilized throughout the performance audit involved review and research into existing documentation. During interviews we were provided documentation and information concerning L&I performance and other relevant subjects; this included, among other things, references to statutes, regulations, and policies. This was particularly true with respect to interviews of L&I personnel. Much information concerning Washington workers' compensation is publicly available, not only directly from L&I and the BIIA, but also from various stakeholders involved throughout Washington workers' compensation claims management process.

The audit team also was given access to the L&I information systems (LINIIS and ORION), as well as the "intranet" or web-based information system provided to L&I personnel. This internal network included access to reference material involved in claims management.

The audit team had frequent phone and email exchanges with L&I personnel, including numerous written follow up questions directed to L&I staff, which was a source of additional documentation and reference information.

3 REVIEW OF CLAIM FILES

3.1 PROCESS

The goals for the claim review were:

- ensuring that the project team is collecting the data that are needed to supplement the electronic data in order to have the data elements needed to answer the research questions (some questions could not be answered by claim data);
- ensuring that the project team reviews enough files and the right mix of files to answer the research questions that involve comparisons between self-insured, Retro and non-Retro employers; and
- ensuring that the project team reviews enough files so the results are credible.

The audit team consisted of the two lead investigators for the project (Bryant and Krohm) supported by two experienced claims adjusters. The team read background documents regarding the L&I claim system and processes. Additionally, interviews with non-L&I stakeholders were conducted and analyzed, to discover anomalous practices that would be useful to focus on in file review. Before commencing the actual file review the team had a period of training on maneuvering through and capturing data from the LINIIS and ORION claims systems. The cooperation of L&I staff in answering questions about what we were finding in the files was tremendously helpful.

For file-review data, we reviewed 264 State Fund (SF) files and 144 self-insured (SI) files. Note that in selecting the files, we did not distinguish between "allowed" and "denied" cases, and there were only a very small quantity of denied cases in the sample. We did a follow-up review of an additional set of denied SF and SI cases, to evaluate the quality of the adjudication decision in the "denied" context. We determined that 46 SF cases (50/50 Retro/non-Retro split) and 46 SI cases were a sufficient sample. For SI cases, it was clear that L&I review of the denial decision was not in-depth, at least from the record, and was essentially cursory in nature and reliant upon the TPA rationale. For SF cases, it was clear that CMs took basic steps to review the evidence of record in making their decision. For these reasons, due to lack of variance around a predominant pattern of findings, review was stopped at 46.

Our general approach in reviewing claim files involved: 1) a preliminary phase; 2) a comprehensive phase; and 3) a follow-up phase. The purpose of the preliminary phase was to test the validity of the methods for review. Following preliminary testing, the file review checklist was modified to accommodate identified issues and help ensure more thorough and accurate reviews.

Preliminary Phase. We sampled 40 claims for an initial review. It was essential that we learn the most efficient techniques for examining digital files, how to interpret terms and classifications correctly, and confirm the efficacy of the "checklist" to be used during review. We also tested our process for documenting findings on each file reviewed, and developed audit work-papers. After the preliminary review, we modified our checklist, and returned to L&I for one half-day of additional testing in the immediate lead-up to the comprehensive file review, for final confirmation of the efficacy of the checklist and preparation for training of the file-review team for maximal efficiency. During this entire phase we made maximum use of experts in the Quality Assurance Section to advise us on terminology, procedures and exceptions noted.

Comprehensive Phase. The focus of the comprehensive phase was on state-fund claims. The strategy on reviewing practices in self-insured claims, as well as results from comparing the electronic data between state-fund and self-insured claims, is discussed below. During the comprehensive review the team reviewed 264 State Fund files, testing for the items listed on the checklist. The rationale behind the sample size is provided below. The team utilized a checklist, which was based loosely on the L&I internal review standards, but modified to focus on measuring system performance at selected junctures in the claim process. During each day of reviews and in a debriefing at the end of the day the team shared questions and tried to coordinate our use of the checklist. We sampled 144 self-insured claims. The sampling methodology is described below.

It is important to emphasize that we studied a process. Individual errors or deviations became important only if we detected a widespread pattern of inconsistent claims handling. Minor, individual deviations from procedure that did not rise to the level of a consistent pattern of behavior, or that did not appear to affect the claim outcome were not noted as a cause for concern.

Follow Up Phase. This part of the analysis responded to issues that required more in-depth study. We examined denied claims primarily to determine the level of review afforded by the CM. The SI and SF process is quite different, but the legal standard for denial is the same between the two groups. We sampled 92 denied claims, 46 SI and 46 SF.

3.2 SAMPLING

For State Fund claims, we sampled 264 files with total medical costs > $5,000 with accident years between 2010 and 2013. We selected 264 as our sample size because it represents a sufficiently large sample to accomplish the statistical analyses if the characteristics of the data fall in the reasonably expected range from data collected in the file review. The required sample size depends on several factors: 1) the nature of the statistic being measure (e.g., population proportion, cardinal values, or ordinal values); 2) the characteristics of the statistic itself (e.g., mean and variance) and the actual difference, if any, being compared (e.g., between Retro and non-Retro); 3) the statistical confidence one wants to assign to any difference being the result of chanced sample variation (e.g., 90%, 95% confidence level); and, 4) the probability that if there is a real difference of a certain size, one will identify the difference. We made some reasonable range of predictions about the expected values and performed power calculations over the range to estimate the required sample size.

For our sample of 264, we pulled a 55/45 split of retro/non-Retro. When the underlying population is large, the sampling should use two equal size groups for statistical testing, e.g., a 50%/50% mix of Retro and non-Retro employers for maximum efficiency in statistical testing for difference between the groups.[2]

As for which files to include in the samples, we determined that selecting from those files where total medical costs exceed $5,000 was the best approach to ensure fair representation of the full range of CM decision making on the claims, e.g., responding to complex and prolonged treatment, permanent disability rating, use of independent medical exams, and the need for vocational services. The file-review team did preliminary testing of claims to determine the appropriate level that would provide a more

[2] Given that, then the smallest standard error is achieved by drawing samples of equivalent size. The size of the underlying population of each doesn't matter. Even if the actual population were, to use an extreme example, 95% Retro and 5% non-Retro, as long as the population is large relative to the sample size, you get a smaller standard error by having similar sample sizes.

complete view of claims management services in sampled claims. Based on distributions from 2011, claims with medical cost > $5,000 represented 19% of all state-fund allowed claims, and accounted for 80% of total dollars. Thus, we sampled from important claims accounting for the majority of dollars while excluding files with few "actions" on which to base an evaluation of performance.

It is important to note that these file samples were only part of the evidence for the aforementioned performance characteristics. We combined the file review results with results from our analysis of the L&I claims databases. This allowed for rigorous analysis using the larger and more complete electronic dataset, providing a view into outcomes of particular actions (or inactions). For example, the data analysis gave us measures of the frequency of vocational services and claim details which we could not have reliably calculated from the file reviews.

For many, but not all, of the required research questions, we relied on electronic data to test for differences in claims handling between self-insured and fund employers. However, some of the questions could only be answered from file reviews, e.g. evidence of potentially biased decision making.

For the file reviews, the unique aspects of self-insured claim handling was an important context for developing an appropriate methodology. There are no legal differences in handling claims between self-insured and State Fund employers. The law regarding timing, validity, and benefits must be followed. If non-compliant with the law, a particular self-insured decision may be protested, and if so the protest is filed with L&I, and possibly, appealed to the BIIA. Timing and legal compliance of these decisions is tracked within the L&I database, and consistency was tested through analysis of the electronic record. Important for our analysis, the initial allowance/denial decision must be formally issued by L&I. In those cases where an SI employer is recommending an "allowance" order, it is our understanding that very little L&I independent fact finding and review occurs, which is understandable because the SI employer, who by statute must have a claims-management function, is asserting review of the claim and recommending allowance. Denials, on the other hand, are a context that can be used to compare consistency between claims handling and decisions by SI and SF employers.

Additionally, SI employers make treatment decisions, provide vocational rehabilitation services, evaluate permanency (and can issue a PPD order, although we understand that this is relatively rare), and make recommendations regarding pensions. Allowance decisions of medical-only claims are not reviewed. Order dates are available in the electronic record, which were analyzed to determine variances between state-fund and self-insured practices.

4 CUSTOMER OPINION SURVEY

4.1 OVERVIEW

Many questions posed by the RFP sought information on perceptions of employers and workers which could only be answered by querying the parties directly. Divergence between processes and outcomes and the perceptions of processes and outcomes might suggest important points for L&I education and intervention. For example, L&I's internal targets for completing various processes may be out of sync with the perceptions of some stakeholders, e.g., the timeliness of resolving protests.

The complex nature of questions posed by JLARC and the desire to compare perceptions across several subgroups, particularly by employer status (self-insured, Retro, non-Retro), required surveying multiple groups and attaining sufficiently large samples of completed interviews to reveal statistically valid

differences, if any, between the several groups.

We conducted opinion surveys of employers and injured workers, covering specific topics of interest to the audit team. The surveys (hereinafter called "Opinion Surveys") were conducted by phone as well as through online entry by some respondents.

For the survey, question format and wording were critical to success. We used focus groups to confirm the proper wording of the survey questions. For the focus groups we provided incentives to encourage participation. In addition to the focus groups we checked out understanding of the process with L&I experts to ensure the correct terminology for various situations, e.g., what is the best term for describing coverage provided by the State Fund, or the best term for an independent medical examination.

4.2 Managing the Survey

The survey contacts and recording of responses was managed by Q Market Research ("Q") as follows:

- Design, develop, and refine survey instruments
- Programming. The survey was programmed into Computer-assisted telephone interviewing (CATI) and an online tool. Q prepared a telephone instrument and an online survey instrument. The letter that was sent to the respondents explaining the reason for the survey offered him or her the option to complete the survey using the online survey tool. The CATI interview for the worker was also programmed into Spanish.
- Administration. The survey was again pre-tested on a sample and then finalized and administered. Interviewers were carefully trained, based on the lessons learned from the pre-testing activities.
- Finalizing the data. After data was collected, it is cleaned (coded and edited) and tabulated, and delivered to the research team for analysis.

4.3 Summary of Survey Responses Results

4.3.1 Employers

- Self-Insured (SI) Employers –Opinion Survey of risk manager staff/relevant HR person from employers; sample size = 165 actual responses (150 targeted)
- Insured employers, non-Retro rated (NR) –Opinion Survey of risk manager staff/relevant HR person from; sample size = 547 actual responses (450 targeted)
- Insured employers, Retro rated (R) –Opinion Survey of risk manager staff/relevant HR person from retro (including both group and individual retro) employers; sample = 697 actual responses (600 targeted)

4.3.2 Injured Workers

- Injured Workers (IW) for SI employers –Opinion Survey of IWs; sample size = 429 actual responses (425 targeted)
- IW for NR –Opinion Survey of IWs; sample size = 454 actual responses (425 targeted)
- IW for R – Opinion Survey of IWs; sample size = 658 actual responses (650 targeted)

4.4 Sampling Method

The first step in the sampling is matching employers between the three groups. This effort creates similar groups for comparing responses from self-insured employers to insured employers and, within

insured employers, similar groups of Retro and non-Retro employers. We used a propensity-score method, as described in Part 6 of this Appendix.

From the matched groups of employers, we then drew samples of employers for interviews.

From the matched groups of employers, we identified all claims meeting our selection criteria (date of injury within range and medical payments greater than $5,000). We then randomly sampled from among these claims at a rate that obtained sufficient samples to complete the target number of injured worker interviews for each group of employers. Injured workers were pulled for the sample regardless of whether they were represented by an attorney. Note: L&I injured worker surveys exclude attorney represented individuals, on the basis that they are prohibited, as a party, from making direct contact with such individuals; our project is not under these same constraints. If an individual responded to a call, "I can't discuss this with you on the advice of my attorney," we recorded the response as such. Also, we explained in introductory material that the information was anonymous and not part of any official record.

For the sample we drew claims from the years 2011 – 2013. The distribution of claims across the three injury years and the three groups of employers was carefully monitored so that the completed surveys match the targets within each subgroup. While we selected claims with total medical cost of $5,000 or greater, L&I surveys focus on claims with time-loss durations greater than 30 days. We determined, however, that selecting from those files where total medical costs exceed $5,000 is the best approach to include representative samples of the various required groups, as well as ensure large representation of other features of the claims process, e.g., vocational services and use of IMEs. Based on distributions from 2011, claims with medical cost > $5,000 represent 19% of all state-fund allowed claims, and account for 80% of total dollars. Thus, we sampled from important claims accounting for the majority of dollars while excluding files with few "actions" on which to base an evaluation of performance.

4.5 Results of Survey Contacts

Workers were mailed a letter explaining the purpose of the survey and asking them to fill out a survey on-line or contact the survey firm for an interview. Workers that did not respond received a follow-up postcard. If workers still did not respond, the survey firm called them. Up to 9 calls were made in an attempt to contact the worker.

Employers were also contacted by mail, explaining the survey and offering the call-in or on-line options. A follow-up postcard was sent. Finally, each employer not responding was called by Q Research. We also received assistance from the Washington Self-Insured Employer Association which sent an email request to members asking them to respond.

4.6 Comparisons with L&I Surveys

The methods that we used in the survey for the JLARC audit require different approaches to sampling employers and claims and conducting the survey than those used by L&I in conducting their customer opinion surveys, which are managed for L&I by Ipsos. Both approaches are well suited and appropriate for their specific purposes. But, the differing requirements necessitate differences in methods. First, the JLARC purpose is a bit different from the L&I-Ipsos objectives and consequently the surveys are designed differently, especially the sampling design. The primary focus in the L&I-Ipsos surveys is on how customers' perceptions change over time, specifically against the baseline at start. Our survey has a similar focus, but it adds a primary focus on whether different groups of employers or workers perceive

they are treated differently along the dimensions of self-insurance and, for insured employers, participation (or not) in Retro-rating plans.

The three statuses (self-insured, Retro-rated, or non-Retro rated) are characterized by different employer and worker characteristics. Since these characteristics are likely correlated with some of the issues at the heart of the performance audit, we needed to be sure we controlled for those characteristics, otherwise the comparisons among the groups would not be reliable. This makes the sampling more complex. The solution was propensity score matching. We end up with two pairs of matched samples for both the employer and worker surveys. We matched self-insured employers to insured employers (both Retro and non-Retro). Separately we matched within insured employers, Retro-rated to non-Retro rated. It is not correct to pool all insured employers in our sample and match them to self-insured, nor is it strictly correct to pool all the insured employers (workers) and compare the Retro to non-Retro rated employers (workers). The Retro and non-Retro insured employers in the sample matched to self-insured employers cannot, under the strictest interpretation, be used in the comparison between Retro and non-Retro employers (workers).

Second, self-insured employer claim data available for this study is not as complete as L&I's high quality data on State Fund claims. Consequently, we could not reliable use measures like time-loss and Kept-on-salary (KOS) when selecting the samples. The L&I-Ipsos survey approach uses time-loss and KOS as a set of criteria (which is correct for their sampling of State Fund employers). Our approach was to use the total medical cost as a selection criteria and set a threshold of medical cost that allowed us to focus on the 20% of cases that are more serious (80% of total cost). The resulting sample is quite similar but not identical to the L&I-Ipsos sample. We will have somewhat fewer small time-loss claims and slightly more large medical-only claims. This choice allows us to select very similar claims across all groups, especially when comparing self-insured and State Fund claims.

Third, the L&I-Ipsos approach excludes claims in which workers had attorney representation. We quite explicitly wanted to include represented claims since both workers and employers are more likely to have experienced challenges on these claims. It would also be more difficult to make reliable comparisons between, for example, self-insured claims and insured claims if the portion of attorney-represented claims differed based on whether the employer was self-insured or in the State Fund Because attorney representation is also correlated with the existence of protests, the potential problems of biasing the comparison could get worse as we examined claims with disputes. Whether or not disputes were handled consistently and equitably across the three employer types (self-insured, Retro, and non-Retro) was a high priority issue for JLARC. The L&I-Ipsos sample does not include attorney-represented workers in the surveys because of L&I concerns about potential *ex parte* communications.

The inclusion of attorney-represented claims could lead to differences in how worker perceptions compare between our results and those of L&I-Ipsos. Attorney representation is likely to be the result of more complex issues or disputes. Both of these characteristics are likely associated with more dissatisfaction with the claims process. Hence the JLARC audit, all else equal, will probably show lower customer satisfaction. This will apply to the worker survey, not the employer results. We do not believe the L&I-Ipsos sample excludes employers if one or more of an employer's time loss claims is represented.

Fourth, and quite important, we did not restrict our sample to claims that were "active" in the prior quarter. We include inactive and closed claims (including those that were denied) from all claims between 2010 and 2013 that met a certain severity level. Again, this is, in part, because it helps ensure that the samples are comparable between self-insured, Retro-rated, and non-Retro-rated workers and

employers. We take this different approach, also, because the basic differences in the purpose between the two surveys: the L&I-Ipsos method is focused on changes over time from a baseline. Our survey is focused on a certain period, not trends over time. Excluding inactive and closed claims would have made for highly skewed sampling from earlier years relative to later years. Readers should keep in mind that this choice can also affect the perceptions of the respondents, most importantly those of workers. Ipsos interviewed all workers very close to recent activity on their claim. Our survey interviewed some workers whose experience is further in the past, and the longer period of recall may affect their perceptions. The direction of any recall effect is not known. The most likely effect is to reduce more extreme views, both positive and negative. Trying to contact workers with inactive or closed claims is also more difficult, especially because the contact information may not be current.

Finally, our survey relies on "mixed methods." We allowed the workers and employers to choose to enter their survey responses on-line in an interactive environment or respond to a telephone interview. The mixed method approach has two advantages. First it can substantially reduce the cost of achieving sufficient samples. This was important in this instance because the scope of the audit, to compare across groups, required relatively large samples. Second, the mixed methods may help improve response rates among usually harder to reach populations.

4.7 Responses

Response rates were calculated as follows:

$$\text{Response Rate} = \frac{\text{Completes}}{\left(\text{Completes}\right) + \left(\frac{\text{Completes}}{\text{Completes} + \text{Not Qualified}} \times \left(\text{Not Contacted} + \text{Refused}\right)\right)}$$

The following table summarizes the results.

	Workers	Employers
Completed interviews	1,541	1,409
Refusal and mid-terminations--respondents who ended the interview before completion regardless of qualification	328	271
Respondents who do not meet the screening criteria and those respondents who would have qualified but their quota group was full	12	122
Applies to all final dispositions that do not fit any other category. For example, answering machine, wrong number, etc.	2,290	1,262
Response rate	37.2%	49.9%

5 Best Practices Survey

There is no universally recognized set of standards for handling workers' compensation claims. While certain practices are widely shared, workers' compensation systems exist in most states in a competitive business environment, and thus practices are proprietary to advancing particular business interests. Several of the questions involved in this performance audit of the Washington workers' compensation claims management function involved evaluating efficiency, as well as comparing results. To establish benchmarks that could be used to answer some of the questions required for the audit, we assembled a

"panel of experts" to participate in a survey. The survey asked general questions to the experts, who provided answers designed to address general claim management organization and performance. The panel did not address whether any particular Washington result conformed to "best practices"; rather, the panel was used to help derive a consensus benchmark that could be used to evaluate performance in general, including Washington performance.

The survey involved 14 respondents. All participants had lengthy careers in workers' compensation claims management. Experience included both front-line claim management experience, as well as supervisory experience. Most participants were involved in managing claims in the private, non-government context. The average professional experience for the respondents was 33 years, with extensive multi-state experience. The survey posed 25 questions about the claim management process, including:

- Number of Days for Lost Time Claims
  - In your opinion what time intervals would represent "best practice" goal for the claims adjuster (or nurse case manager) to make actual voice contact with the injured claimant? (0 = less than 1 business day to make actual contact; 21 = 21 or more business days to make actual contact)
  - In your opinion what time intervals would represent "best practice" goal for the claims adjuster (or nurse case manager) to make actual voice contact with the employer of injury? (0 = less than 1 business day to make actual contact; 21 = 21 or more business days to make actual contact)

- On average, how long (from the date of receipt of the accident report) does it take for an adjuster to communicate with the claimant that the claim is denied?

- How frequently would an adjuster (or nurse case manager) interact with an employer on strategies for returning the injured worker to the job within the physician's duty limitations? (For this question, assume that lost time payments are about to begin)

- % where IME needed re ability to return to work
  - As an estimate, in what percentage of lost-time claims, with disability over 60 days, is an IME needed by the adjuster to confirm or challenge the treating physician on the following issues. (0% = IME never needed; 50% = IME needed half of the time to confirm or challenge the treating physician; 100% = IME needed in every case on the particular issue)

- Generally speaking, how reliably can an adjuster predict, after 60 days of lost time, that a worker with a moderately severe injury (major sprain to a joint, tendon tear, etc.) will not likely return to work at the employer of injury ?

- Vocational evaluation (e.g. job skills assessment; ability to work)
  - What percentage of lost-time claims usually require the following:

- Vocational retraining plan
  - What percentage of lost-time claims usually require the following:

- Total number of open cases per front-line adjuster
  - In your opinion, based on average adjuster training and experience and assuming average case complexity, what would be a standard caseload per workers' compensation claims adjuster.

- Number of open lost-time cases per front-line adjuster
  - In your opinion, based on average adjuster training and experience and assuming average case complexity, what would be a standard caseload per workers' compensation claims adjuster.

There was a high degree of agreement of opinions expressed on most questions. The complete survey instrument is attached to Appendix 8.

6 DATA ANALYSIS OF L&I CLAIM DATA

6.1 INTRODUCTION

Our methods for conducting data analytics started first with becoming acquainted with the L&I data warehouse characteristics. We did this through review of the data dictionary and extensive interviews and questions with L&I research staff. After appropriate confidentiality agreements, we went through a formal data selection and transfer process. A significant dataset, covering claims with accidents between 1/1/10 and 12/31/13 was developed. We used an industry standard, open-sourced statistical software package, known as "R," to develop a database for inquiry.

From the onset of this study we recognized the challenge of obtaining, editing, and analyzing a very large and complex dataset constructed by L&I. The study could not have been completed without the full cooperation of L&I in supplying the correct data elements and assisting us in their interpretation. The cooperation in interpreting the data by L&I has been superb.

The data quality controls begin with correctly understanding the definitions of data elements and the way data are captured, edited, and recorded. This process begins with in-depth discussions with the L&I data warehouse managers. As described below, we conducted our own edit checks of data received from the L&I database to measure the conformance of records to data definitions and plausible values. Also described below is the process by which suspect data were evaluated for inclusion. In the case of anomalies, we consulted the appropriate authority within L&I for an explanation; following this we applied consistent standards for either reconstructing the record or excluding it from the analysis database, and documented these actions in case of later review.

There are significant distinctions that cut across several of the comparisons in the RFP between self-insured and state-fund handled claims. We discovered a large number of data elements on records of state fund claims that are not available for self-insured claims. This impacted cross-states comparisons to a certain extent, and in coordination with the Washington Self Insurer Association we sought inter-state data from TPA members with experience in states neighboring Washington; we were not successful in persuading TPAs to supply comparative data. We do believe that we were able to gain enough comparative data from other sources to conduct a solid analysis.

Through interviews we noted that the nature of the L&I review for self-insured claims differs from the state-fund context. This review process of TPA decisions is considered "oversight" and not original claims investigation. Reported exceptions include the denial process and the segregation process, and to some extent the closing process. The role played by L&I in these processes were described as more substantive than "simple" oversight; we note that these were "reported" exceptions because our observation, through file reviews, did not indicate a true difference in actual practice; in other words, we did not observe in the course of file reviews that L&I performed a review of these decisions that made a difference in the outcome.

Many of the characteristics analyzed in the context of legal decisions by L&I involve various aspects of timeliness. The RFP specified a number of these, which was driven by the difference between various dates available in the claim record. To prepare for this analysis, we first verified that these dates had the proper logical relationship (for example, date of bill payment is after date of bill receipt).

After this step, we computed lag times between relevant dates for each measure. In prepping for these queries, we found that some values fit the strict data definitions, but needed to be excluded as atypical, such as claims with long processing lags due to an initial denial decision being overturned for the first time after several levels of appeal. If extreme outlier values distorted results, we trimmed data that exceeded or fell below 3.5 standard deviations from the mean.

Lag times. A lag-time is the elapsed time between the conclusion of one event and the occurrence of a subsequent event. For lag time measures we looked at the distribution of time lag days for each group. This included various standard measures (means, percentiles, standard deviations). When data values were excluded as atypical, it was noted.

Proportions. Some of the legal-decision questions under analysis referred to statutory measures that specified a timeliness benchmark; we computed the proportion of cases that met the timeliness standard in addition to running measures of distribution. For many timeliness measures a reasonable goal was to maximize the share that conform to the standard, and thus resources were not devoted to improvement beyond meeting the standard. Thus it would be possible to view multiple aspects of the time to issue a particular decision, for example:

- Mean (average) time overall to issue the decision was 36 days;
- Median time was 26 days;
- 86% of claims measured met a timeliness standard of 30 days.

We established a “target” standard or standards, which was based on a number of factors, including statutory requirements, stated policies, and industry best practices. The target could also be a mean or median, and the proportional analysis would be based on what percentage of values is within certain ranges from that target, similar to a standard-deviation presentation. We were flexible in utilizing those standards that are most “resonant” with stakeholders, determined through review of L&I law and policy as well as acceptable norms.

6.2 MATCHING

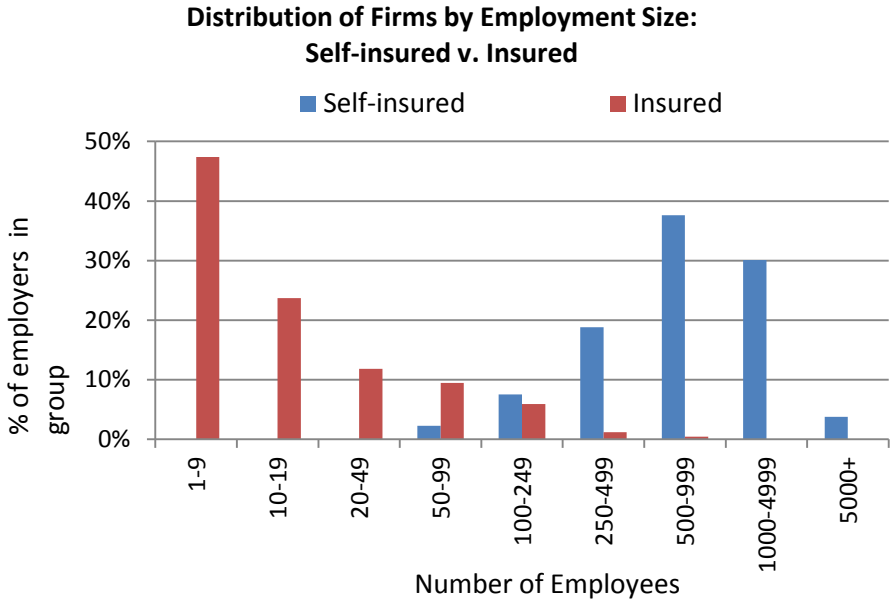
When an analysis requires comparison of measures, like opinions or performance, between two groups, it is important to control for factors that may lead to erroneous or unsupported conclusions. This audit aimed to compare various measures across different groups of employers in two pair-wise comparisons: self-Insured employers to Insured employers and Retro-rated (State Fund) employers to non-Retro-rated (State Fund) employers. Matching employer types was an important methodology challenge. JLARC's objective was to determine if L&I activities lead to actual differences in claims handling or employers' and workers' perceptions of claims handling. It was important for us to distinguish differences driven solely by, for example, employer differences, versus those driven by activities specific to L&I.

This challenge was present both with respect to data analysis in general, and also with respect to compiling a valid sample for conducting employer and worker perception surveys and analyzing results.

Our strategy involved first selecting employers and carefully matching the employers across the different dimensions to make the statistical analysis as accurate and precise as possible. The next step

involves sampling claims (and workers) from within the matched samples of employers. The impact of this two-stage approach is to create convincing, highly defensible inferences about the impact of different government processes (State Fund, L&I, and BIIA), independent of differences in the underlying employers and claimants. We describe this below.

Sampling employers. Regression techniques meant to control for differing characteristics when comparing outcomes between two or more groups can be improved on in many situations. Most importantly, when: 1) membership in one group over another involves some element of choice (here, whether to self-insure or choose a Retro program); or 2) the overlap between the two groups is limited (e.g., self-insured vs. insured and employer size), then standard regression approaches cannot be reliably used without likely creating biased results. This is shown visually below using the single dimension of firm size.



Source: WorkComp Strategies

In this example, we present a hypothetical distribution of firm size for self-insured and insured employers. Most employers are small, fewer than 50 employees. But virtually no firms smaller than 50 employees self-insure, in part because they do not meet minimum financial requirements. At the other extreme, virtually all firms larger than 1,000 employees self-insure (in our hypothetical example). If one uses standard regression techniques to control for characteristics, in this case firm size, the method extrapolates firm size beyond the ranges in which it is comparable between the two groups of employers. That is, the effect of firm size on a measure of interest, like time to return to work, may not matter for very large firms in the way it does for very small firms. Consequently, it is difficult to control for the effect of firm size when comparing very large and very small employers.

The state-of-the-art approach, propensity score matching, is to first match employers exploiting the unobserved process by which they make the decision to be members of one group over another. That approach uses logistic regression to model the probability that an employer will choose, for example, to self-insure, based on a range of available characteristics (size, industry, or injury experience, etc.). Indeed, the biggest methodological problem we face is matching employers for the comparison groups.

As one adds different dimensions (size, injury experience, industry, location, availability of re-insurance, state-specific factors such as the public nature of the filings, etc.), it quickly becomes complex, if not impossible, to judge which firm in one group is the best match to a firm in the second group. This is notwithstanding the perhaps very "personal" or individualistic part of decision to self-insure, including the degree to which the company is willing to assume risk.

We performed propensity score matching in some measures to address this issue. Using this technique, the regression coefficients from the logistic regressions can be combined into a single score, referred to as a propensity score, which is used to match employers. This method has been tested and proved to be more efficient and to produce better matches than other, formerly used approaches.[3] Some outliers in both groups (e.g., very small insured employers) may be excluded because no near matches can be found. We used this matching process as part of the first stage of comparison of measures across self-insured, Retro, and non-Retro employers.

We matched employers the following dimensions:

- Employer size (hours for insured employers and employees for self-insured)
- Experience rating (for insured employers) and pseudo-x-mod for self-insured[4]
- Primary class code (NAICS); we assigned a primary class code to self-insured employers if not available from claims data)
- Employer has exposure in more than one class code in year (Y/N) (imputed for self-insureds)
- ZIP Code (Several instate geographic regions and out-of-state headquarters)
- Multi-state employer (if we can determine this dimension)
- Primary NAICS Code (2-digit) (assigned by us)
- Years in business (<1, 1-2, 3-5, 6-10, 11+)

These data were readily available from our electronic database.

The next step in propensity score matching is selecting a method for choosing among all available matches in one group when matching to an employer in the second group. We chose the "best match" based on the closest propensity score. We also defined a range outside of which we would not match. That is, if no match is found within +/-X of the original employer's propensity score within the other pool, we dropped the original employer from the analyses. Since we are matching two groups of employers, rather than strictly a "treatment" and a "control" group, matching was done without replacement. Each of these choices requires some experience with the data to understand the distributions and the degree to which employer characteristics overlap. Consequently the precise choice was dependent on review of the electronic data available to us. These decisions were documented and explained in the interim review meetings with JLARC, including estimating impacts from the decisions.

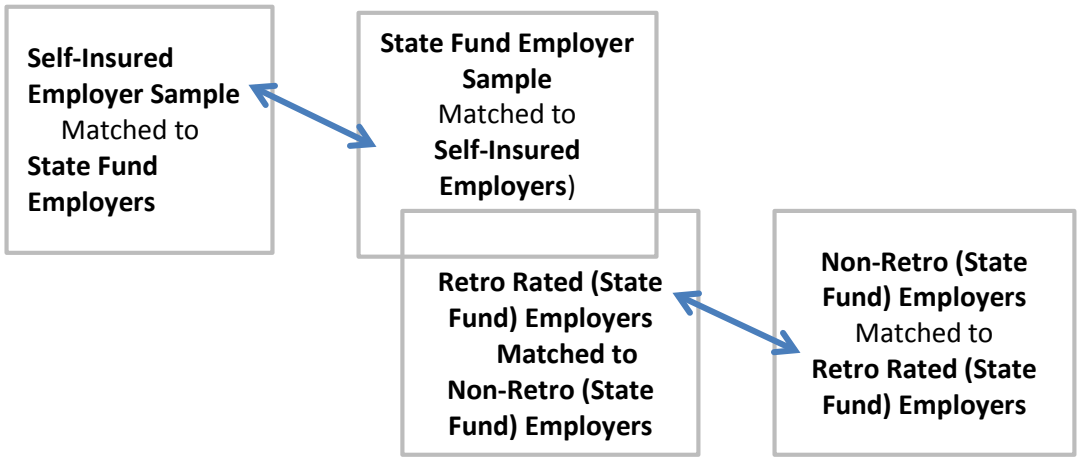
Note that the above discussion of matching applies equally to comparing self-insured and State Fund employers and Retro and non-Retro employers.

This process results in two sets of paired samples. The first pair of samples will have matched similar Insured and Self-insured employers based on the method described just above. The second pair of

[3] See, e.g., Heinrich, et al., "A Primer for Applying Propensity-Score Matching," Inter-American Development Bank (2010), and sources cited therein.

[4] The pseudo X-mod will be created by modeling the x-mod and frequency of injuries for insured employer by NAICS code and translating this to self-insured employers.

samples will match the Retro insured employers to the non-Retro insured employers. These two sets of paired samples are diagrammed in the figure below.



In the above diagram, two of the four samples, Insured employers (matched to Self-Insured Employers) and the Retro Rated Insured Employers (matched to the non-Retro Rated Insured Employers) overlap. That is, some employers may be in both samples. This does not pose any statistical problems, since the two pairs of samples are analyzed independently.

As indicated in the first diagram, there are some employers that will not match sufficiently closely to another group to be included in the analytic, matched samples. This is appropriate in this particular analysis because the interest of JLARC is to examine whether the different regimes (Self-insurance, Retro-rating participation, and non-Retro-rating participation) result in different outcomes for workers and employers because their employer status leads to differences in claims handling by L&I.

The stated objectives in the audit did not include, for example, analyzing specifically how small employers are treated relative to large employers or new employers relative to established employers. While our approach is not specifically designed for these types of distinctions, the approach used in our study does allow additional dimensions to be studies with confidence.

Sampling Claims and Workers. Matching employers using propensity scores should result in a very similar pool of claimants and claims. Ordinarily with propensity score matching other characteristics, beyond those matched on, will also be very close in terms of means and distributions. However, some dimensions of the claimants and claims may still differ in important ways that we feel might affect our inferences about the L&I's claims handling activities being studied.

By studying features within the claims process we can offer a much richer picture of differences in legal decisions than just the main distinctions between self-insurance, Retro employers, and traditional state fund employers. These sub-issues may help explain the gross differences we may find across groups.

The combination of propensity score matching and regression control is the best way to evaluate whether there are differences in important measures (e.g., consistency, fairness, timeliness, etc.) across the different groups of employers on legal decisions and protest handling. The data can be used to drill down and study what particular issues drove the protests or appeals, e.g., wage, PPD, or Pension. If differences are found, this approach will give, as near as possible, unbiased estimates of the size of any differences. As noted above, a solid measurement of the differences between groups has been accompanied by some explanation of the reasons for the differences and whether these causes are benign or need correction.

General Data Description. As indicated earlier, our dataset consisted of claims with dates of accident between 1/1/10 and 12/31/13. We used the calendar year approach as opposed to the development year approach because of the extremely long durations in Washington. For example, if we had measured all activity in a certain calendar year, we would be combining claims with many different years of activity. This would have made the dataset unworkably large. When development of a claim to maturity was needed for analysis (such as measuring ultimate durations of TTD), we were able to rely on L&I actuarial data. The L&I research and actuarial services teams were extraordinarily helpful.

Some general characteristics of the dataset: For all employer groups 2013 is likely to underrepresent the ultimate total of accepted claims; this is due to late reporting of claims and lengthy investigations for some claims. Note that SI data reporting can be delayed, which we expect resulted in lower claim counts for 2013, which were small relative to 2010-2012.

All reported claims by injury year

Measure	Number of claims
Total number of claims	569,262
Injury year 2010	144,037
Injury year 2011	142,127
Injury year 2012	144,482
Injury year 2013	138,616

All reported claims by injury year and medical only or timeloss

Year	No timeloss or medical	Medical only	Timeloss
Injury year 2010	13,564	98,769	31,681
Injury year 2011	14,133	97,236	30,705
Injury year 2012	13,583	100,977	29,716
Injury year 2013	10,643	88,808	25,458

Note that the total of these columns does not match the total above for all reported claims; some reported claims end up being excluded for various reasons, e.g., duplication, erroneously reported, etc.

Accepted claims by injury year and medical only or timeloss

Year	No timeloss or medical	Medical only	Timeloss
Injury year 2010	5,110	90,041	31,302

Injury year 2011	4,739	87,774	30,247
Injury year 2012	5,190	90,115	29,102
Injury year 2013	3,575	76,320	22,722

Note: 2013 results will increase as more claims are reported and investigations concluded.

Accepted claims by injury year and injury or illness

Year	Illness	Injury
Injury year 2010	5,288	121,170
Injury year 2011	5,174	117,598
Injury year 2012	4,840	119,625
Injury year 2013	3,470	106,937

Accepted claims by injury year

Year	Illness/Injury
Injury year 2010	126,458
Injury year 2011	122,772
Injury year 2012	124,465
Injury year 2013	110,407

Accepted claims by injury year and SF or SI

Year	SF		SI	
Injury year 2010	86,929	(69%)	39,529	(31%)
Injury year 2011	85,422	(70%)	37,350	(30%)
Injury year 2012	87,733	(70%)	36,732	(30%)
Injury year 2013	85,639	(78%)	24,768	(22%)

Note that many SI claims are reported long after they occur.

All SF claims by injury year and Retro or non-Retro

Year	non-Retro	Retro
Injury year 2010	55,870	43,117
Injury year 2011	56,391	41,378
Injury year 2012	58,545	42,672
Injury year 2013	52,433	37,430

BIIA data. In additional to published statistics by the BIIA data staff, we also received a large dataset from BIIA. This was in the form on an Excel spreadsheet. It contained data and a data definition lexicon of final orders from BIIA for the years 2012 and 2013. By using final orders some cases will have had dates of injury well prior to 2012; many cases decided in 2012 will have been filed in 2011 and even earlier. In addition, some appeals filed in 2012 and 2013 also were not included because they were not yet concluded. This was a cross section of decisions from 2012 and 2013 regardless of the date of injury or the date the appeal was filed. We also received reports developed by BIIA showing duration lags for key throughputs, such as time to decision. The dataset also included "issue" information, meaning those issues identified by the staff when the final order was issued. No opinion was given on whether any particular issue was more crucial to the case than others; rather all identified issues in a case were listed. BIIA data staff were very helpful in interpreting the data. We were able to use this issue information to gain insight into prevalence of certain issues. The data also included whether the appealing party was representing by counsel when filing the appeal. The data also included outcome information, including a flag by BIIA as to whether the particular order was a "reversal" or not. In this way the data provided a view into outcomes on appeal. The reversal information did not include partial reversals, however; i.e., if a claim was appealed on several issues, one of which was determined to merit reversal, then the entire claim was considered "reversed."

7 COMPARATIVE DATA ANALYSIS

The audit team conducted phone and in-person interviews with officials from other state workers' compensation programs. The audit team also collected data from several states. Some information was obtained through special records requests, but most was available on websites. We made personal requests for data from many states. States particularly helpful in providing information were North Dakota, British Columbia, Saskatchewan, Montana, Ohio, Oregon, and Idaho. We focused on the following key comparative data points:

- Denial Rate
- Time to initial Payment
- TTD Duration
- Time to Closure

We used WCRI and NCCI data to provide further comparisons; in some instances comparisons with up to 37 states was available. We conducted interviews with Idaho, Ohio, and Oregon self-insurance managers to gain comparative insight into their self-insurance programs.

Our targeted selection of jurisdictions was based on the following considerations:

- Proximity to Washington. Neighboring states are usually regarded as interesting comparisons by policy makers.
- Preference for monopolistic systems. The inclusion of BC, Saskatchewan, North Dakota, and Ohio was appropriate because of the shared insurance approach.
- Similarity in economy and size, as used to identify candidates for comparison shown in the Methodology Appendix to the RFP.
- Professional relationship with management in the jurisdiction. This refers to our ability to persuade an agency to perform custom analytics.
- Jurisdictions reported in the WCRI CompScope™ and the NCCI disability durations reports on selected measures.

Given these diverse selection criteria, we hesitate to call our selection "representative." There are two glaring problems with all comparisons of WC data across states. First, there are many factors that would explain persistent differences among states, e.g., disproportionately high employment in high-risk industries, the proportionate number of self-insured or high-deductible employers, or variations in causation standards and claim waiting periods. Even if there were a match on one important characteristic it would be very rare to find a match on multiple characteristics. For example, compared to Washington, the Oregon body of self-insureds includes fewer entities and fewer very large corporations, while British Columbia has a small number of self-insureds concentrated in a few large employers and a few industries. Neither is representative of Washington's situation. A second problem in comparisons is the different ways data are defined, collected, edited and reported by jurisdictions.

An additional constraint in terms of inter-jurisdictional comparisons involves the unique aspects of workers' compensation programs in the US and Canada. Each jurisdiction has an individualized set of laws and regulations, resulting in difficulties in drawing strict comparisons. There are many procedural and legal differences that complicate particular comparisons of jurisdictions, e.g., number of permanent total disability claims or percentage of denied claims. Notwithstanding these methodological challenges, we did find a large number of meaningful measures of Washington's performance relative to other jurisdictions.

When performing comparisons of Washington workers' compensation program with other state programs, a major caveat is the unique treatment in Washington of self-insured employers. Specifically, Washington has two systems for controlling the process of claim adjudication: one for State Fund claims and one for self-insured claims. Many other states will report on regulation of claims activity as a whole – all claims, both insured and self-insured. For example, in the 2012 *Report on the Oregon Workers' Compensation System*,[5] the table on page 24 reports on the "Insurer claim acceptance and denial, median time lag days," but we confirmed that the data in that table includes "traditional" insured claims, self-insured claims, and state-fund (assigned risk pool) claims – in other words, all claims.[6] In our performance comparison (Chapter 5 of the report) we made adjustments to the Washington data to better compare it to other states, and disclosed major methodological differences in multistate comparisons.

In conducting our inquiry, the supplying states were asked to document any factors that might deviate from the stated request, e.g., first payment date is supplied voluntarily by a subset of self-insured, government self-insured are not counted in the data, or denials exclude certain types of denials (duplicate claims, out of state employment, etc.). The states were asked to cite any statutory standard or administrative goal for first payments, e.g., 80% of lost time claims paid within 14 days of date of injury. The response on this request for elaboration and documentation was generally poor.

Denial rates are seldom computed and published, by the insurance industry, self-insurers, research organizations, or government agencies.[7] Also, denial information must be carefully defined, since denial statistics may or may not include summary denials arising primarily outside of the claims management process (e.g., lack of employer coverage, claimant not an employee, duplicate claim, etc.).

In Washington, there was no need to contact self-insureds or their TPAs regarding the payment promptness and denial statistics. Both are available for the entire population of self-insureds via the electronic database. These data were tested and appear to be relatively sound. We combined the self-insured data with the Washington state fund data for analysis with other jurisdictions. Note that we did seek TPA data regarding the four comparative questions set forth above, in addition to the question of time to provider payment for initial treatment, for neighboring states to Washington (OR and ID). We worked on repeated requests with the Washington Self Insurer Association Executive Director, who was supportive of the request. Unfortunately, no TPAs were willing to supply the information. Fortunately, the above described methods resulted in sufficient comparative information.

[5] Available at http://www.cbs.state.or.us/external/imd/rasums/2362/11web/11_2362.pdf.

[6] Note that we have received data, based on a special request, from OR that is self-insured specific, thus simplifying (and improving) some comparisons considerably.

[7] Oregon is the shining exception; they publish denial statistics for insurance each year. Minnesota had a special project on denial rates in the early 1990s.