Analyzing WCB Injury Rate Data:

How to properly analyze and interpret your injury rate data.

Converge Consulting Group

People • Process • Performance

Converge Consulting Group Inc.
www.converge-group.com
403 266 0061

Converge Consulting Group Inc.

Mailing: 2617-20 Street SW
Calgary, Alberta
Canada
T2T 4Z4

Phone: 403.266.0061
Web: www.converge-group.com
www.operationalexcellence.ca

Analyzing WCB Injury Rate Data

How to properly analyze and interpret your injury rate data

Background

On October 01, 2010, the Alberta Government publically released workplace injury & fatality records. This made the Province the first in Canada to release a comprehensive list.

Anyone can now access information about worksite injuries and fatalities involving more than 140,000 employers insured by the Workers Compensation Board (WCB).

Information for each employer includes:

- number of lost-time claims
- estimated number of employees
- lost-time claim rate
- number of fatalities
- if the employer holds a Certificate of Recognition, and
- industry and province-wide lost-time claim rates for comparison purposes.

The records can be found online at:

www.employment.alberta.ca/employerrecords

As may be expected, the release of this information has generated a great deal of interest. People are naturally interested in knowing:

- whether trends exist in the data, that is whether injury and fatality rates for organizations or industries are getting better or worse, and
- whether some organizations or industries perform better than others.

The need for analytical guidelines

However, no guidelines or instructions are provided for ensuring the data is *analyzed properly. How do you determine if your statistics are ge*tting better or worse over time? How do you tell if your performance is above or below that for some other organization?

It is these questions this paper will answer. We outline the proper approach for analyzing employer injury rate data (or any other organizational performance data) as released by the Alberta Government and the WCB.

There are other issues and concerns with using WCB data. However, for the purposes of this paper, data is taken at face value. The focus is strictly on proper analytical approach, not on data validity.

Converge Consulting Group People • Process • Performance

Converge Consulting Group Inc.
www.converge-group.com
403 266 0061

Overview of the Data

The data is presented over the Employment Alberta website. An image of the results for the City of Calgary is presented below.

Data Output and Presentation for the City of Calgary

Government of Alberta

Five Year Employer Lost-Time Claims and Fatalities

Date: 9/28/2010

Industry Name: CITIES - 95104

Year	COR*	Lost Time Claims	Person Years Estimate**	LTC Rate***	Fatalities
Employer Name: CITY OF CALGARY					
Common Business Name:					
City: CALGARY					
Account Number: 1B					
2005	Yes	503	14267.70	3.53	5
2006	Yes	591	14504.00	4.07	3
2007	Yes	625	15193.38	4.11	3
2008	Yes	677	16558.92	4.09	4
2009	Yes	665	17295.97	3.84	1

2009 Industry Total	1277	41534.50	3.07	6
2009 Provincial Total	26096	1539737.86	1.69	110

* COR indicates if an employer has a Certificate of Recognition any time during the year.
** The person years estimate is calculated by AEI.
*** LTC Rate is the number of Lost Time Claims per 100 Person Years (LTC Rate is not calculated if the employer has less than 40 Person Years).
The UNASSIGNED Account is a general WCB-Alberta holding account used in cases where an employer has not been identified on a claim
The Industry Total Person Years Estimate is inclusive of all employers within the industry.
**** "Under Review" indicates data for unreported years are being verified.

Source: Government of Alberta Employment & Immigration

The presentation of data in this way, while extremely useful, can encourage improper data analysis. For example, there may be tendency to compare the City of Calgary LTC results over the past 5 years (average is 3.93) with the industry average (3.07). Direct comparisons of individual results with averages, while commonly used, yields worthless and more often than not, harmful results.

Differences, between the City of Calgary and other organizations or any combination of organizations such as an industry average, may be attributable to random fluctuations (system/statistical noise) or to important differences in organizations and their respective safety systems (system/statistical signals). Separating the signals from the noise (the wheat from the chaff) is what proper statistical analysis is all about. It cannot be done by simply comparing averages.

Data Definitions as provided by Alberta Employment

Data Element	Description
Employer Name	Name of company, sorted alphabetically within industry.
Common Business Name	Operating name or the commonly known name of the business when it is available. This name may or may not be the same as the employer name. .
City	The city of the employer's mailing address as provided to the WCB. Although the address may be outside Alberta, this report includes only Alberta worksite claims and person-years.
Account Number	An identifier to distinguish an employer's WCB account or multiple WCB accounts.
Year	Occurrence year of the claim and person-years associated with that operating (calendar) year. The report will collect and display the past five operational years.
Certificate of Recognition (COR)	Indicates if an employer had a COR anytime during the reported calendar year.
# Lost-Time Claims	Total number of accepted lost-time claims, based on definitions elsewhere in this document, as reported by March 31st of the year following the occurrence year.
Person-Years Estimate	Estimated person-years worked calculated by AE&I from WCB's wage and insurable earnings data. One person- year can be assumed to be 2000 hours worked per year.
LTC Rate per 100 PY	The number of lost-time claims per 100 person-years. LTC Rate is not calculated if the employer has less than 40 person-years, so this field would be blank.
Fatalities	The number of fatalities accepted in that calendar year.
[Year] Industry Total	Statistics for the industry in the most recent year reported.
[Year] Provincial Total	Provincial statistics for the most recent year reported.

Source: Government of Alberta Employment & Immigration

Analyzing the Data Right

Analyzing the data right, means taking an analytic, rather than an enumerative, approach to the analysis.

- Enumerative studies count populations of objects within a static frame with the aim of describing current conditions. A poll is an example of an enumerative study - it counts the number of people that claim they will vote one way or another (if the election was held tomorrow).
- Analytic studies analyze processes, the cause & effect relationships and levels of performance of a dynamic system with the aim of explanation and prediction. This is what we are after in examining injury rate data, understanding a dynamic process or system and specifically, the causes and levels of injury rates. [1].

The primary tool for understanding process behavior is the control chart[2]. Control charts:

- presents the data in context,
- preserve the time order in which the data occurred,
- provides a means to test for the homogeneity of the data without which, no summary descriptions or inferences can or should be made, and
- provide empirical tests that separate signals from system/statistical noise.

The proper control chart to use in analyzing the data from a single organization, such as the City of Calgary, is the *XmR* Chart (also referred to as the *ImR* Control Chart). Most statistical packages have routines to calculate the necessary statistics and display Control Charts[3].. the following observations are made concerning the *XmR* control chart (:

- The chart actually consists of two charts (*X* and *mR*). The upper chart (*X*) plots the values of interest, in this case the LTC Rate for the City of Calgary over time. The second, lower chart, (*mR*) plots the differences between LTC rates for the years provided.
- Data points going beyond the upper or lower control limits (red lines) in the upper (*X*) control chart are signals of important differences from the background systemic/statistical noise. No such points arise in the Calgary data indicating that no special causes or signals are present - all year to year changes are random fluctuations or statistical noise.
- Patterns in the data of the upper control chart may reveal the emergence of trends either upward (things are getting worse) or downward (things are getting better). With only five data points, interpretations are tenuous. All data points would have to display continuous increases or decreases for there to be a conclusion that trend *may* be present.

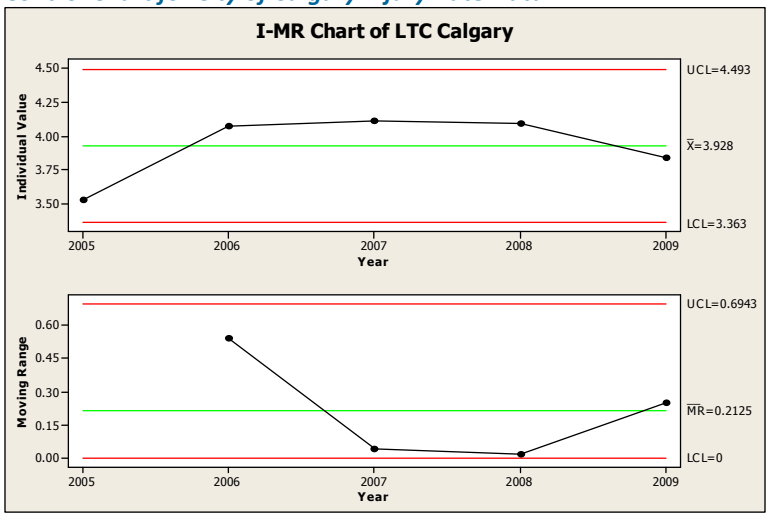
[1] More information on the difference between enumerative and analytic studies can be found in, On Probability as a Basis for Action by Edwards Deming

[2] Proper rules for control chart preparation and calculation of control limits must be followed or the control chart will be no more useful than any other data analysis technique.

[3] Without going into the details of control chart preparation, analysts must be careful that the correct method of calculating sigma (the basis of the control limits in the control chart) is used. It is incorrect to use standard deviation calculations to estimate sigma. The correct method is to base sigma calculations on the d_2 bias correction factor for the average Range. To check your software, take the average Range (in the lower mR chart) and multiply it by 2.66. Adding the result to the Average of the X Chart should equal the upper control limit for the X Chart. If it doesn't, your software is calculating sigma incorrectly.

Converge Consulting Group Inc.
www.converge-group.com
403 266 0061

Control Chart for City of Calgary Injury Rate Data



Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

- Data points rising beyond the upper control limit in the lower chart indicate a lack of homogeneity in the data - a sign of (likely) significant change in the process or system within the time period for which data is available. That isn't present in this data. Thus, the interpretation is that the data is homogenous and comes from a consistent cause and effect system. No significant changes in the system affecting performance have occurred at the City of Calgary over the last five years.

Is five data points enough for the *XmR* Chart?

In a word, no. Like most statistical tools, *XmR* Charts prefer to have more data than less. The greater the number of data points, the greater the certainty in the limits displayed on the chart.

How many data points?

Number of Data Points	Certainty of Control Limits
2-5	Soft Limits
6-10	Limits Gel Use with caution
11-30	Limits Solidify Use with confidence
30+	Limits Solid Anything over 50 data points will not provide any additional certainty on the estimated control limits.

Source: Based on descriptions in Six Sigma Practitioners Guide to Data Analysis by Donald Wheeler.

With only five data points provided by Alberta Employment & Immigration, the limits in the *XmR* Chart are considered 'soft'. This means that the chance of the *mR* Chart detecting a signal of non-homogeneity is very low. Which also means that if a signal is detected, it is almost certainly a 'real' signal.

Despite the lack of data, the *XmR* Chart is still the best analytical option for the analysis of these types of smaller data sets. This is because:

- the limited ability of the *XmR* Chart to detect a lack of homogeneity in the data is still far superior to simply pretending or assuming the data is homogenous - which is the case with other statistical analysis tools.

Converge Consulting Group
People • Process • Performance
Converge Consulting Group Inc.
www.converge-group.com
403 266 0061

- the *XmR* Chart provides context, presenting the data in its original time order, and
- the *XmR* CHart provides a clear, concise and easy to understand picture of process behavior.

This makes the *XmR* Chart the best option for proper analysis of Alberta Employment & Immigration injury rate data. Analysis of the data should proceed with an understanding of the levels of certainty in the control limits given the number of data points at our disposal.

As new data becomes available and added to the analysis, the confidence in the control limits will also grow. The addition of data in future years will also provide an method of determining the impact of health & safety changes for any single organization or for a broader industry.

Comparing two or more organizations.

Are we better than the other guys? It's a natural question. Most managers want to know how their performance stacks up against the competition or at least other organizations.

A tale of two cities

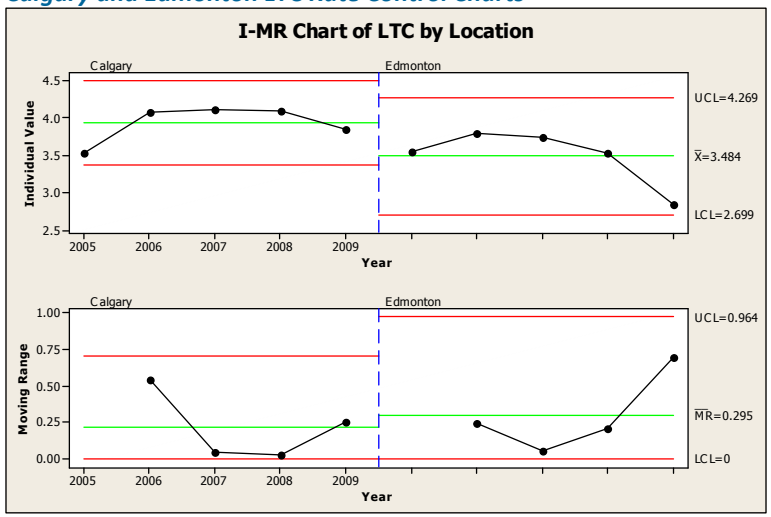
To see how to assess this, let's compare the City of Calgary to another city in Alberta - Edmonton. Edmonton is the only other city in Alberta of comparable size. The key question for Calgary, are we better or worse than Edmonton in terms of Loss Time Injury Claims (LTC)?

Because the control chart represents our best representation and analysis of the data for any one city, it follows that comparing the control charts of each city is the best way to gauge the comparative performance of these two cities.

In practice, this means plotting the respective control charts beside one another. Interpretation steps include:

- Examine the lower (*mR*) charts to ensure the data for both cities is homogenous. If it isn't, things have changed for one of the comparison subjects. In this case, however, the data for both Edmonton and Calgary appears homogenous. As such, we are entitled to go to the next step in the analysis.

Calgary and Edmonton LTC Rate Control Charts



Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

- If any data points from one *X* Chart, lies above or below the control limits from the comparison control chart, we have a signal that the two systems are different in some important way that is worthy of further investigation. We don't know what the cause of this difference is. We only know it is large enough to be worthy of investigation. That is the case here. The 2009 data point for the City of Edmonton is below the lower control limit for The City of Calgary.

- Had we found no detectable differences between Calgary and Edmonton (had all data points been within the control limits of each other's *X* Chart) it only means we failed to detect an important difference. It doesn't mean there are no differences between Edmonton and Calgary. We know without doing any analysis that Edmonton and Calgary are different (that's why they have different names). It only means we have failed to detect any important differences in LTC worthy of investigation.

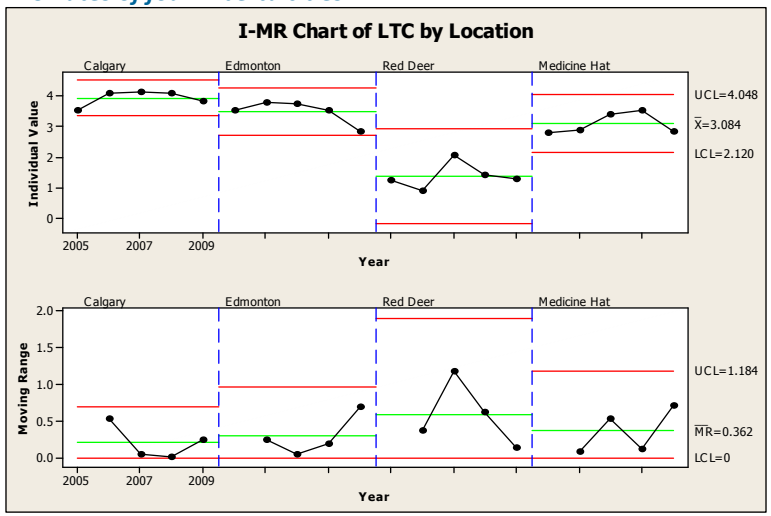
With only five data points, these interpretations are somewhat tenuous because our limits are quite soft. However, the very large difference between Edmonton and Calgary, with Edmonton's average line very near that of Calgary's lower control limit, provides reasonable evidence that the two cities are detectably different.

Comparing more than two organizations.

What works for two, works for three or more organizations as well. We are only limited by the width of the paper. To demonstrate this, control charts were prepared for two additional Alberta cities - Red Deer and Medicine Hat. Calgary and Edmonton are Alberta's major metropolitan areas, each with populations in excess of a million. Red Deer and Medicine Hat are or an order of magnitude smaller, each with populations near 100,000.

The method of interpretation is the same as in the comparison of only two cities. First we examine the homogeneity of the data for each city (lower charts). In this case, the data for each city appears homogenous so we are okay to proceed.

LTC Rates of four Alberta cities



Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

Next we look at the data points for each of the four cities to see if any has data points above or below the control limits of any of the others.

In conducting this analysis we observe the following:

- There is a detectable, important difference between Edmonton and Calgary (as already noted).
- Red Deer is almost wildly below the rates for the other three cities with its average line below the lower control limit of any of the other centre's. We don't know why this is so, but we do know it is worth investigating.
- Medicine Hat has detectably lower LTC than Edmonton or Calgary. Again, we don't know why. It is much smaller than either Edmonton or Calgary, but the same size as Red Deer.
- Larger cities have higher, detectable LTC's than smaller cities. Again, this is pattern worth looking into.

The City of Red Deer makes an instructive point. Control chart analysis can tell us if differences in performance are detectable and worth looking into, or not. They cannot tell us the source of these differences. Some possible reasons why Red Deer has such low LTC rates include:

- Red Deer is smaller. Smaller centers seem to have lower LTC rates so this is a contributing factor, but this alone is not sufficient because it doesn't explain the differences between Red Deer and Medicine Hat which are of similar size.
- Red Deer has a better employee safety system than the other centres.
- Environmental conditions are different in Red Deer than in the other centres.
- Red Deer has a different system of filing claims to the WCB, perhaps discouraging claimants or otherwise fudging their data.
- Red Deer performs different functions than the other cities.

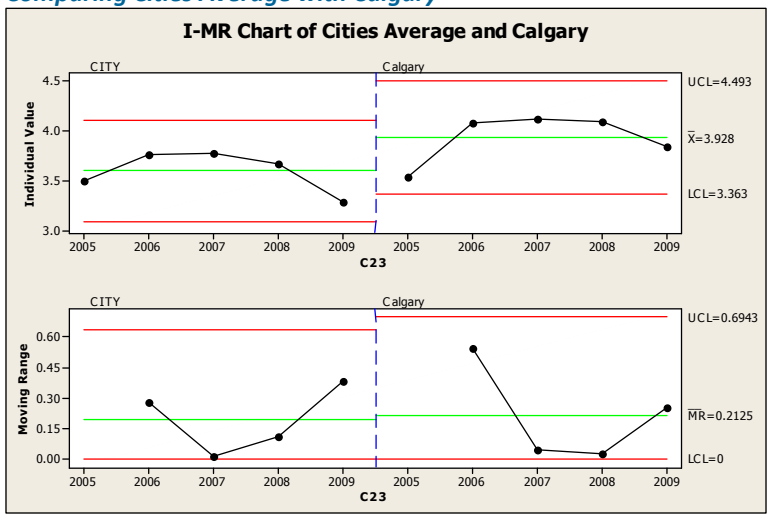
All of these are valid possibilities. Once important differences have been detected, organizations need to follow up with root cause analysis to understand why the differences exist and take rational action to elicit improvement.

Comparisons with the industry average

There is a limit to how many companies we can compare with this control chart technique. Paper width has its limits, as does human capacity for information. Placing twenty control charts alongside one another is bound to be a little confusing (although the confusion can be alleviated somewhat by sorting the order in which the charts appear by the value of the average line for each of the organizations).

So how to compare your organization with the industry average? One approach would be to plot the industry averages on a *XmR* Control Chart and compare this with a chart of plotted values for your organization. In essence, this simplifies the comparison back to two *XmR* Charts, one for industry averages and one for your organization.

Comparing Cities Average with Calgary



Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

Unfortunately, comparisons to the average in this way is an error that can lead to faulty, and at times, disastrous conclusions. Averages, as with all summary statistics, remove variation from the data - variation that is used to calculate control limits. This makes comparing an *XmR* Chart of industry averages, with an *XmR* Chart of a specific organization, the equivalent of comparing apples and oranges.

We are left without any reliable way of evaluating the differences in the data. Comparing a set of values to a

corresponding average, on a control chart (or anywhere else) is an example of *lying with statistics*[4].

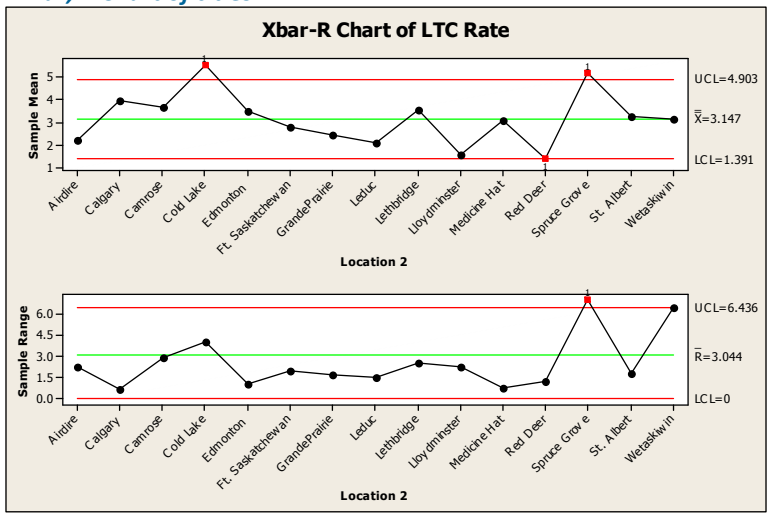
X-Bar, R Analysis

The right way to analyze this data is use a different control chart, the *X-Bar, R* (also called the Averages Chart). This control chart rolls-up the yearly data for each organization in the industry (each organization is a sub-group in control chart terminology) and then compares these cities against the overall average. All of Alberta's thirteen cities are included in this analysis.

Notice that now, instead of years along the bottom axis, we have the various cities that comprise the industry. These cities are now compared to the average and control limit lines to see if there are detectable important differences between each city and the industry average. The basic interpretation approach is similar to that outlined previously.

- First, examine the bottom *R* Chart for homogeneity in the data set. In this case, Spruce Grove displays a lack of homogeneity in their 2005 to 2009 data with a range value that exceeds the upper control limit of the *R* Chart value of 6.436. Wetaskiwin is close to exceeding this upper limit as well. Normally, we would remove Spruce Grove's data at this point and redo our analysis, however, it the data is left in here for exposition purposes.

X-Bar, R Chart of cities



Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

- In the *X-Bar* Chart we see that our industry average of LTC is 3.147. As was the case with the *XmR* Chart, control limit lines are calculated to either side of this average line. We see that Cold Lake and Spruce Grove lie above the upper control limit of 4.903. This means the average performance of each of these cities between 2005 and 2009 is detectably worse than the industry average comparison group. With Spruce Grove we already know not to take this result at face value because of the demonstrated lack of homogeneity. With Cold Lake, however, the high LTC rates are detectably higher than those of the industry, something that should be investigated.
- Likewise, Red Deer is detectably better than the others in the comparisons group with data lying below the 1.391

[4] In honor of, *How to Lie with Statistics*, by Darrell Huff, 1954. The uselessness of comparing single values directly against averages is a point to which more managers need to pay attention. Most benchmarking activities concerning operational performance, health & safety, employee and customer satisfaction, make direct comparisons to industry averages (or the averages of a sub-group such as the top quartile). Such analysis is worse than a waste of time and money because it leads to flawed conclusions that produce management decisions and actions that harm performance rather than enhance it.

lower control limit. Again, a difference that should be investigated.

- For all other cities, their performance as measured by the LTC rate metric is not detectably different from that of the industry overall. As was the case previously, this doesn't mean there are no differences between each city and the industry overall, only that none of the differences are likely to be of any practical or economic importance.

What's with the Spruce Grove data?

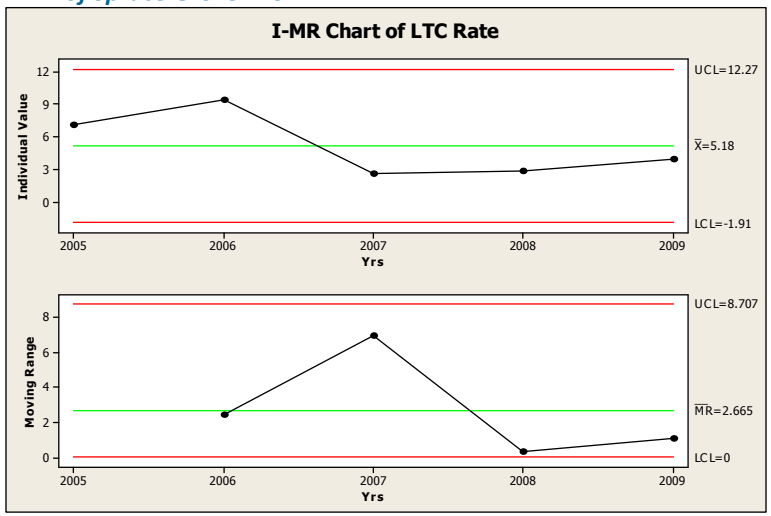
So what about Spruce Grove? The *R* Chart indicated a lack of homogeneity in their data. What does this mean? Well, it means that their data didn't likely come from a single cause and effect system. Something likely changed in the period between 2005 and 2009 that affects their data.

To understand this and examine it more thoroughly, an XmR Chart of Spruce Grove data should be prepared. It would be expected that either the *X* or the *mR* Chart would show an out of control point.

Unfortunately, the data does not display this out of control condition. Why? Recall that we are dealing with only five data points and that this produces some soft limits - limits that will often fail to detect a signal.

Which is why there is no substitute for the *XmR* Chart - outside of the 'statistics' it also presents the data graphically, allowing us to see what the 'statistics' miss. In this case, notice how close 2006 came in the *X* Chart to exceeding the upper control limit and the large drop the following year indicated by the spike in the *R* Chart for 2007. It seems likely that something changed in 2006/2007 that caused a large drop in the LTC Rate for Spruce Grove. Yogi Berra was right, *You can see a lot just by looking*.

XMR of Spruce Grove LTC



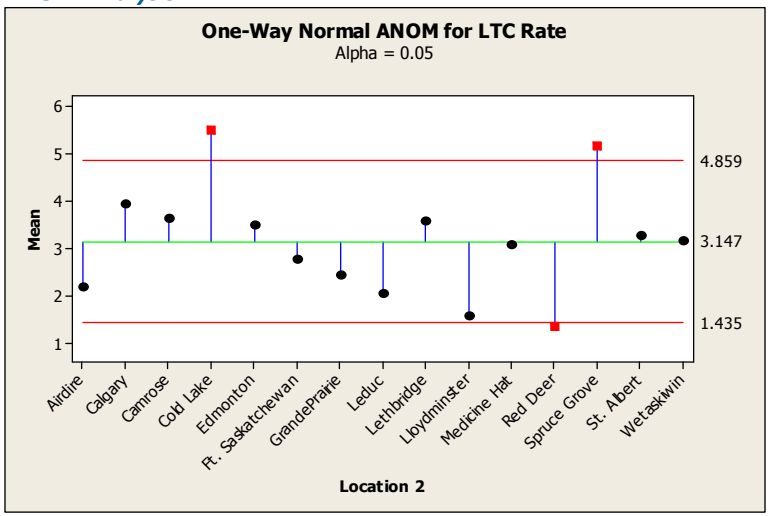
Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

ANOM Analysis

ANOM, Analysis of Means, is an alternate means of analysis that is close in principle to the *X-BAR, R* Chart. ANOM provides an exceptionally easy to understand graphical representation of the differences between a city on the one hand, and the overall industry average on the other.

The downside is that data homogeneity is assumed (that's why there is no lower graph). Before using this chart, therefore, the analyst should have examined the *XmR* chart for each company or organization in the industry average to ensure the assumption of homogeneity holds.

ANOM Analysis



Source: Data: Workers Compensation Board.
Analysis: Converge Consulting Group Inc.

Outside of this, the ANOM Chart has the advantage of:

- an easier to understand graphical representation of the data. The burden of assessing homogeneity falls to the analyst, leaving managers and executives free to focus on the ANOM chart itself and the differences between the industry average (3.147) and the LTC rates for individual organizations,
- allows the analyst to specify an alpha level for the statistical test. The control chart sets the alpha level at practical certainty. If differences are detected in the control chart, you can be certain, for all practical purposes, that the difference and its root causes should be sought out. On the ANOM chart, this sensitivity can be adjusted by the analyst. In the example, the alpha value is set to 0.05 - meaning there is a 95% probability that the difference is worthy of additional investigation. That's why the control limits on the control chart are wider than the control limits on the ANOM chart - practical certainty is harder to get than 95% sure.

Critical Conclusions

The release of injury rate data by the Workers Compensation Board and the Alberta Government has generated a fair amount of interest by organizations, employee groups and the public. Both groups are to be commended for publishing this data - transparency and accountability are impossible without publication.

Publishing data, however, is not without its dangers. The greatest of these is poorly done analysis that leads to flawed conclusions and actions that increase the probability of workplace injuries. Poor analysis would include one or more of:

- superficial analysis of the data that relies on management or expert opinion to judge differences (*our results look too high*),
- analysis that fails to understand the distinction between enumerative and analytic studies and so, applies enumerative methods to what is an analytic problem,
- the use of hypothesis or statistical significance tests to examine differences among time periods, organizations or other groups of data,
- the use of symmetric functions (such as the average or standard deviation) that assume homogeneity, destroy the variation necessary for analysis, and present the data out of context ignoring the time order of the data.

No tool, no methodology, can compete with the control chart (*XmR* & *X-Bar, R*) in the analysis of injury rate data. Effective examination of trends or of other important patterns including

making meaningful comparisons with other organizations (or groups of organizations) simply require its use.

This is true not just for injury rate data, but for any organizational performance data. Benchmarking of:

- operational performance,
- financial performance,
- health and safety performance,
- environmental performance, or
- employee or customer engagement,

without a Control Chart, is at best, a waste of effort. At worst, it is blunt incompetence that will work against the very performance the organization wishes to improve.

Hopefully, this paper will help those interested in reducing injury rates in Alberta to analyze the recently released injury rate data correctly and make meaningful and useful conclusions as a result.

Converge Consulting Group Inc.
www.converge-group.com
403 266 0061

Converge Consulting Group Inc.

Mailing: 2617-20 Street SW
Calgary, Alberta
Canada
T2T 4Z4

Phone: 403.266.0061
Web: www.converge-group.com
www.converge-group.net