Red, Yellow, or Green Light?

Signaling Performance

in the Workers' Compensation Board of British Columbia's Executive Dashboard

by

BRYAN ANDREW READ

B.Sc. (Actuarial), University of Calgary, 1998

B.Comm. (Risk Management and Insurance), University of Calgary, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER of SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Commerce and Business Administration)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 2003

In presenting this thesis in partial fulfilment of the requirements
for an advanced degree at the University of British Columbia, I
agree that the Library shall make it freely available for reference
and study. I further agree that permission for extensive copying of
this thesis for scholarly purposes may be granted by the head of my
department or by his or her representatives. It is understood that
copying or publication of this thesis for financial gain shall not
be allowed without my written permission.

Department of *Commerce and Business Administration*

The University of British Columbia
Vancouver, Canada

Date *March 24, 2003*

ABSTRACT

In 2002 the Workers' Compensation Board of British Columbia's management began a project to move its executive-level Balanced Scorecard metrics into a dashboard-type intranet application. This thesis begins by discussing this improvement in the communication of Balanced Scorecard results. It then explores, in detail, two different approaches to assessing performance. The first approach outlines the most common methods being used to set targets. These targets are measurable goals that the organization strives to achieve as part of its corporate strategy. The second approach involves the use of historical data to assess current performance. This section of the thesis discusses how the most recent value, short-term and long-term trends can be ranked against prior period measurements, and then combined into an aggregate performance assessment for the business metric. This result can then be used to communicate performance that is better or worse than historically experienced. Both the target setting and historical ranking approaches can be used in the dashboard to trigger an indicator light (red, yellow, or green) that signals performance to an executive. A red light signals that the organization needs to do additional work to explain performance or to initiate projects to improve results. Ultimately, it is through both setting targets and ranking performance that the WCB or any other organization will move closer to achieving its strategic objectives.

Key Words: Balanced Scorecard, Dashboard, Key Performance Indicator, Business Metric, Performance Measurement, Performance Assessment, Target Setting, Trend Analysis, and Workers' Compensation Board.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I. INTRODUCTION

In this chapter we describe the Workers' Compensation Board of British Columbia, the problem that it was facing, and the project that forms the core of this thesis. Finally, we review some related research.

1.1 WCB of BC Overview

1.1.1 The Organization

The Workers' Compensation Board of British Columbia (WCB) was established in 1917 to provide compensation to British Columbia's workers for injuries incurred while performing work-related activities. In return for these benefits, injured employees agree not to file lawsuits against their employers. Coverage is provided for almost all of the 2 million workers currently employed in BC, employed by almost 170,000 businesses.

The WCB receives annual premiums from the employers, which is used to cover the claim costs and the administrative expenses of the board. The WCB receives no financial assistance from the BC government; it is a completely self-funded non-profit organization. In 2001, WCB's revenue (premiums and investment income) totaled $1.5 billion, while claim payments and administrative expenses totaled almost $1.8 billion.

Over the past 85 years the WCB's activities have expanded and now include setting and enforcing health and safety standards; assisting injured or disabled workers and their dependants; and assessing and collecting premiums from employers to maintain operations. Its workplace health and safety mandate is to "prevent workplace injuries, diseases, and fatalities; rehabilitate injured workers and return them to productive, safe employment; provide fair compensation for workers suffering from an occupational disease or injury; provide sound financial management for a viable workers' compensation system; and protect the public interest." (WCB of BC 2002)

Looking at the WCB's organizational structure in Appendix A, it can be seen that a Panel of Administrators oversees the Senior Executive Committee, and that the organization is made up of 4 key functional areas: Prevention; Rehabilitation and Compensation Services; Finance and Information Services; and Human Resources and Facilities.

1.1.2 The Organization's Information and Balanced Scorecard

In the early 1990's the WCB created a data warehouse to centrally store a summary of its electronic operational data. This data warehouse contains both raw and summary data for WCB claim and premium activity. Although a primary source, the data warehouse is not the only source of data in the organization. Some of the departments still use other data storage systems that are distinct from the data warehouse; for example, much of the human resource and financial data is in PeopleSoft.

Throughout the organization, many measures of performance (business "metrics") are tracked on a regular basis, using data from the various sources. Prior to the late 1990's there was no formal approach to measuring organizational performance. In 1998 the WCB designed and developed a Balanced Scorecard and over the past few years it has been expanded to include over 50 metrics. The executives use the Balanced Scorecard as a snapshot of the organization's performance on a monthly basis. It contains business metrics from many different parts of the company and is organized into 4 "perspectives": operations, finance, human resources, and customer service. The business metrics chosen in each of these categories are listed in Appendix B. Appendix C, meanwhile, contains a subset of the Balanced Scorecard metrics in a table from the WCB's 2001 Annual Report. The Annual Report communicates all of the most important metrics, providing the 2001 target, the 2001 actual result, and the 2002 target.

The Balanced Scorecard is a monthly paper document that is used by the Senior Executive Committee and the Panel of Administrators to monitor the organization's performance. The Balanced Scorecard measures business factors that are directly linked to the organization's strategic plan. Thus, the senior executive committee uses the Balanced Scorecard to determine if their corporate strategies are succeeding or failing. Because it is so important to monitoring the health of the organization, the Balanced Scorecard is an evolving tool and the WCB encourages ongoing improvements.

1.2 The Problem and the Project

1.2.1 Problem Definition

As with many businesses, the WCB is committed to researching and developing new tools that will improve the effectiveness of its executive level decision-making. These tools are important because

given that the organization manages over $1.5 billion in claims annually even small improvements at the top can result in significant benefits.

There are reams of data at the WCB, flowing from numerous organizational processes. This data is continually being synthesized into business metrics by the WCB's information systems and staff. Of these metrics, only a select group is used in the current WCB Balanced Scorecard. The executives use these metrics (the "Key Performance Indicators" or "KPIs") to monitor the organization's health and to make key strategic decisions.

However, the existing Balanced Scorecard was simply a paper document with a set of charts, containing no analysis, no conclusions, and no information about how the organization was working toward better results. It was obvious to the WCB that the communication and use of the Balanced Scorecard required improvement.

In addition to this, the performance of most of the KPIs in the Balanced Scorecard was not being assessed. Without a target or a comparison to historical data, there was no assessment of performance. Furthermore, without an assessment of performance, there was no clear understanding of what the organization was doing well and what needed improvement.

Even when a target was chosen for a particular KPI, it was not clear how this was done. Was the target an arbitrary value, simply agreed to by a handful of people in a meeting somewhere? Was it understood and approved at the executive level? Was there buy-in from those directly impacted by or responsible for the results? Was the approach and information used to choose the target credible? Was the target really a "make it or break it" value, or should there be some in-between level of performance? It was clear that the Balanced Scorecard needed to be improved.

1.2.2 The Project

It was determined that the WCB needed a centralized view of the Balanced Scorecard's Key Performance Indicators (KPI) in a "dashboard" type application and within their "intranet" environment (an internal internet). The purpose of the tool was to effectively communicate the organization's key results, including an assessment of performance.

Put another way, and with regards to the structure of this thesis, the project sought to:

1) Improve the communication vehicle of the existing Balanced Scorecard, ultimately allowing the executives to make more effective business decisions. (See Chapter II)

2) Create a flexible and structured methodology for setting KPI targets. (See Chapter III)

3) Create a performance evaluation and signaling process using the KPI data. (See Chapters IV and V)

Understanding where the business has been and where it is going is the responsibility of everyone in the organization. But, it is the executives who choose the direction of the company through the creation of corporate strategies and policies. By assessing performance effectively, the leaders of the organization can react faster to the ever-changing business environment.

1.2.3 The Team

Designing and building the dashboard application for the Workers' Compensation Board required a cross-functional team, including information technology, business, and academic representatives. In the initial design of the application, not all of the business areas were represented; however, in the future phases of the project all key departments will be involved, and the application will be adapted accordingly.

The "dashboard project" was conceived and sponsored by Sid Fattedad, the Vice-President of Finance and Information Services at the Workers' Compensation Board. The business analyst was Ernest Urbanovich, and the Information Technology support was provided by team leaders Sandi McConach and Fernando de Nobrega, and programmers Nitin Gandhi, Rob Sclater, and Thomas Erl. In addition, Martin Puterman, a UBC Commerce professor of Operations and Logistics, and myself, Bryan Read, a UBC M.Sc. student, provided research and development support. The primary members of the team were Sandi McConach, Nitin Gandhi, Ernest Urbanovich, and myself.

1.2.4 Significance of this Research

The purpose of this thesis is to communicate the method and results of our project with the Workers' Compensation Board of British Columbia, plus a proposal for performance assessment. It is hoped that this thesis will contribute to the growing body of research related to the Balanced Scorecard. Specifically, this thesis should prove useful to anyone searching for a reasonably objective method for

assessing the performance of Balanced Scorecard KPI results, relative to either a target or the historical data.

The objectives of this applied research thesis are:

- *To examine* the alternatives and considerations involved in assessing the performance of complex business metrics.
- *To educate* the reader on the intricacies related to choosing performance criteria.
- *To propose* a methodology for performance assessment that can be used in the design of a Balanced Scorecard or dashboard application.

By applying the Balanced Scorecard and dashboard methodologies to the WCB project we were able to design a highly effective executive application for the WCB Balanced Scorecard. There are a number of benefits associated with our application design over the available commercial products and the general Balanced Scorecard theory. Specifically, our application has the added flexibility of being able to add comments to explain performance and improvement initiatives. In addition to this, instead of forcing the organization to choose the targets against which performance will be measured, we have created a process for ranking performance using just the historical data. This primarily objective approach can be automated and can provide a performance assessment of current values relative to the past values in the series. It takes into account not only the most recent value but also the short and long-term trends. It is not known if this method is currently being practiced by other organizations, but it is a useful and effective method for triggering a performance indicator light and is an alternative to setting targets.

Corporate metrics are not effective if poor results are not distinguished from good ones. A consistent and structured approach is necessary to ensure a useful performance assessment is being made. Whether such an assessment involves setting targets or using historical data, a formal process for evaluating metric values must exist. At present, this formalization is not a significant part of the Balanced Scorecard Methodology, and hence the value of this thesis is that we present just such a structured approach to assessing performance.

1.3 Literature Review

1.3.1 Balanced Scorecards

In 1992 a Harvard Business School professor and a business consultant jointly published an article recommending a different approach to measuring and managing an organization's performance. This new concept was named "The Balanced Scorecard" (Kaplan and Norton 1992), and since that initial article there has been a significant volume of research on this topic, particularly by the original authors.

The Balanced Scorecard is used by an organization to translate corporate strategies into clear and measurable goals. The Balanced Scorecard provides a comprehensive picture of the business operations and facilitates the communication of the business' strategic objectives (Kaplan and Norton 1996).

Typically, a Balanced Scorecard has 4 or 5 "perspectives" (or groupings). The most common ones are financial, customer, operational, and innovation and learning. These perspectives support the core philosophy behind the Balanced Scorecard: in assessing performance, organizations should balance their short-term targets (ROE, Net Income, etc.) with their long-term goals (customer satisfaction, employee skills, product quality, etc.). (Kaplan and Norton 1996)

Each perspective is a collection of two to five key business metrics. These metrics are known as the "Key Performance Indicators" or "KPIs", and their values represent the health of the organization. For each KPI, a "Performance Indicator Light" denotes performance. The Balanced Scorecard methodology suggests using traffic light colors to communicate three levels of performance: good-green, average-yellow, poor-red. (Kaplan and Norton 1996)

Despite developing a very thorough approach to "balancing" the assessment of performance in the organization and tying performance metrics to strategy, very little of Kaplan and Norton's work addresses how to actually assess performance. In their 1996 book there are only a few pages dedicated to setting targets. They do suggest that the organization should set "stretch targets" that are challenging, but achievable with significant effort. Stretch targets are usually the long-term plans (3-5 years) from which annual targets are built. How exactly to choose the stretch targets is vague, however, simply recommending the use of benchmarks and getting the buy-in of those accountable for the results (Kaplan and Norton 1996).

The Balanced Scorecard concept has been implemented by many organizations, including early adopters such as Mobil, Cigna, and Brown and Root (Kaplan and Norton 1996). More recently the Workers' Compensation Board of Saskatchewan (WCB Saskatchewan 2002), BC Ferries (BC Ferries 2001), the Insurance Corporation of British Columbia (Henrich 2000), Minnesota Department of Transport (Minnesota DOT 2002), Nova Scotia Power, Inc. (Niven 2002), and, as of 2003, every department in the US federal government (Ballard 2002) have implemented some form of Balanced Scorecard. In fact, according to a 2001 survey by the UK firm Business Intelligence, 57% of 200 companies from more than 20 countries had adopted a Balanced Scorecard (Miyake 2002).

In the early 1990's the Balanced Scorecard methodology was applied to the WCB to help the Board of Directors monitor the overall effectiveness of the organization. In late 1997, the scorecard was adopted by the WCB's Senior Executive Committee for use throughout the WCB as the standard for organizational measurement. In 1998 the Royal Commission on Workers' Compensation in BC reviewed the scorecard and its metrics. This report made recommendations for the improvement of the WCBs Balanced Scorecard, including the addition of several KPIs, the need for improved data collection, the inclusion of definitions and analysis, the incorporation of trends, and the establishment of baselines and targets (Macfarlane and Weltz 1998). It is the last two of these recommendations that this thesis addresses.

1.3.2 Dashboard Applications

Dashboards are computer-based applications that present business metrics in a central and easy to access way. Typically, these applications display values in tables, graphs, charts, and dials. A big advantage of the dashboard application is that it is a central source for performance information. General Motors uses a dashboard with performance lights to monitor the progress of its ongoing projects (Mayor 2001).

Dashboards can also be used to manage and communicate Balanced Scorecard metrics. There are now more than 30 software vendors offering Balanced Scorecard dashboard applications, including Cognos, Crystal Decision, Oracle, Peoplesoft, and SAP (Marr and May 2002). These applications range from generic software applications to highly customized ones where a team of consultants will tailor a

product specifically for an organization. Produced by QPR, "ScoreCard" is one such example of a dashboard software package that utilizes the Balanced Scorecard methodology (QPR 2002).

Many of these software vendors are "Balanced Scorecard Certified" by the Balanced Scorecard Collaborative (BSCol). The BSCol has established a set of functional standards that applications must conform to in order to be certified. The mandate of this organization is "to facilitate consistent and appropriate use of the Balanced Scorecard globally" through the "harmonization and standardization of the methodology of the Balanced Scorecard as envisioned by the creators of the concept, Dr. Robert Kaplan and Dr. David Norton." (Balanced Scorecard Collaborative 2002)

1.3.3 Performance Assessment and Target Setting

Organizations have been setting financial targets for many years. Financial planning and budgeting occur at least annually in most organizations, and so it is a well-established process. In 2001, a team from the Cranfield School of Management and Accenture in the UK did a review of worldwide planning and budgeting practices. Their report outlines what they believe to be the best practices in planning and budgeting. Their conclusions include "forecasts should be assumption based not opinion based", "the focus is the competition not the budget", "reviews should be action- not explanation-oriented", and "organizations should be strategically managed not financially managed." In addition, they go on to give cases of organizations that use benchmarking and forecasting as part of their planning process (Neely, Sutcliff and May 2002).

Benchmarking and forecasting are used extensively by many organizations when converting their strategic goals to measurable targets. The Minnesota Department of Transport in the US is just such an organization, recently producing a 40-page report outlining their targets and target setting practices for 2003 (Minnesota DOT 2002). There are also numerous organizations that exist to provide benchmarking analysis. The Housing Corporation, a provider and regulator of social housing in the UK, has done a significant amount of research on benchmarking and uses it to promote continuous improvement in the construction industry (Housing Corporation 2002). It has even started a "Benchmarking Club" (Housing Forum 2002). Related to Workers' Compensation, one of the mandates of the AWCBC's (Association of Workers' Compensation Boards of Canada) is to gather data from all of the Canadian WCB's and provide benchmarking services (AWCBC 2002).

Research into qualitative forecasting is ongoing, and although many methods exist to project future values, there are always events or influences that are not expected. This thesis builds on only the simplest methods of forecasting that can be found in any introduction to a forecasting textbook (Makridakis. Wheelwright and Hyndman 1998).

1.3.4 Trend Analysis and Signaling

Financial analysts and stock market analysts use a number of methods to analyze and signal trends, but they tend to be more interested in prediction than assessment. Financial analysis typically includes trend line techniques, curved trend channels, oscillator techniques, and moving averages (Sklarew 1982). There are software products, such as WizeTrade (WizeTrade 2002) and NeoTicker EOD (TickQuest 2002) that utilize these methods to provide real time analysis on investments. These financial techniques are useful for metrics like stock prices that are recorded daily at a minimum, but are less effective for less frequently measured metrics like monthly KPIs. Nonetheless, some of the techniques were explored and are transferable, although most did not influence this thesis.

Unfortunately, we were unable to find published references to the Historical Ranking Process developed in the second half of this thesis. Specifically, there were no references to using percentile calculations to rank current values against prior values in a time series as a method of assessing performance. This could be due to the focus of most financial techniques on predicting values rather than making an assessment of the current performance. We believe that the Historical Ranking Process is new.

II. DASHBOARD DESIGN

Building on the existing Balanced Scorecard metrics implemented and tracked since 1998, the team identified the weaknesses within the existing system and developed a process that is believed to be superior to the existing one. The Executive Committee and all of the stakeholders in the organization will significantly benefit from these improvements. Here in Chapter II we identify the opportunities that existed, outline the new process flows, define the role of performance indicators, and discuss the dashboard application that was developed.

2.1 Opportunities for Improvement

As mentioned in the introduction and illustrated in Figure 1, the existing Balanced Scorecard system at the WCB involves numerous departments extracting data from different sources and compiling, summarizing and charting the data (using MS Access, Crystal Reports, and MS Excel). These charts are then brought together by one employee, printed, replicated, and finally, distributed to the executives. An electronic copy is also made. This copy is posted in the WCB's intranet as a PowerPoint file. The report is produced monthly, and as of June 2002 it included approximately 50 metrics.

The opportunities for improvement in the current system are six-fold. The existing process would improve with *centralization*, reducing the significant *resources* required, and shortening the long *turnaround* time. The *accuracy and consistency* of the reports are difficult to verify, there is currently no *accountability* for Key Performance Indicator results, and there is no clear *performance communication*. The new system will be web based, mostly automated, and interactive, while addressing all of these issues.

Centralization of Information: Currently, not all of the data is available from one central source, or warehouse and so neither are the charts and graphs. Progress has been made since the early 1990's, and it is a long-term corporate goal to centralize all of the data used to track the key business metrics. With the dashboard all of the charts will be automatically produced by the system, and then displayed together through the intranet (Figure 1).

Resource Commitment: The manual calculation of Key Performance Indicators requires significant human resources to extract data, make the appropriate calculations, and then update the charts.

10

Centralizing the data and automating the calculations and charting will significantly reduce the human intervention.
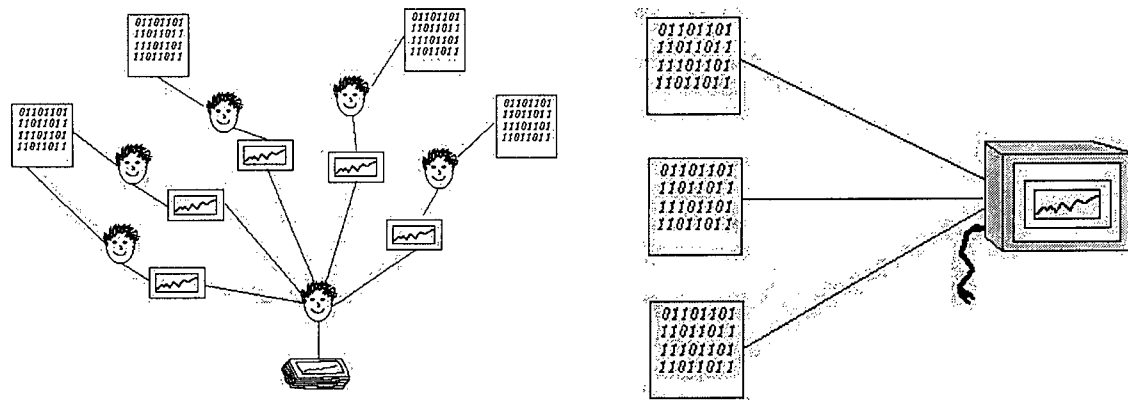
Figure 1 - Current Process and New Process Structure

Turnaround Time: The current process takes between 2 and 3 weeks each month to prepare. With an automated system, the monthly results will be available overnight. With the utilization of today's technology, information is available significantly faster, allowing for more timely decisions.

Accuracy and Consistency of Reporting: Because many different people across the organization create the charts accuracy and consistency are almost impossible to monitor. By automating the process accuracy and consistency should improve, given that the automated reporting method will be approved and validated by the appropriate owners.

Accountability for Performance: Under the current system not all of the KPIs are "owned". Under the new system every KPI will have an owner. This person will be responsible for communicating all changes in the performance of the KPI, providing an analysis of what is driving and/or influencing the results, and detailing the initiatives that are ongoing in the organization to improve performance.

Performance Communication: In the past, KPI communication was simply through the use of graphs and charts. Some KPIs had target performance levels, while others did not. For those with targets, simple calculations were used to determine the percentage deviation from target. However, no assessment was made whether the KPI was exceeding expectations, within acceptable limits, or failing

11

to perform adequately. Performance assessment is a key principle of the Balanced Scorecard methodology, and it is discussed later in this chapter and then in-depth in the rest of the thesis.

2.2 Process Flow Design

In this subsection there are three flow diagrams, the first of which represents the new flow of information at the WCB. The second and third diagrams illustrate the owners' and executives' interaction with the dashboard, respectively. The flow diagrams were created using elementary flow chart elements.

2.2.1 Information Flow Diagram

The information flow diagram below (Figure 2a) maps the flow of the data from the warehouse to the charts and, finally, the performance indicator lights. This flow diagram is for the approach detailed in Chapter IV and V - using historical values to assess performance. The diagram begins in the same way that the process has always begun; utilizing data from the data warehouse has always involved data extraction, metric calculation and metric charting. In many cases the 12-month moving average (12MMA), or some other smoothing technique has also been used prior to charting. Our new system branches off from the old one, however, and now involves the calculation of measures (or perspectives) on the smoothed metric data.
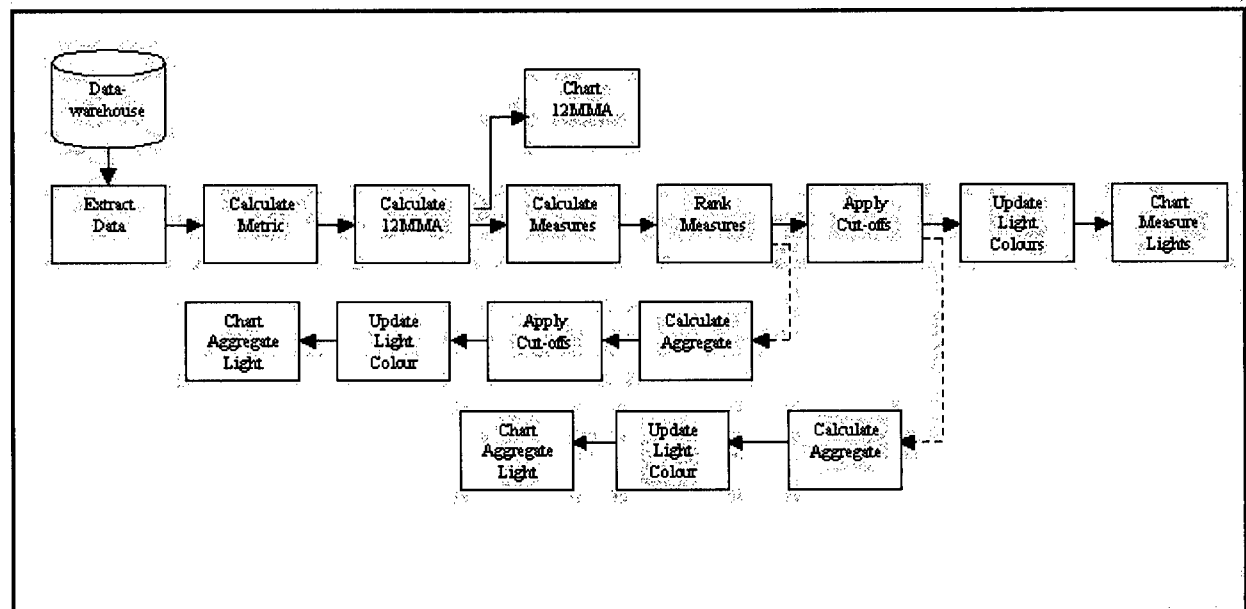
Figure 2a - Information Flow Diagram for the Historical Ranking Process

These measures are then ranked, using a percentile method against all previous measurements at all previous periods. This results in a set of scores in [0,100], against which we can apply cutoffs to determine a set of performance indicators or lights (red, yellow, green), and we can calculate an aggregate value. Applying a cutoff to the aggregate value results in an aggregate performance light (red, yellow, green). All of these lights can be charted over time to illustrate how performance has evolved.

If a performance light is triggered by a comparison to a target instead of ranking current performance against past values, then the flow chart is much simpler. After the "Calculate 12MMA" step, the actual value is just compared to the target and then the performance light is set accordingly.

2.2.2 Owner's Update Process Diagram

Each KPI is assigned an owner, who may or may not be an executive. This person is responsible for analyzing the latest results posted for the KPI, updating the analysis in the dashboard, and managing the initiative updates related to the KPI (projects to improve performance). The second diagram, Figure 2b, maps the newly created owners' process.
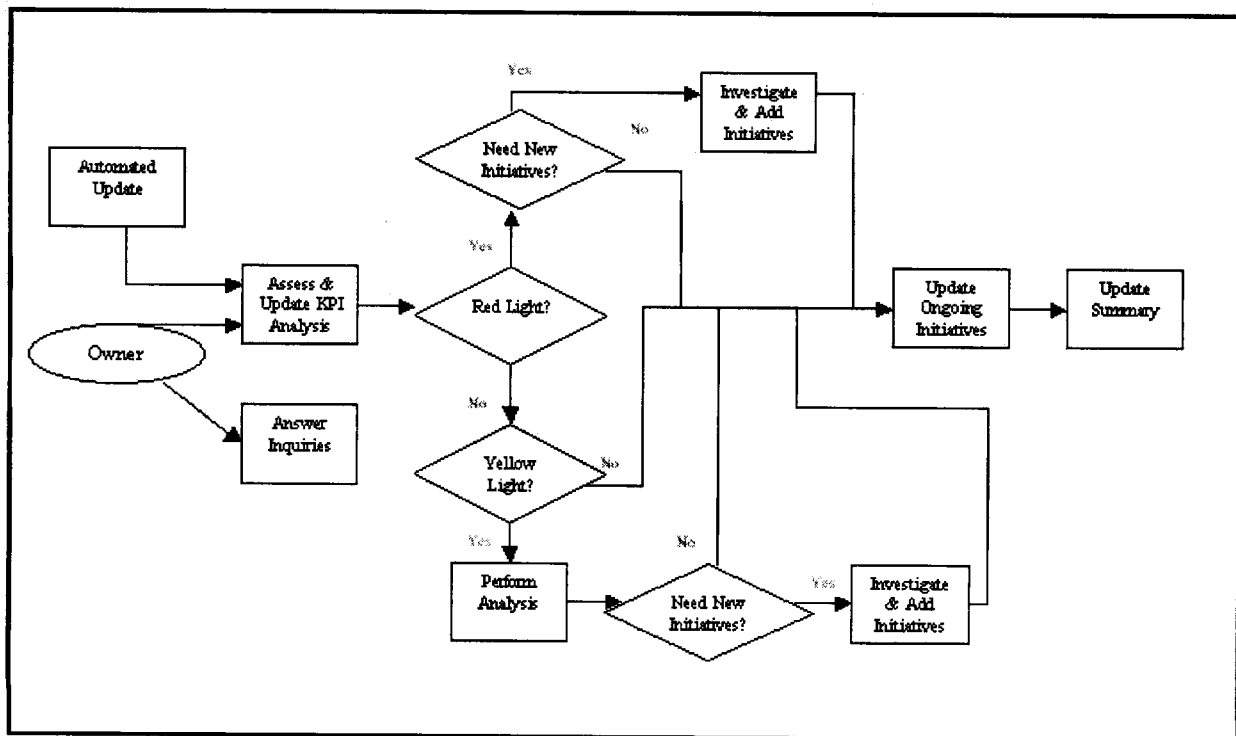
Figure 2b – Owner's Update Process Diagram

13

The process begins with the owner and the automated update of the KPI in the dashboard (data extractions and calculations are set to run automatically each month). Upon receiving email notification of the update, the owner performs and updates the written analysis in the dashboard. If the overall indicator light is red, the owner must decide if new initiatives are needed. If so, the owner investigates and adds new initiatives. If the light is yellow, further analysis may be required to determine if the yellow light is likely to turn red in the near future, or if it is just a natural fluctuation. If performance is determined to be deteriorating, new initiatives may be required to deter a future negative result.

At the corporate level, resources for new initiatives will typically be prioritized to improve those KPIs that are red before those that are yellow (although the importance of the KPI relative to the other KPIs will be considered too). In the ideal world there would be no red lights. All KPIs with a yellow indicator would be dealt with promptly enough so that they would never reach the red status. In fact, the Historical Ranking Process detailed in Chapter IV reinforces this continuous improvement process, requiring continuous improvement for the indicator light to remain green.

The owner's process always ends with the owner updating the existing initiatives and the summary. Finally, at all times the owner is expected to answer enquiries about the KPI from the executives in the organization. These enquiries may include questions about the calculations, assumptions, analysis, indicators, or initiatives.

2.2.3 Executive's Investigation Process Diagram

The third flow diagram, Figure 2c below, illustrates the executives' process. This process has been designed with "exception management" in mind. The executive (usually a Vice President of the organization) has limited time, and ideally he/she should only have to spend time on the areas of the organization that require guidance, not the areas that are doing well on their own. The executive, once notified of the update, would examine those KPIs that are red. Starting with these worst performing metrics, the executive can read the summary written by the owner. If the executive is not satisfied with what the organization is doing to improve the performance or needs more information, then he/she can look at the more detailed screens: drivers, inter-related factors, drill down, and initiatives. If the executive is still not satisfied, he/she can email the owner to address any issues or questions.

This cycle continues for each and every red KPI relevant to the executive, until all have been dealt with. After all of the reds, the executive can do the same for the yellows. The greens, however, do not need to be explored at every update and can be reviewed less frequently.
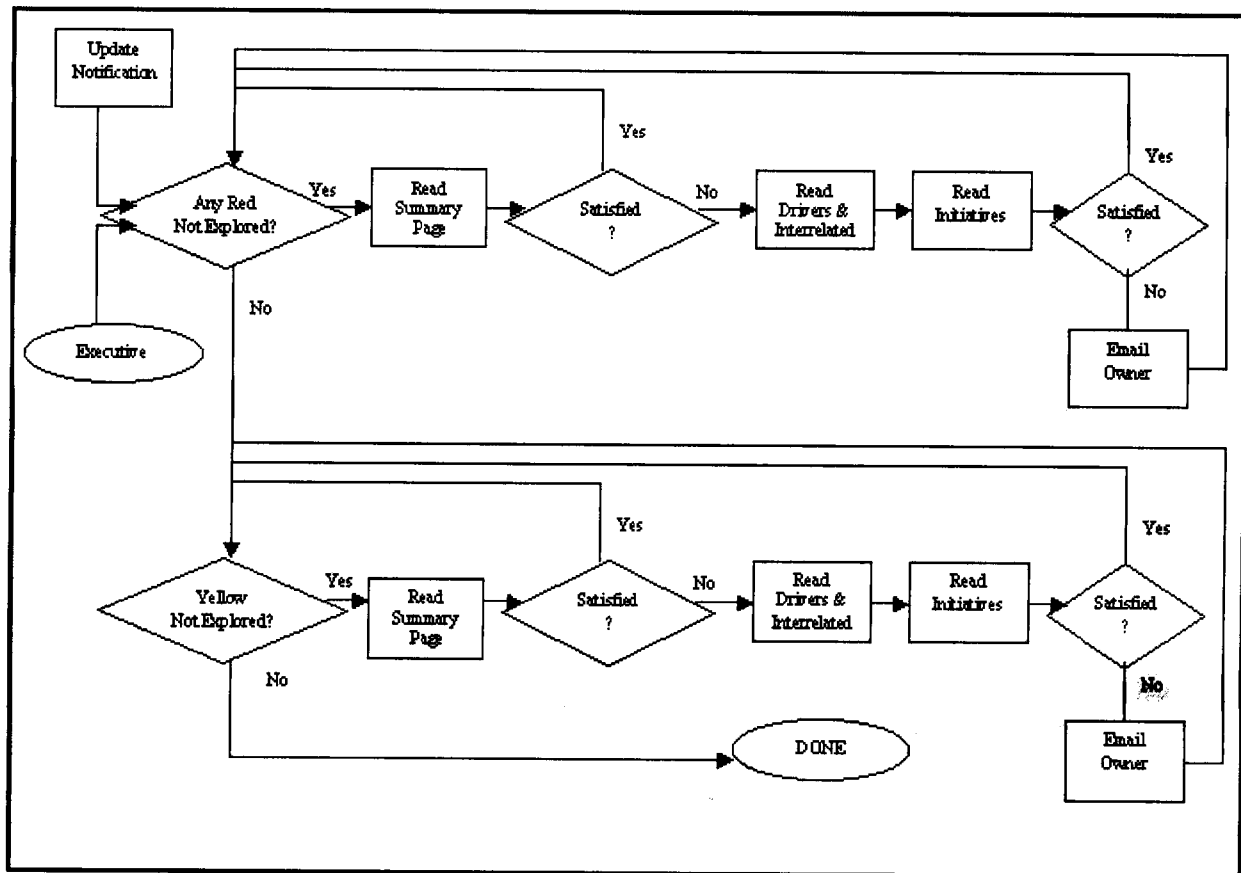
Figure 2c – Executive's Process Diagram

2.3 Performance Measure Indicators

An indicator is "a signal for attracting attention", or "a device for showing the operating condition of some system" (Wordnet Dictionary 1997). Performance indicators in business provide a simple representation of the effectiveness of a complex process. They allow for exception management, as executives can selectively target only those areas of the organization that require improvement. Areas that are performing well can be given less attention, thus resulting in a more effective and focused management approach.

The Balanced Scorecard methodology suggests using a traffic light indicator for each of the businesses' KPIs. If the light is red, the performance is poor. In other words, the organization is either significantly off target from its monthly objective, unlikely to catch up to its targets by year-end, or the performance is much poorer than expected given historical values. Green has the opposite interpretation of red. Green communicates that the performance is good, that the organizational monthly target is being exceeded, or that the organization is on track to exceed its year-end targets. Green may also mean that the organization is doing much better than it has generally done in the past. Finally, a yellow light is used for the "in between"; performance is not poor, but it is also not good. This may mean that the value is within some tolerable deviation from plan, or is an average result, historically speaking. However, the yellow light can also serve as a warning that performance is currently off track and is unsatisfactory, but that it can be amended given analysis, understanding and early intervention.

The reaction of the organization to each of these indicator lights depends on the color. A red light requires immediate attention, and initiatives should be developed to get the company back on track for this dimension. If initiatives already exist for the KPI, sufficient time needs to be given for them to take effect and improve the results. In the case that a target has been set, the red light most likely suggests that it is unrealistic that the annual target will be achieved by year-end, and so the organizational focus will shift to improving results rather than achieving the target.

The yellow KPI indicator requires either that the KPI be closely monitored for the deterioration of results or that an analysis needs to be done to determine if the condition causing the yellow light is trending better or worse. If the KPI is expected to turn red in the near future without intervention, corrective action or new initiatives need to be created to improve the results.

A green indicator does not require an action. After the reds and yellows have been dealt with, if there is time, the executive will review the greens. The only action that might occur is the rewarding of the owners in the form of year-end bonuses or recognition.

It should be noted that the above is a simplification of the real world. Choosing, designing and implementing initiatives is not an easy task. For many KPIs, reactions to poor performance are limited, partially by the nature of the measure, partially by a lack of understanding, and partially by laws, rules

and regulations. It takes time to implement programs, and it takes time for the programs to impact the results. However, as the business learns and innovates, it develops more methods of dealing with poor performance because, despite limitations, the organization cannot allow poor performance to get worse.

2.4 Intranet Application Design

In this section we discuss the structure of the dashboard application that was built for the WCB by our cross-functional team. By describing each of the primary screens, it should become clear how the aforementioned improvements and process-flows tie in.

As can be seen in the initial screen below, Figure 3a, the user is first greeted by the 'Executive View' screen, which contains a count of the red, yellow and green KPIs for each of the 4 business perspectives (financial, human resources, operational, and customer). In addition, the Vision and Mission of the organization are displayed, as they are integral to the Balanced Scorecard methodology.

Corporate Scorecard

Decision NET

| Executive View | Financial | H.R. | Operations | Customer |

Vision
Workers and workplaces are safe and secure from injury, illness, and disease.

Mission
The WCB mission is to:
- Prevent workplace injury and disease
- Provide adjudication of benefit entitlements
- Facilitate recovery and safe return-to-work
- Efficiently deliver service to workers and employers
- Effectively maintain a financially sound, sustainable system

Balanced Scorecard - Overview

Financial Perspective				Human Resource Perspective			
Total No.	⊗	–	✓	Total No.	⊗	–	✓
3	2	0	1	4	1	1	2

Operational Perspective				Customer Perspective			
Total No.	⊗	–	✓	Total No.	⊗	–	✓
6	3	1	2	4	1	2	1

Worker's Compensation Board of B.C's corporate balanced scorecard. See Scorecard Overview, for info on why the scorecard has been deployed.

Figure 3a – The Executive View

17

From this screen, the user can choose to take a closer look at any one of the business perspectives. For example, if the 'Operations Perspective' is chosen (Figure 3b), the next screen contains a list of the KPIs for that perspective, along with each organizational strategy and performance status (red, yellow, or green). Because of the number of KPIs used by the WCB, these metrics are naturally grouped into 'KPI Categories', thus simplifying understanding and discussion. The general rule applied here to create the category indicator is that if at least one KPI is red, then the category indicator is turned red. Otherwise, if at least 2 are green, the category light is turned green. In all other situations the light will be yellow.
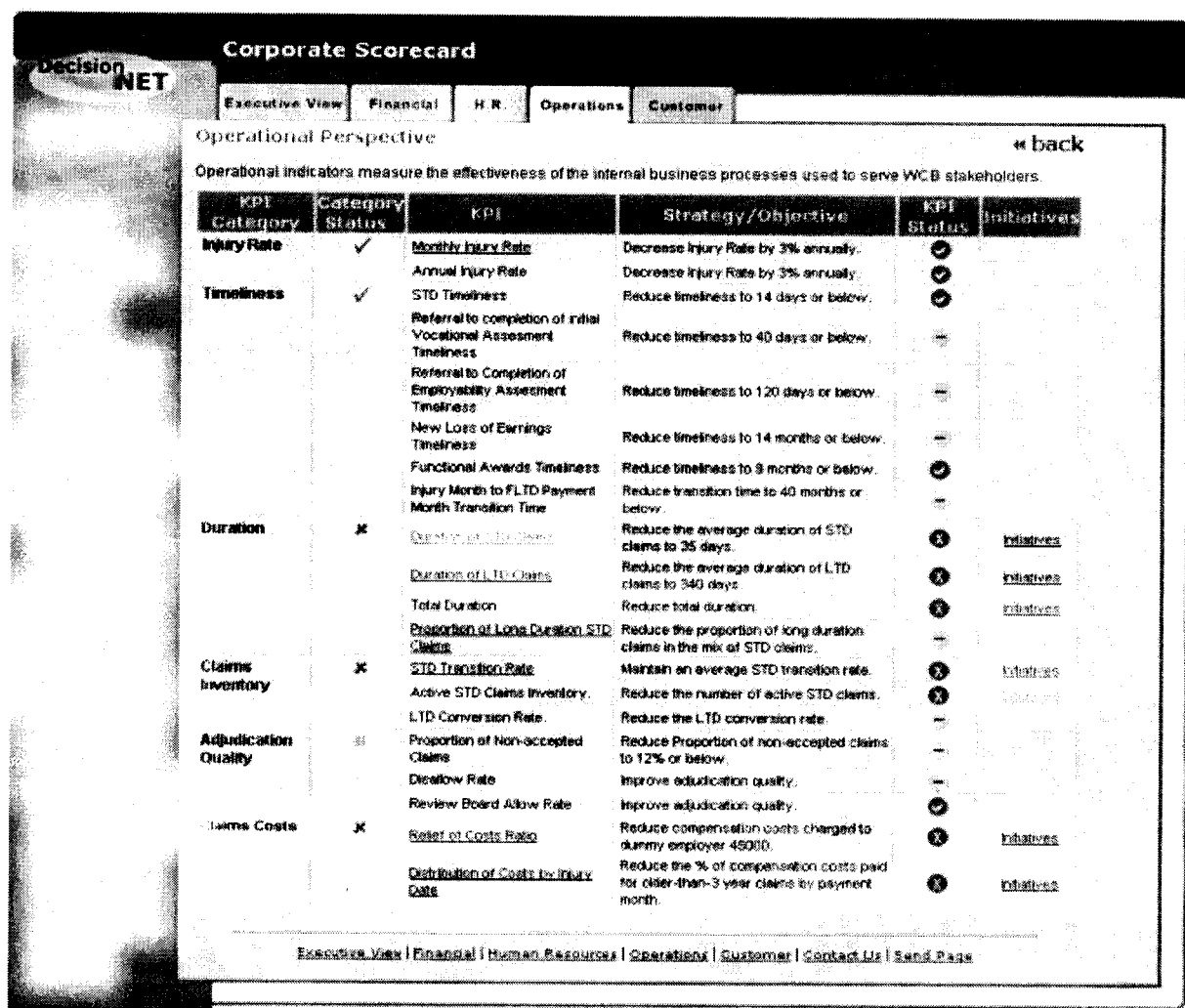
Figure 3b – The Operations Perspective

This 'Operations Perspective' page allows the user to link to any specific KPI for more information. So, if the "Duration of Short Term Disability (STD) Claims" is chosen, then by clicking on the appropriate link, the "Duration of STD Claims" page appears (Figure 3c). At the top of this screen is a definition of the KPI: "the average number of days paid out per short-term disability claim". There is also a synopsis of what is driving the performance of this metric.
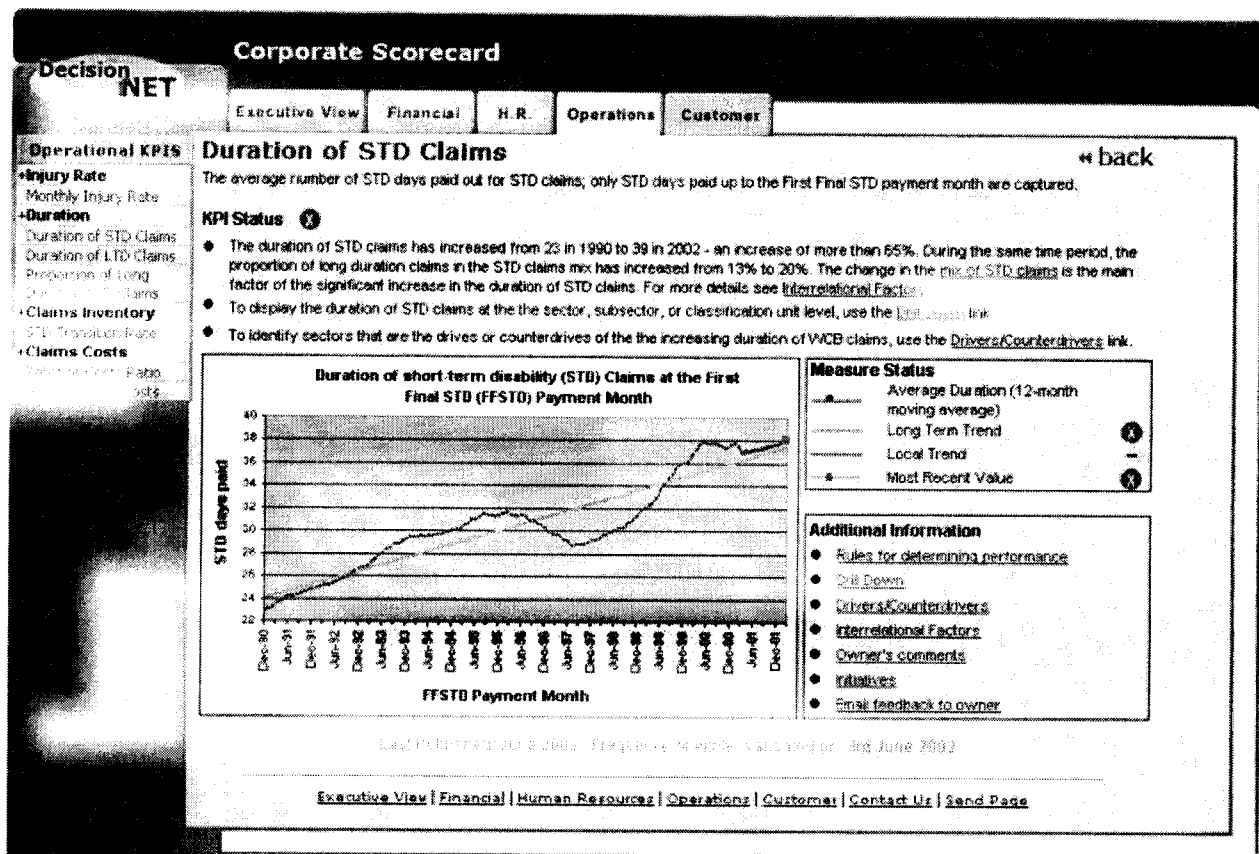
Figure 3c – The KPI Detail View: Duration of STD Claims

At the center of this page is the charted historical data that is automatically retrieved from the data warehouse. This representation can be any sort of chart or graph that effectively communicates the historical information. The WCB's data warehouse contains historical information back to 1987 for the operational metrics. This wealth of data can now be effectively utilized in KPI evaluation, as detailed later in Chapter IV. It is this automated process of performance evaluation that determines the colors of the performance lights on the right hand side. (In the prototype screen, there are only 2 lights for

19

the trends, one for the 12-month and one for the global trend. Chapter IV recommends also including a 5-year trend performance indicator when the data volume permits.)

Also on the 'KPI Detail View' screen is a set of links to more information. These include calculation and performance formulas and assumptions, owner's comments, and the ability to directly email the owner. In addition, there is a screen specifically devoted to outlining the current projects and initiatives in the organization that are meant to improve this metric's performance. The ability to drill down or slice and dice the metric data is also available, along with a screen outlining the driving factors behind the performance (subsets of the original data), and a screen communicating the relationship of the performance of this metric to other metrics in the organization (outside the original data). All of this additional information can easily be accessed from the main KPI screen.
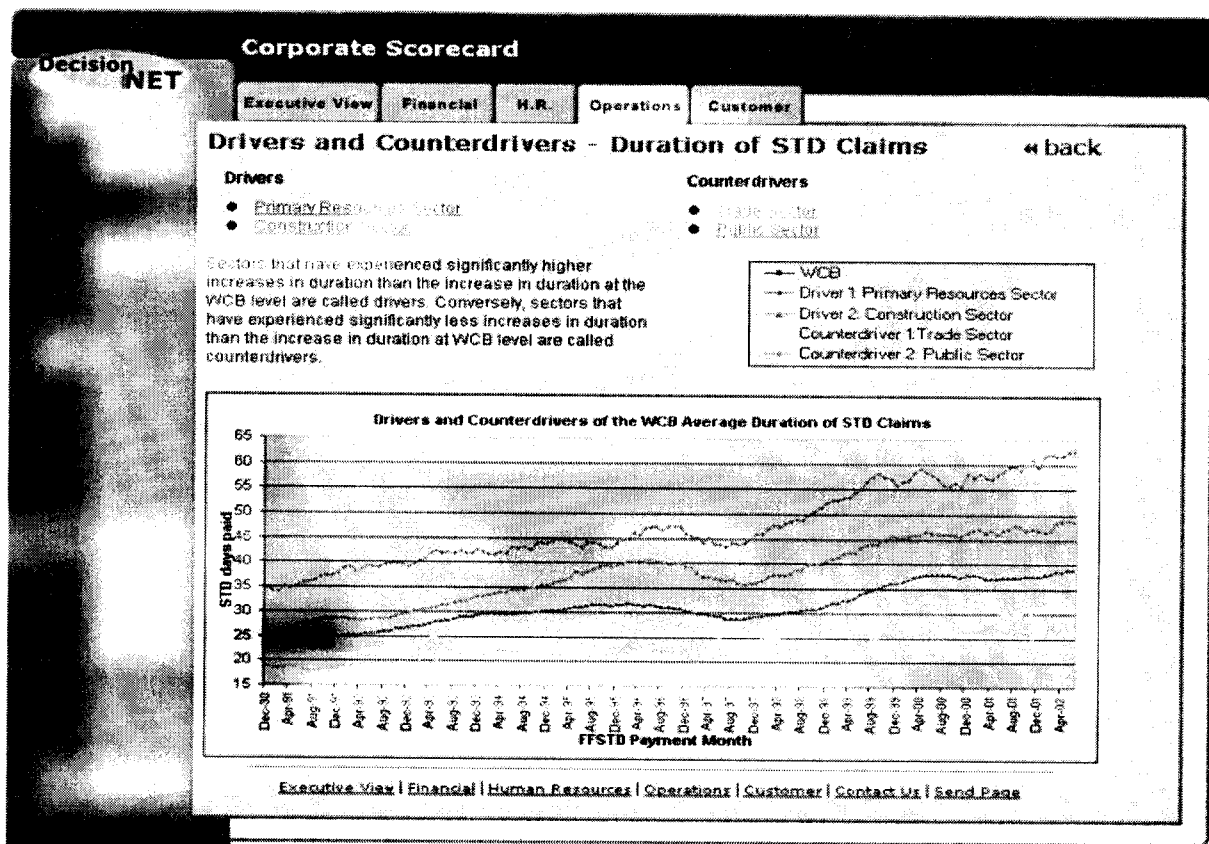
Figure 3d – The Drivers: Duration of STD Claims

Finally, looking at the "drivers" and "inter-relational factors" screens, we can see the details behind the poor performance (the red light) of this KPI. For the Short Term Disability claim duration, the

organizational level data can be broken into subsets, such as by industry. The values from each of these groups are either higher or lower than the aggregate value. In this example, those that are significantly higher are drivers of the performance, while those significantly lower are considered counter-drivers. So, the conclusion here is that the primary resources and construction sectors are driving the poor performance of this KPI, while the trade and public sectors are doing better than the norm (Figures 3d and 3e).
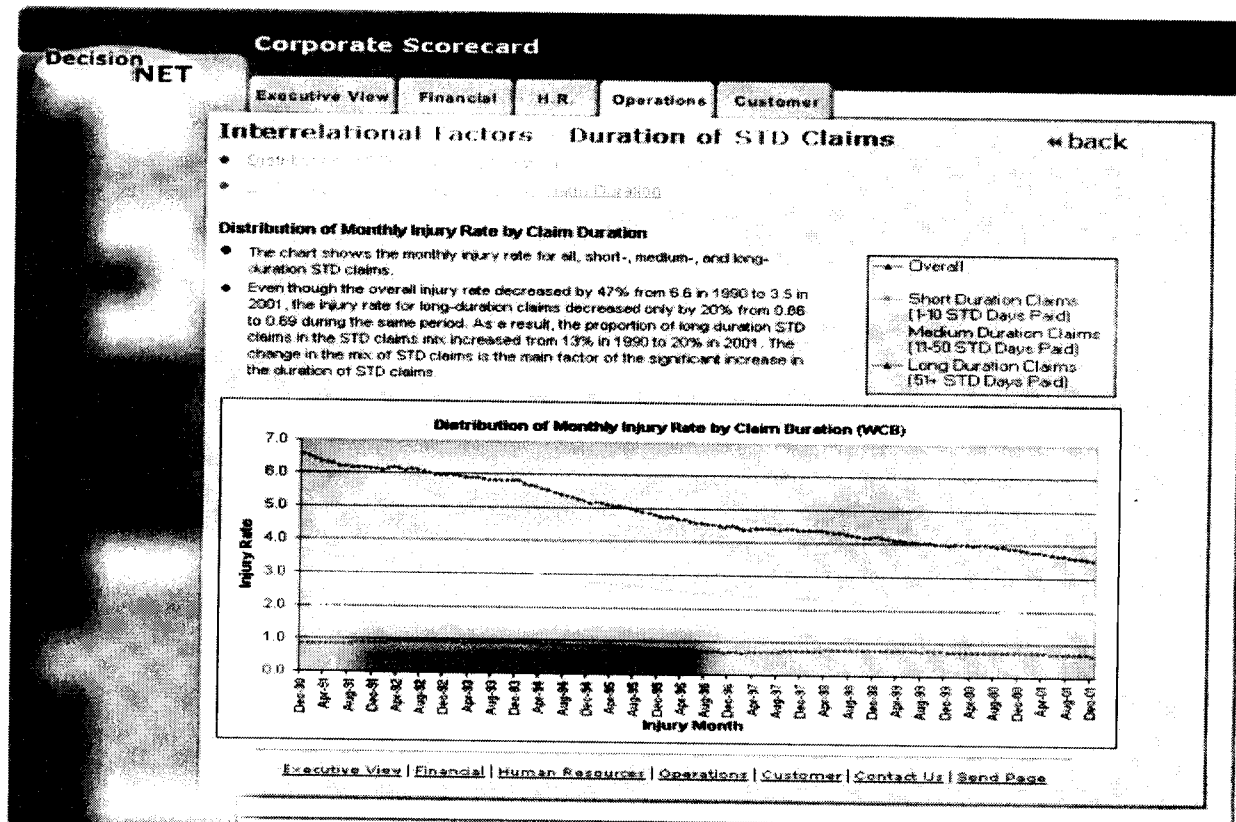
Figure 3e – The Inter-relationships: Duration of STD Claims

Although there are many metrics that can be communicated when looking at the inter-relational factors, it is important to communicate only the key ones. For the "Duration of STD Claims", the most important relationship is with Injury Rate (the number of injuries per 100 employee years). The injury rates for the short, medium and long duration claims are changing at different rates; in fact, the proportion of short duration claims has decreased dramatically over the past 11 years, while long duration claims have actually increased. The mix of short and long duration claims has changed, thus prompting an upward trend in the average "Duration of STD Claims" over the history of the available

21

data. Put another way, over the past 10 years or so, the WCB has been very effective at reducing the number of short duration claims, thereby driving up the average claim length. This is one of the main reasons for the increasing average duration and is briefly outlined in the KPI's summary.

With the dashboard the WCB will be able to better manage its information. In an organization with as much data as the WCB, it is vital for the organization to develop not only methods for transforming data into information, but also for distilling that information down to what is most important, particularly before it is communicated to the executives. This application takes a giant step in that direction, dramatically enhancing the effectiveness and the communication of the WCB Balanced Scorecard, and improving the management effectiveness of the executives.

III. PERFORMANCE EVALUATION – SETTING MENINGFUL TARGETS

Targets are used by an organization to quantify strategic objectives and to motivate employees. The planning process itself helps the members of the organization to better understand and appreciate where the organization has been and where it should be going. In this chapter we discuss the structuring of such targets; we examine some of the more common approaches used to plan targets; and we end by addressing some additional considerations.

3.1 Structuring the Target Planning

We begin by looking at why targets are used, who generally sets them, and when and how they are chosen. We follow this with a discussion about the number of "zones of performance" and the method of communication. Finally, we end this section with a brief paragraph on setting monthly targets.

3.1.1 Initial Target Setting Questions

Why Set Targets?

Organizations set targets to motivate continuous improvement and to focus the employees on the strategic goals of the business. Setting targets allows the organization to communicate its strategy in measurable terms. If a target is not being met, then actions are triggered within the organization in an attempt to improve performance. In order to affect results, existing initiatives may simply need fine tuning, but new aggressive initiatives may also be required.

Target setting must begin somewhere. For profit-oriented organizations it starts with the shareholders' expectations, and these are usually linked to the ROE (Return on Equity) or a similar business metric. If, for example, the organization needs a 15% ROE, this figure will affect the planning of many other metrics in the organization, as many metrics of the Balanced Scorecard are inter-related.

For non-profit organizations, on the other hand, the "bottom line" is important but the ROE is not (generally the business will want to break even). However, other priorities, such as customer welfare and satisfaction or quality and fairness of service are just as important for non-profit organizations.

23

Who Sets Targets?

The WCB is regulated, and its Board of Directors determines many of the high-level strategic decisions that the executives implement. When setting targets for the KPIs, the reactions of the stakeholders and regulatory body must be considered. In addition, employees directly and indirectly responsible for meeting performance targets must buy-in to the target levels chosen. Performance targets must be seen as reasonable, achievable, and good for those accountable and the organization. Despite all of this, it is ultimately the executive team that is responsible for setting the targets and for effectively communicating their importance.

When Are Targets Set?

In most organizations, planning the expense budget occurs annually, and the Balanced Scorecard KPI targets should be set at the same frequency. For an organization such as the WCB, where there are no competitive forces at work and the organization is as stable in any given year as the economy, an annual planning process is sufficient for setting targets. However, targets must be flexible. If there is a significant change in the environment, the business targets might need to be re-evaluated and re-set. It is not effective for an organization to keep the same goals once they become unreasonable or are achieved very early on in the year.

How Are Targets Set?

Making a plan or setting a business target requires both objective and subjective information. The target must be not only realistic and attainable but also challenging enough to require effort. Ideally, a target should also be beyond the natural fluctuation of the metric so that it is not likely to be "accidentally" achieved.

In this chapter we consider three approaches to setting target levels:

- *Forecasting Approach:* Compare to expected future values.
- *Improvement Goals Approach:* Compare to relatively arbitrary targets set by the executive team as part of the organizational strategy.
- *Benchmarking Against Others Approach:* Compare to the best results from other, related organizations.

24

3.1.2 Performance Zones and Communicating Performance

Performance Zones:

We define "performance zones" as the different levels of performance to be communicated by the dashboard application. In a single-threshold approach, there are only 2 zones; performance is simply either good or poor, depending on whether or not a target is met. This is the most common approach used by organizations when setting targets. A typical example is the budget. It is a "make it or break it" approach, requiring only one target, such as "produce 55 units this month".

The Balanced Scorecard methodology generally promotes the use of a third zone, somewhere in-between the classifications of poor and good. This zone can be used as either a warning light to suggest that targets are not quite being met, or as an indicator that performance is average. In the case that a warning indicator is used, the reaction of the organization would be more urgent than if an "average" indicator is used. In both of these cases a new line must be drawn, and so two targets are required, such as an upper cut-off of "55 units" and a lower cut-off of "20 units".

Another possibility for metric performance evaluation is to use more than 3 zones. Defining many zones of performance allows for the communication of results at various levels of good through various levels of poor. For example, a 5-zone system would require 4 targets, such as "55 units", "45 units", "30 units", and "20 units". With this structure, a result greater than 55 might be "Excellent" performance; between 45 and 55 would be "Good"; 30 to 45 would be "Average"; 20 to 30 would be "Warning"; and below 20 would be "Poor".

The last approach tends to be less popular than the 3-zone system in dashboard applications. This could be due to the fact that executives want to keep the interpretation of the performance indicator as simple as possible. Typically, colors are used to depict the different levels of performance, and the Kaplan and Norton recommendation is the traffic light system (red, yellow, green). It is a very familiar and easily understood system for everyone in the organization, and it requires minimal training: "Green is Good", "Red is Bad", "Yellow is In-Between".

Communication Method:

There are two common methods used to communicate metric performance in a dashboard application. First, there is the traffic light system already mentioned. However, the next question is

usually, To what degree is performance poor? If it is yellow, is performance closer to the red or to the green light?

A possible solution to this is to communicate performance on a dial, somewhat similar to a speedometer. Figure 4 below contrasts the traffic light and dial systems; it can be seen that the dials offer more information than the lights. (The dial zones are green, yellow, and then red - from right to left. In this example, the size of the yellow zone was chosen arbitrarily to communicate average results.)

Green Light	
Yellow Light	
Yellow Light	
Red Light	

Figure 4 – Comparing Performance Lights and Dials

At the time of this thesis, the WCB has decided to go with the simpler traffic light system, with the three performance zones represented by the colors red, yellow, and green. For some KPIs the yellow light will be a warning, while for others it will indicate average performance (each choice will be communicated within the dashboard).

3.1.3 Setting Monthly Targets

Usually the target setting process will begin with the executive choosing "stretch targets". As part of a 5-year plan achievable but aggressive targets are chosen for the long-term. These targets are next broken into one-year targets. Once the year-end targets are known for the given year, they must be divided so that progress can be tracked monthly. Fortunately, given that all of the KPIs are smoothed (the 12-month moving-average) the impact of seasonality can be ignored.

The approach to setting monthly targets is varied. A common approach is to interpolate an equal improvement for each month of the year. However, it may be more realistic to set lower targets at the beginning of the year and higher ones nearer the end of the year so that improvement initiatives and projects have time to be fully realized. The same could be said for the 5-year plan.

3.2 Three Approaches To Setting Targets

This section introduces the most common methods for setting targets: forecasting historical trends, setting improvement goals, and benchmarking against other organizations. Forecasting can involve many techniques, and essentially assumes that the organization wants to measure performance against the expected values. Improvement goals set by executives target better results than would normally be expected but usually use only the prior year's values as the reference point. The benchmarking approach uses results from a variety of similar organizations to set its targets, as long as comparable metrics are available.

3.2.1 Forecasting Approach - Expected Performance

Effective forecasting is one of the most challenging aspects of business planning. There are many complexities to consider, such as choosing a forecasting method and deciding how to handle cycles, seasonality, and interventions (one-off events). In addition, many factors and events can occur that are not expected, due to either internal or external environmental changes.

The goal of this target-setting approach is to assess actual results against the results that the organization "expected" to achieve (Table 1). Once a series has been projected, there is the issue of how to handle expected changes to the internal and external environments. Factors (such as regulatory changes) that are expected to have a significant impact on future performance must be estimated. These estimates should then be used to adjust the projected values to create a more "realistic" projection.

After projecting the series, it is important to define a tolerable deviation from the "expected future". These thresholds are the targets and represent the "acceptable future". With some projection methods it is possible to determine a confidence interval, which takes into account the forecasting error. This information can be useful when deciding on the range for acceptable performance.

Where do you expect to be?
1. Historical Trend: *project data series using statistical techniques, while accounting for cycles and seasonality.*
2. Internal Changes: *make an estimated correction to the forecast for the cumulative impacts from ongoing initiatives or projects, policy and procedural changes.*
3. External Changes: *make an estimated correction to the forecast for the cumulative impacts from the changing environment, such as regulation and law changes, demographic shifts, bankruptcy, and unemployment rates.*

Table 1 – Target Setting Using The Forecasting Approach

Common choices for projecting a time series include simple linear regression, exponential smoothing, and ARIMA.

Regression tends to be the easiest of these three methods and is commonly used for time series' that are relatively stable. Simple linear regression first estimates a "best fit" trend line. The typical formula is: $y_t = \beta_0 + \beta_1 x_t + e_t$ where β_0 is the intercept, β_1 is the slope, and e_t is the forecasting error. This formula is then used to project the series into the near future. Unfortunately, for metrics that are cyclical, contain seasonality, or have a significant amount of noise, this regression model is not very effective.

Exponential smoothing is a little more complicated for the non-quantitatively inclined members of the organization to understand than the simple linear regression method, but it is a more flexible and a more effective approach. Simple exponential smoothing begins by multiplying the current value of the time series by a smoothing constant. One minus the smoothing constant is then multiplied by the prior period's forecast, and the two numbers are added together to get the next period's forecast. The formula for this method is: $S_{t+1} = \alpha X_t + (1 - \alpha) S_t$ where S_{t+1} is the forecast, X_t is the current period performance, and α is a smoothing constant between 0 and 1. The simple version of this method is effective with series' that are cyclical or otherwise irregular. There are several other, more complex versions of exponential smoothing that may or may not be better for forecasting, but it depends upon the time series. In particular, the Holt-Winters method has been shown to be quite effective.

Another choice is ARIMA. This method is the most complicated process of the three mentioned in this section, and it definitely requires the assistance of forecasting software such as SAS or SPSS and expert analysis. ARIMA takes exponential smoothing to the next level, combining moving average

terms (MA: lags of the forecast errors) with auto-regressive terms (AR: lags of the differenced series). If the series is non-stationary then it must be integrated (I: differenced series). If the series variance is not constant then a logarithmic transformation might be necessary. From an organizational perspective, ARIMA it is far too complex and, thus, it is not practical for the dashboard application. Any metric forecasting method used must be understandable to and accepted by the KPI owners, and ARIMA is not easily explained.

Another important consideration when choosing a forecasting method is that ARIMA and multiple regression allow for transfer functions or explanatory variables (the use of additional data series' to improve the projection of future values). For example, economic indicators (such as unemployment rates and GDP) might be correlated with investment performance or claims experience. For some factors, using explanatory variables is a quantitative alternative to the highly subjective after-projection adjustment suggested earlier.

Finally, it must be recognized that there are several important limitations to forecasting. First, the quality of the data must be reasonable. It must be valid, consistent, credible, reliable, verifiable, accurate, and where possible indisputable. Some metrics possess only a limited amount of data, and this can significantly influence the effectiveness of the forecast. Another problem is interventions. An intervention is the historical occurrence of a non-typical event that is not likely to occur again and that cannot easily be predicted. Where possible, interventions should be removed from the data, even if this means adjusting or truncating the time series.

3.2.2 Improvement Goals Approach - Desired Performance

The second, and most common, approach to setting targets for an organization is the improvement goals approach. In typical organizations the executive team determines that an improvement in the bottom line is necessary. The 2002 year-end goal might take the form: "to achieve 1997 performance targets" or "to see a 10% improvement over 2001 year-end." Generally, this is an arbitrary figure, the intent of which is in line with the vision of the organization but the magnitude may or may not be reasonable (depending upon the trend and future environmental changes).

It is vital that for any improvement goal that the organization has a comprehensive plan in place to ensure that the improvement is attainable. Initiatives and projects within the company are necessary

for the organization to experience "better than expected" results. The type of project depends upon the KPI. For example, from a human resources perspective an employee education program on ergonomics might be necessary to improve sick leave results; social activities can have a positive influence on employee moral and productivity; or, a more strict enforcement of safety rules can reduce the workplace accident rate.

3.2.3 Benchmarking Against Others Approach - Leading Performance

It is very difficult to compare the WCB with other organizations as it has a monopoly in BC, and so there are no competitors. However, there are WCBs in other provinces and in the United States, and most of these use business metrics to monitor their performance. In addition, the results from other government organizations may also be useful in setting achievable benchmark targets.

For the WCB, the most obvious comparison is with other Canadian WCBs. The most often chosen are Ontario and Alberta due to their similar regulations. For the human resource metrics, other large, regulated, unionized organizations, such as BC Ferries and ICBC, can be used too.

However, even if metric data is available from other organizations, it is likely that the methods of data collection and metric calculation are very different, and regulations are not the same in every jurisdiction. Some provinces offer broader or narrower coverage than others, while some have larger or smaller claim benefits. The differences between the WCBs are probably countless. However, despite all of this, in some cases values are still comparable and can be used in target setting.

The Association of Workers' Compensation Boards of Canada (AWCBC), a nation-wide organization established to facilitate the exchange of information between WCBs, has tried to bridge the comparison gap by creating standard metrics to be reported by all Canadian WCBs. The organization collects the results of the predefined metrics and provides comparison reports, tables, and charts. Unfortunately, however, there is only a small overlap between the WCB Balanced Scorecard KPIs and the AWCBC metrics at this time.

For those KPIs that can be benchmarked, it may be done in several ways. Commonly an organization will rank itself against others in a group, and the target might be as aggressive as to be the "best in class", or something more liberal, like to be "in the top third of performers". However, because of the

30

lag time for organizations to publish their results, ranking does not give the organization a clearly defined *target to* shoot for. Another option, then, is for the organization to use the prior year's values of the best performing companies when planning targets, making adjustments for the changing environment if necessary.

3.3 Contrasting the Three Approaches To Setting Targets

Each of the preceding methods of setting targets (forecasting, improvement goals, and benchmarking) uses different information to accomplish the same objective: to focus the organization on improvement. By successfully achieving better and better KPI results the organization will move toward its strategic goals. Note, however, that individually each method has its own strengths and weaknesses.

To start with, the forecasting approach utilizes both historical information and knowledge about the changing future environment to project where the organization expects to be going. This projection is a useful "target" if the executive team is comfortable with current results and is not seeking improvement. The biggest challenge with this method is the forecasting accuracy. Is the missed target the result of unexpectedly poor performance, or, is it the result of poor forecasting?

The improvement goals approach is likely to be the most commonly used method of planning targets. An advantage of this method is the improved employee buy-in. There is a clear goal set by the executive, and the performance assessment does not depend on subjective forecasting. In addition, although setting improvement goals is itself a highly subjective approach, it allows the executives to incorporate qualitative information and their "gut feel" into the target-setting process. The biggest limitation of this method, however, is that often the improvement target is set relative to only the prior year's result. This method could be improved with a better understanding of the KPIs, including the historical trends and cycles. If executives continually set unrealistic targets, employee motivation could be affected, and the effectiveness of the Balanced Scorecard will suffer.

The third approach discussed was benchmarking against others. By measuring performance relative to other organizations undertaking similar activities, the organization can set targets that are being achieved by others. Benchmarking tends to be more effective between competitors in the same market place than between independent geographical monopolies. The biggest problem with the

31

benchmarking approach, though, is that most business metric calculations are complex and methods are likely to be different across organizations. In addition, a lag in the publication of the results of other organizations means that the organization could be chasing an "invisible target".

3.4 A Combined Target Setting Approach

Given that individually the effectiveness of the target setting approaches outlined above is limited, we should consider a method for combining them. It is by combining these three target-setting approaches that we can improve the performance of the executive's dashboard, and ultimately, the organization's ability to achieve its vision.

Independently, the three previously mentioned approaches answer the questions: Are we better or worse than we expected to be, given the evolving internal and external environments? Are we achieving our corporate improvement goals? Are we a top performing organization? A combined approach, on the other hand, will allow us to answer the question: Are we meeting our strategic goals and moving toward our vision?

But how should this information be combined? This is an opportunity for significant further research. However, for the purpose of this thesis we propose starting the process by first forecasting where the organization expects to be with adjustments for environmental changes. This projection should then be adjusted by the improvement goal, which is chosen using benchmarking information and knowledge of what can reasonably be done to influence the results (the potential impact of initiatives). As a final consideration, targets for different KPIs should be homogenous with other KPIs.

```
┌─────────────────────────────────┐
│        1. Historical Trend       │
│    ┌─────────────────────────┐   │
│    │    2. Internal Changes   │   │
│  ┌─┴─────────────────────────┴─┐ │
│  │     3. External Changes      │ │
│┌─┴─────────────────────────────┴┐│
││ 4. Improvement Goals / Benchmarking ││
│└─────────────────────────────────┘│
```
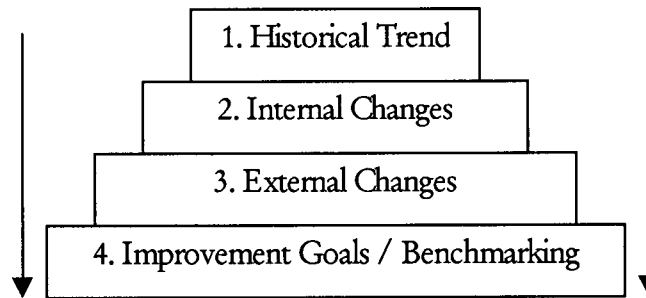
Figure 5 – Combining the Target Setting Approaches

Figure 5 above is a simple chart representing the combined approach to setting KPI targets. Below in Table 2, meanwhile, is a more detailed outline of the combined approach.

A. Where do you expect to be? **1. Historical Trend:** *project data series using statistical techniques, while accounting for cycles and seasonality.* **2. Internal Changes:** *make an estimated correction to the forecast for the cumulative impacts from ongoing initiatives or projects, policy and procedural changes.* **3. External Changes:** *make an estimated correction to the forecast for the cumulative impacts from the changing environment, such as regulation and law changes, demographic shifts, bankruptcy, and unemployment rates.* **B. Where do you want to be?** **4. Improvement Goals:** *set quantitative improvement targets with respect to the organizational strategy, which might be to seek aggressive improvements in some KPIs and status quo on others.* **5. Other Organizations:** *adjust improvement goals with respect to other similar organizations' past results, making goals more aggressive where necessary.* **6. Inter-Relationships:** *make additional adjustments to KPI targets to ensure that there are no conflicts between KPI targets. Many of the KPIs are inter-related, and conflicting targets will cause confusion.*

Table 2 – Target Setting Using The Combined Approach

If targets are set such that red lights are too many or too frequent, their effectiveness in the dashboard will be reduced as people become desensitized. Frequent red lights can also impact employee moral and motivation. So, it is important to keep the overall picture in mind when setting targets, making them challenging but not too aggressive. Ultimately, what we are looking for is a target that will signal a need for action if it is not met. A target that is not met will trigger a red light, and this light will attract

the attention of the executives who will want to know why the light is red and what is being done about it. KPIs must be actionable, and through corrective action it is expected that performance will improve.

3.5 Additional Considerations When Setting Targets

In this last section of the chapter we address some of the miscellaneous issues related to setting performance targets. We begin with a discussion on the inter-relationships between KPIs and the impact of mid-plan changes. Next, we consider the idea of tracking the performance lights over time, and then introduce two alternative concepts for performance measurement: using historical deviations from plan, and trend target setting.

3.5.1 Inter-Relationships Between KPIs

No discussion about setting performance targets in a Balanced Scorecard would be complete without discussing the inter-relationships between business metrics. There is no established scientific method to dictate how to ensure all strategic targets are consistent. However, when choosing targets for any one KPI consideration must be given to other metrics' targets. All corporate targets must be synergistic, i.e. non-conflicting.

As an example, if the corporate strategy leads the executive to set aggressive goals to reduce the injury rate, which requires spending more on existing and new programs, then targeting an aggressive reduction in the operational expenses as well may be unrealistic and unachievable. These two strategies, and thus the two targets, will work against each other, increasing tension in the organization and reducing the chances that either target will be achieved.

3.5.2 Adjusting Targets Due to Environmental Changes Mid-Plan

An important concept in the Balanced Scorecard methodology is that targets should be flexible and appropriate. Given that the business environment can change rapidly, it may be necessary to amend performance targets mid-term. Without a meaningful target, the measurement process could lose credibility and the organization could lose focus.

To give an example, if regulations change early into the year, and it is obvious that planned targets will not be achieved or will be too easily achieved, then it makes sense to re-plan the year-end target for

each affected KPI, resources permitting. The caveat here is that the target has become unrealistic due to "uncontrollable" circumstances. For target setting to be an effective and integral part of achieving the organization's strategic objectives, targets should not be adjusted or changed lightly, particularly not to artificially make performance look better than it should look.

3.5.3 Tracking Past Performance To Past Plans

Additional information that might be useful to the executive includes a chart of the historical performance lights. Over the past 48 months, how have the actual results compared to the targets? If there is chronic poor performance, then either the planning process needs to be improved, the accountable employees need to be motivated, a better understanding of the controllable and uncontrollable factors driving the KPI need to be explored, or additional resources are required.

3.5.4 An Alternative: Significant Deviation From Target

When single corporate targets are used regularly, another approach can be taken to signal performance. By tracking the difference between the actual values and the target over time, the deviations from plan can be analyzed to determine whether future deviations from target are significant.

The concept here is that some deviation from plan will likely occur due to unexpected influences in the outside and inside environments. If we track historically what that deviation from plan has been, then we might be able to measure the "normal deviation from plan" by calculating the standard deviation of the differences. This expected deviation could then be used to differentiate acceptable from unacceptable results. Red, yellow and green lights could be set using only a single target and the known deviations. For example, if targets have been set, and historically the actual values have been within 5% of the past targets 80% of the time, this percentage could be used as the threshold. A red light would be triggered if current month value is more than 5% higher than the target, a green light if it is more than 5% below the target, and a yellow light if in-between.

The complication to this approach, however, is that every year's plan will be different and that this information must be tracked. (The WCB only has this information for a couple of its KPIs, and it is incomplete.) In addition, plans are frequently set as year-end targets and are not allocated to each month. Even if a monthly plan is available, many times it is a year to date figure, and this information is very challenging to work with. Given that the WCB has now built the dashboard application, and

that the targets will be set monthly, the data required for this analysis should be available in 3-5 years. This concept is put forth here as an idea, but it has not yet been explored in any depth.

3.5.5 An Alternative: Trend Target Setting

This chapter focuses on setting targets. But what about the trend? It is possible that a monthly result may exceed its target, turning a light green. But, if that result is part of an extremely poor trend then perhaps the communication should be different.

One option is to set target "slope values" for each of the KPIs. However, this is probably not a realistic recommendation as it is a very difficult concept to sell to an executive, and ultimately it is similar to setting monthly targets at the desired rate of change.

There is another possible approach to using the trend information though. The Historical Ranking Process outlined in the next two chapters proposes a method for evaluating performance using just the historical data. It uses not only the most recent month's value, but also the trends in the data. By ranking the current measures against past measures and then using cut-offs, performance can be qualified as red, yellow, or green. Red is defined as a poor result relative to past performance, yellow signals an average performance, and green communicates a good performance.

IV. PERFORMANCE EVALUATION – HISTORICAL RANKING PROCESS

In this chapter we use historical data to evaluate the performance of business metric values. We begin by discussing the data at the WCB, the models we used to assess performance, and our method for selecting the necessary model parameters. To demonstrate our methods we use a simplified version of one of the WCB's Balanced Scorecard metrics.

4.1 Data

Data is the source of the performance information that the executives will use to make important decisions in the organization. It is vital that the data is integrative and stable. In addition, smoothing the data yields a more stable picture of performance, with reduced noise and no seasonality.

4.1.1 Data Integrity

Every analysis using historical data must first begin with an assessment of data integrity. The key issues are quality, consistency, and to some extent, quantity.

Quality data: The information used must be correct. Survey results must be unbiased; calculation and data entry errors must be insignificant. Data errors can significantly impact decisions within the organization if not corrected or accounted for. Another common problem is data that has become distorted during a conversion from one information database to another. This movement of information, despite tight controls, can result in data corruption. Where possible, end results should be validated against other sources and previous studies.

Consistent data: Over time the business environment changes and calculation methods can change. In an industry as heavily regulated as the Workers' Compensation Board, a government bill can change the way the company does business. Changes in coverage or changes in premium structure, for example, can result in information recording one thing for several years, and then recording something else afterwards. Changes in the environment can result in changes to the distribution of the underlying factors, the "mix". This is a particularly important consideration when the metric is an average result, combining numerous subsets into one value. The increased concentration of any one subset can change the overall picture without performance necessarily being good or bad.

Data quantity: For an effective analysis when using historical data, a reasonable amount of data must be used. For businesses experiencing any sort of seasonality or business cycle, one year of data is not enough to make conclusions. For more complex businesses tied to the economy, business cycles of 3-5 years are not uncommon. The bottom line is, obviously, the more information one has available, the more effective the analysis is likely to be. As a rule of thumb, we suggest at least 3 years but preferably more than 10 years worth of monthly data (36-120 periods).

Overall, the data at the WCB is in reasonable condition. The complexity of their business information requires that an internal data analyst be utilized for the retrieval of the appropriate data. An analyst will also ensure that the information is interpreted correctly. There are many subtleties to the data in a complex organization, and it is easy to retrieve data that is not what it appears to be.

With regards to quantity, the WCB has a large volume of information for most of its business KPIs. The Balanced Scorecard has been used for almost 5 years, and for the operational metrics the data warehouse has up to 15 years of history. Unfortunately, due to a new information system in 1999, the human resource metric information has only 3 years of history.

4.1.2 Data Smoothing

For many business metrics, the raw data used is generated on a monthly basis, a measurement of the experience between the first and the last day of each month. An example of this data is the number of claims reported per month. For other metrics, however, a snapshot taken of total or aggregate amount at the end of a particular month is used. This could be the total amount of capital tied up in investments.

In some cases, the monthly measurement is converted and tracked on a "year-to-date" performance. Here the metric tallies information from January of the current year to the most recent month-end. Where possible this data should be adjusted back to current-month values for the Historical Ranking Process performed in this chapter.

Although many metrics are tracked monthly, some data such as WCB payroll information is not; in fact, payroll is recorded on a pay-period basis instead. Pay-periods at the WCB are bi-weekly, which means that there are 26 measurements each year (compared to 12 when we measure monthly). Pay-

periods are also used to track other human resource metrics at the WCB such as leave hours, management ratio, turnover, and productivity. There is no easy way to convert this raw data to a monthly basis, and so the Historical Ranking Process discussed in this chapter must be adapted for this type of data. We go into a little more detail about this later in this chapter.

Given that we have monthly raw data, we need to smooth the variability caused by seasonal and other monthly fluctuations. There are several methods that can be used to remove seasonality: differencing, seasonality factor adjustments, and moving averages. The most accepted process in business metrics is to use moving averages, and in specific, a 12-month moving average.

To convert the raw data to a 12-month moving average, we simply take the value each month, add the previous 11 months and divide by 12. Each of the 12 months is weighted equally, and in effect, this results in a one-year performance measurement as of any given month-end.

4.2 Models

In this section we develop three slightly different models for evaluating performance using historical data. We start by defining four measures that represent the four business questions we want to answer, and follow this with a simple approach to measuring trends. Next, we look at the three different historical ranking models, the options for ranking performance, and the choices for cut-offs. Once we have evaluated the most recent performance of the four measures, we look to combine this information into a single, overall indicator of performance. All of these rankings and indicators can then be charted to give a dynamic view of performance over the history of the metric. This will prepare us for the next section, in which we discuss how to choose the appropriate model and parameters.

4.2.1 Measures

It is important to start by defining the measures we are going to use in our models. The simplest measure by far is the most recent month's value. This value is the standard measure for determining performance, but it is an incomplete measure of performance, as it does not provide any information about the trend. Is performance improving or deteriorating?

The simplest indicator of a trend is the slope of the time series over the most recent months using linear regression. Another option is to use both the slope and the intercept information to project a value, using this result to evaluate performance. Both are explored later.

The next question is, how many months should be included in the slope measure? The answer depends on the number of periods of data available and the business question one wants to answer. For many of the metrics that originate from the WCB's data warehouse, more than 15 years of monthly data is available. Given this volume of data, it makes sense to evaluate several trend measures. For our basic model structure, we chose two trends: a short term and a long term. Given the general business definitions of "short term" and "long term", we can define them as 12 months and 60 months respectively.

Finally, there is also value in the trend that exists given the entire data series. To ensure that there is an indicator that utilizes all of the data available, a global trend is used as a fourth and final measure.

These four measures will be used to answer the following four business questions:

Global Trend (GT): How is the overall trend performing?

Long Term Trend (LTT): How is the 5-year trend performing?

Short Term Trend (STT): How is the 1-year trend performing?

Most Recent Value (MRV): How is the most recent month performing?

Figure 6 below gives a visual representation of these measures.
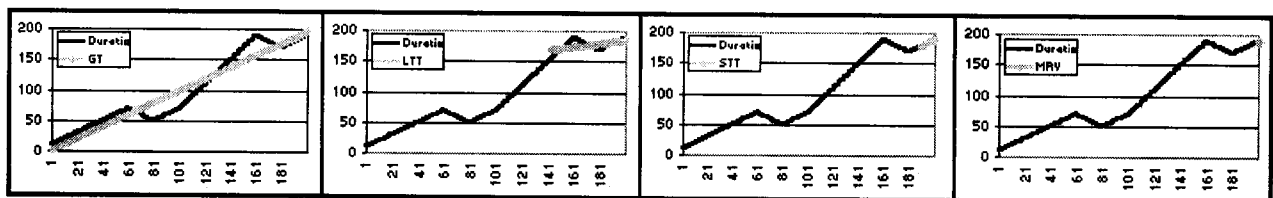
Figure 6 – The Four Measures: GT, LTT, STT, MRV

4.2.2 Linear Regression

To quantify these trend measures, we calculate slopes and intercepts using linear regression and the method of least squares to find the best-fit lines to historical data. In other words, we minimize the sum of the squared distances between the actual and predicted values to determine the trend line. MS

Excel does this very easily for us, and we can calculate the values just by using the "slope" and "intercept" functions.

And so, we define the following for use in the next section:

ActVal = actual current month value = y_t

$\text{Slope}_{t-p-1,t}$ = slope (or trend) = β_1

$\text{ProjVal}_{t-p-1,t}$ = projected value = Y_t = $\beta_0 + \beta_1 * t$

Where:

t = current period

p = number of periods to include in the linear regression.

4.2.3 Model Options and Measurements

Given that we now have a set of four generic measures from Section 4.2.1, there are several possible variations that will be useful in answering the aforementioned business questions. The most basic method is to just evaluate the slopes of the trend lines (Model A in Figure 7 below), which simply measures the rate of change of the KPI. Another possibility is to evaluate the deviation of the shorter-term trends from the global trend; this measures the accelerating rate of change (Model B in Figure 7 below). A more complicated model, but one with a useful business interpretation, is to compare the rate of change to each successive rate of change instead of to the global trend (Model C in Figure 7 below).

Figure 7 – The Three Models: A, B, C

Given that we now have the tools to answer the four business questions and that we have three possible variations, we can consider three models: A, B, and C. Using the least squares method, linear regression analysis is applied to this data to find the slopes and intercepts.

Model A: Trend Approach
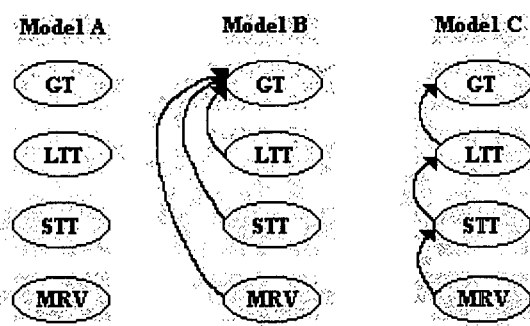
- Global Trend (GT): The slope of the regression line using the *entire data set.*
- Long Term Trend (LTT): The slope of the regression line using latest *5 years* worth of data.
- Short Term Trend (STT): The slope of the regression line using latest *12 months* worth of data.
- Most Recent Value (MRV): The *most recent month's value.*

Model B: Deviation from Global Trend Approach

- Global Trend (GT): The slope of the regression line using the *entire data set.*
- Long Term Trend (LTT): The percent deviation of the slope of the latest *60 months* from the *Global Trend* slope.
- Short Term Trend (STT): The percent deviation of the slope of the latest *12 months* from the *Global Trend* slope.
- Most Recent Value (MRV): The percent deviation of the *most recent month's value* from the projected value of the *Global Trend* linear regression.

Model C: Deviation from Successive Trends Approach

- Global Trend (GT): The slope of the regression line using the *entire data set.*
- Long Term Trend (LTT): The percent deviation of the slope of the latest *60 months* from the *Global Trend* slope.
- Short Term Trend (STT): The percent deviation of the slope of the latest *12 months* from the slope of the latest *60 months.*
- Most Recent Value (MRV): The percent deviation of the *most recent month's value* from the projected value of the *12-month* Trend linear regression.

At this point, a slight change in our approach is required. When calculating the deviations from a slope, there is a danger that the slope may approach zero. Models B and C both have this weakness and can become unstable. However, in many business metrics it is not likely that the GT will go to zero once an increasing or decreasing trend is well established. This means that Model B is likely to be acceptable. Model C, however, uses the 5-year trend as the benchmark for the 12-month trend in the STT measure, and this measure can quickly become unbounded, and thus ineffective. Many of the business metrics we can evaluate with Model C have this problem. But, based on our experimentation, a solution to this is to compare projected values rather than slopes. This new approach uses not only the slope but also the intercept information. Now, given that we are only using this result to evaluate deviations, and that the Model C approach is too sensitive to slopes approaching zero, we use a

modified Model C that compares the projections of the linear regressions rather than their slopes. This results in a more stable model, as there are no WCB KPIs where the projection approaches zero.

In Table 3 are the final models, using the notation defined in subsection 4.2.2 above:

Model A:	Model B:	Model C:
GT: $Slope_{1,t}$	GT: $Slope_{1,t}$	GT: $ProjVal_{1,t}$
LTT: $Slope_{t-59,t}$	LTT: $(Slope_{t-59,t} - Slope_{1,t}) / Slope_{1,t}$	LTT: $(ProjVal_{t-59,t} - ProjVal_{1,t}) / ProjVal_{1,t}$
STT: $Slope_{t-11,t}$	STT: $(Slope_{t-11,t} - Slope_{1,t}) / Slope_{1,t}$	STT: $(ProjVal_{t-11,t} - ProjVal_{t-59,t}) / ProjVal_{t-59,t}$
MRV: ActVal	MRV: $(ActVal - ProjVal_{1,t}) / ProjVal_{1,t}$	MRV: $(ActVal - ProjVal_{t-11,t}) / ProjVal_{t-11,t}$

<center>Table 3 – Model Measure Formulae</center>

Now, let us start by applying these models to the simplified 12-month moving-average data series from Figure 8 below. Here we have the WCB business metric "Duration of Short Term Disability (STD) Claims" as a raw monthly time series (Figure 8a) and a 12-month moving-average smoothed time series (Figure 8b). The third chart (Figure 8c) is a simplified series based on the 12MMA-smoothed series in Figure 8b. It was created using straight lines with slopes 1, -1, and 2, to approximate the progression of the Duration KPI, but without the noise. The simplified series begins at $t=1$ with the value $y_1 = 10$ units, and steadily increases to 190 units by $t=200$. It is this simplified time series that we will use to illustrate our Historical Ranking Process performance assessment method in the following sections, but we will use the real data (Figure 8b) in Chapter V.
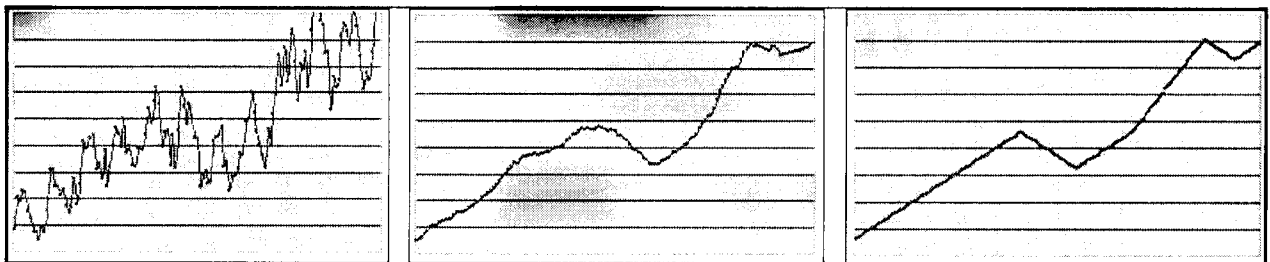
<center>Figure 8a,b,c – Raw Time Series, Smoothed Time Series, Simplified Time Series</center>

Table 4a below contains the evaluation of the measures at period $t = 200$. For Models A, B, and C the GT, LTT, STT, and MRV have been calculated to be:

<center>43</center>

Measures	Value
Model A	
Global Trend	0.97
Long Term Trend	0.29
Short Term Trend	1.00
Most Recent Value	190
Model B	
Global Trend	0.97
Long Term Trend	-0.71
Short Term Trend	0.03
Most Recent Value	-0.03
Model C (Projections)	
Global Trend	196
Long Term Trend	-0.05
Short Term Trend	0.02
Most Recent Value	0.00

Table 4a – Measurement Values (t=200)

The interpretation of these values is:

Model A

- GT: Since time t = 1, the average increase has been 0.97 units/month.
- LTT: Since time t = 141, the average increase has been 0.29 units/month.
- STT: Since time t = 189, the average increase has been 1.00 units/month.
- MRV: At time t = 200, the most recent month's value is 190 units.

Model B

- GT: The same interpretation as in Model A.
- LTT: The average monthly increase over the past 5 years is 71% less than that for the past 17 years.
- STT: The average monthly increase over the past year is 3% more than that for the past 17 years.
- MRV: The most recent month's value is 3% less than the global trend projection.

Model C

- GT: The value projected by the global trend is 196 units.
- LTT: The 5-year trend projection is 5% less than the global trend projection.

- STT: The 1-year trend projection is 2% higher than the 5-year trend projection.
- MRV: The most recent month's value is the same as the 1-year trend projection.

The information from these models will be used in later steps to provide a more complete assessment of KPI performance. Combined, these measures can be used to determine exactly how significant performance is. For example, if the Long Term Trend, Short Term Trend and Most Recent Value are determined to be poor, this is more significant to the organization than if the trends are determined to be reasonable and only the latest month is poor. But, how do we assess performance? How do we distinguish poor from good performance? The next section discusses our chosen approach in depth.

4.2.4 Ranking Performance

Now that we have values for each of the twelve measures, we can look at ranking their performance. The same measure calculations from the previous section can then be made for every month in the data set, going back as far as mathematically possible. This will allow us to compare the most recent measures to the past measures.

The LTT measures the 5-year trend, which requires 60 months of data to begin the calculations. In addition, for Models B and C we are comparing the GT to the LTT, and as both use the same information at time t=60 months, the value will be zero, and the deviation will be very small for periods immediately following time t=60. This means that we need at least 6 years of data just to apply the models. But, to establish a moving picture of performance for the ranking, more than 10 years of data is preferable.

Once these calculations have been made, we next need to assess the performance of the metric at time t. There are a number of methods for doing this. We chose to look at three: Z-score, Percentile, and Min/Max.

Z-score: The Z-score is calculated as the number of standard deviations an actual value is from the mean of the measurements, with the sign indicating the direction. The percentile of the deviation between the actual and the mean can be determined by using the Z-score table. However, the Z-score is based on the assumptions that over time values will have a constant mean and will be normally distributed around this mean, and business metrics rarely support either of these assumptions.

Percentile: Using the percentile method, the current value at time t is ranked against all of the measurements made prior to t. The number of measurements with values less than or equal to the current value is divided by the total number of measurements. One minus this value is the percentage of the measurements that were higher than the current one. (If an increasing trend is a good result for the given metric, then a higher percentile is better than a smaller one. In the opposite case, the opposite interpretation is true.)

Min/Max: With this method we use only the current, minimum, and maximum measurements rather than every measurement made. The score given is simply the maximum minus the actual measurement, divided by the difference between the maximum and the minimum measurements. This gives us an interpolation of how far between the best and worst historical values the current measurement at time t is.

Of these three possible "ranking" methods, the percentile method is the most robust for most business metrics. It includes all of the measurements rather than just three in the case of the Min/Max method, and is not as susceptible to outliers. Compared to the Z-score, it does not require the constant mean or the normality assumption. Although being less effective when the distribution of values is clustered, the percentile method remains an easy concept for the layperson to understand. (For example, if the percentile is 75, then 75% of the historical measurements have been lower, and only 25% were higher. "Better" or "worse" depends on which direction is preferred for the given metric.) The percentile method is the one that we chose to use to rank our measures.

Note, however, that the percentile method "weights" each previous measurement equally. It assumes that the most recent measurements have the same value as ones that are from many years ago. In addition, it does not give any information about the distribution of the measurements. If the distribution is significantly skewed, the percentile could be misleading but then so too would the Min/Max and Z-score methods. The metrics that we address in this thesis are not significantly impacted by this assumption that the distribution of measurements is relatively symmetric.

If we now apply the percentile-ranking concept to the example started in the previous section, we can add 2 new columns to Table 4a to get Table 4b. The first is "n", the number of periods prior to t=200

against which t=200 can reasonably be ranked. Periods prior to t-n+1 are required for the first calculations of the global and 5-year trends. For all measures except the Model A MRV, n is capped at a maximum of 140 to add credibility to the ranking calculation. By the nature of the definition of Model A's most recent value, it only makes sense to compare it to all previous values on record. For the two LTT measures with n=116, this comparison allows for the 60 month long-term trend calculation, plus 24 months of additional development to ensure that the GT and LTT have had time to develop differently.

Measures	Value	n	Percentile
Model A			
Global Trend	0.97	140	71
Long Term Trend	0.29	140	19
Short Term Trend	1.00	140	40
Most Recent Value	190	200	99
Model B			
Global Trend	0.97	140	71
Long Term Trend	-0.71	116	17
Short Term Trend	0.03	140	45
Most Recent Value	-0.03	140	44
Model C (Projections)			
Global Trend	196	140	100
Long Term Trend	-0.05	116	28
Short Term Trend	0.02	140	64
Most Recent Value	0.00	140	14

Table 4b –Ranking Performance (t=200)

The last column in this table is the percentile of this current month's value with respect to all n previous values. For example, in Model A, the GT of a 0.97 unit increase per month at time t=200 is the 71st percentile of the 140 measurements made during the 200 months of data.

The interpretation of the 71st percentile is that 71% of the prior values have been lower than the one being ranked. Or, the converse, only 29% are higher. For a metric in which higher values constitute

negative performance, a "71" means that this global trend has been "better" than this in more than seventy percent of the previous measurable months.

If we look down the percentile rank column, it can be seen that Models A and B consider the GT to be above average, while Model C considers the GT to be one of the highest ever. This difference is due to using the projection method in Model C, and this GT measure is more similar in interpretation to the MRV in Model A than to the two slope related GTs. Looking at the 5-year trend, it is doing quite well in all three models.

The STT performance differs depending on the comparison made, ranging from the 40th percentile to the 64th percentile. This says that, in and of itself (Model A), the STT is better than many of the previous STTs. However, when first compared to the GT (Model B) or LTT (Model C), it is not performing quite as well.

The MRV in the three models each tell a very different story. Model A suggests the MRV is not doing very well; it is one of the highest values it has ever been (100 is the highest percentile possible). However, Model C says that the most recent value is better because it is in line with the STT, and is not currently causing any upward pressure on the metric. When the most recent month is compared to the GT average monthly increase, Model B, the deviation is slightly less than average, historically speaking.

All of these percentiles can be used to diagnose the significance of the most recent performance. However, before conclusions are made, we first need to draw lines that clearly distinguish poor from good performance. To do this we need "cut-off" percentiles that group the results into the red, yellow and green zones. This is discussed more in the next subsection.

Finally, it should be noted that this process of ranking the current measurement against all previous measurements can be applied at every period prior to t as well as at t (using only the data prior to each given date). The ranking of every prior period can be charted, giving a moving picture of performance rather than just the most current value. This is explored in greater depth later in this chapter.

4.2.5 Applying Percentile Cut-Offs

Now that we have a percentile value for each of the twelve measures, we need to make an assessment of this performance. In order to convert a number between zero and one hundred into the indicator light system (red, yellow, green) we need two cut-offs, a lower cut-off (LC) and an upper cut-off (UC).

The choice of the percentile cut-offs is quite arbitrary. Choosing them depends on the portions of the prior periods that should be represented as red, yellow and green. For metrics where higher values are better, the cut-off criteria (x, y) results in reds for percentiles lower than x, greens for percentiles higher than y, and yellows for all others. As an example, (33,67) would mean that results higher than the 67th percentile trigger a green light, while those lower than the 33rd percentile would trigger a red. Results in between the LC and UC trigger a yellow performance indicator light. As a simplifying assumption we take y = 100 – x, where x is between 0 and 50.

At this time, we could not identify an effective method to "optimize" the choice of these parameters, unless an expert from the organization provides a qualitative judgment of performance at every historical period. The error between the model and the expert's opinion can then be minimized. But, this would be a very cumbersome and impractical process when performed on all 4 measures, for all 3 models, for 120 or so periods, for numerous metrics. In addition, it is a very biased method and does not actually optimize the "accuracy" of the model, only the "acceptability".

In the models built, cut-offs can be adjusted until the "appropriate" or expected picture of historical performance is achieved. This first requires knowledge of the next two sections – calculating the overall indicator and charting historical performance indicator lights. It should be noted that some subjectivity is not detrimental to the modeling when used to set the model parameters and is, in fact, necessary. Choosing the parameters sets the sensitivity of the models. Once chosen, the parameters should not be adjusted too often and, to be changed, should require both owner's and executive's agreement.

Once again, if we continue with the sample series from Figure 8c, and add the columns "LC", "UC", and "Light" to Table 4b, we get Table 4c below. To evaluate the performance of each of the measures, we need to decide what portion of the prior values has been good and what portion have been poor. For simplicity, the lower and upper cut-offs are complements and we use the same values for all twelve

measures. For this example, we use (15,85) to illustrate our methodology. In practice, choosing these cut-offs requires an understanding of how the performance assessment should look (discussed in more detail later). A percentile measurement above the upper cut-off will trigger a red light in this metric. If the measurement is below the lower cut-off, a green light will result, while for a percentile between the cut-offs, a yellow light is triggered.

Measures	Value	n	Percentile	LC	UC	Light
Model A						
Global Trend	0.97	140	71	15	85	Yellow
Long Term Trend	0.29	140	19	15	85	Yellow
Short Term Trend	1.00	140	40	15	85	Yellow
Most Recent Value	190	200	99	15	85	Red
Model B						
Global Trend	0.97	140	71	15	85	Yellow
Long Term Trend	-0.71	116	17	15	85	Yellow
Short Term Trend	0.03	140	45	15	85	Yellow
Most Recent Value	-0.03	140	44	15	85	Yellow
Model C (Projections)						
Global Trend	196	140	100	15	85	Red
Long Term Trend	-0.05	116	28	15	85	Yellow
Short Term Trend	0.02	140	64	15	85	Yellow
Most Recent Value	0.00	140	14	15	85	Green

Table 4c – Applying Percentile Cut-offs (t=200)

From Table 4c, it is clear that all of the measures in all of the models at time t=200 are yellow, except for the MRV in Model A, the GT in Model C, and the MRV in Model C. The first two of these communicate the same result: they are both red and so they both indicate that the latest values are higher than they have historically been. The MRV in Model C, however, is green. This stems from the fact that the most recent month is currently in line with the short-term trend, something that has not happened as often as being "worse" than the STT.

The appropriateness of these results is debatable depending on the owner of the KPI; however, as we will see later, the owner will be able to calibrate the parameters (cut-offs) such that the model will be as sensitive as is desired. The next step, though, is to turn this detailed information (4 measures) into one light that triggers a response from the owner and the executive.

4.2.6 Overall Indicator

For each of the three models, there are now four lights for the most recent performance of this metric, each red, yellow or green, and each answer one of the four business questions. But, now we need an indicator light to signal the overall performance of the metric, using an aggregation of the four performance measures already calculated.

There are three obvious methods for combining the information from these four measures into one. The first is to use a simple "If/Then Statement". Another possibility is to use a "Weighted Discrete" approach, while a third is to use a "Weighted Continuous" method.

If/Then Statement: This first method is the most basic of the three approaches, as it only requires a logic rule such as "if at least one of the four measure indicators is red then the overall indicator will be red". Given such a rule, the owner and executive can easily determine whether the performance is poor because of the most recent month, the short-term trend, the long-term trend, the global trend, or some combination of these. However, this method uses simplified information about the performance in each model.

Weighted Discrete: A second possibility for formulating an overall indicator is to use a weighted discrete method. This approach requires that each performance light color be assigned a discrete value and that weights be chosen for each measure. Using this information a weighted aggregation can be calculated. For example, if red is assigned the value -1, yellow 0, and green 1, and if weights (summing to 1) are chosen for each of the four measures, the result is a number between -1 and 1. This new value will then require a new set of cut-off values to determine the color of the overall indicator light; however, this method would be overly complex and requires several additional parameters relative to the other two options.

Weighted Continuous: The third approach we considered involves taking a weighted average of the percentiles. This is similar to the "Weighted Discrete" method above but uses the percentiles rather than parameters representing the light colors. This method still requires a new upper and lower cut-off to determine the overall aggregate light color, but does not require the arbitrary assignment of parameter values for red, yellow, and green.

All three of these approaches can easily be programmed within a computer application such as the WCB's dashboard. From the business' point of view, the easiest to understand is the "If/Then Statement" method. However, in this thesis we use the "Weighted Continuous" because it uses the non-simplified performance information unlike the "If/Then Statement", while requiring fewer parameters than the "Weighted Discrete" method.

One assumption that must be made though is that there is no correlation between the measures being weighted. The appropriateness of this assumption is debatable because each of the perspectives is correlated with each of the others due to the impact of the same information - the most recent value. None of the above approaches are perfect; however, the results are effective, useful, and understandable.

The measure performance evaluation from the previous subsection can now be extended. To calculate the overall (or aggregate) performance indicator we need to decide on the weights for each measure. Table 5 below takes the percentile from each measure in Table 4b along with arbitrarily chosen weights, and then multiplies the percentiles with the weights and adds them up to get aggregate scores for each model. Once again, choosing these parameters requires some information about the true historical performance, and an approach to selection will be discussed in a later section.

Overall	GT		LTT		STT		MRV		Overall			
	%ile	Weight	%ile	Weight	%ile	Weight	%ile	Weight	Result	LC	UC	Light
Model A	71	0.1	19	0.2	40	0.3	99	0.4	63	35	65	Yellow
Model B	71	0.1	17	0.2	45	0.3	44	0.4	41	35	65	Yellow
Model C	100	0.1	28	0.2	64	0.3	14	0.4	41	35	65	Yellow

Table 5 – Overall Indicator Calculation (t=200)

For Model A, the resulting aggregate score is 63, while for Model B and for Model C it is 41. In order to evaluate these scores we again need cut-offs, but this time for the aggregate value. Now, because we are taking a weighted average of four numbers between zero and one, and our weights add to one, the distribution of these values over a large number of periods is likely to be clustered about the middle of the percentile range, around 50. This suggests that the cut-offs chosen will be much closer to the mid-point of the range than the previous choices of cut-offs, where we assume a more uniform distribution. For this example we have chosen (35,65) for our cut-offs. When these cut-offs are applied, all three models show performance at t=200 to be yellow. Here, yellow is an average performance. The metric is not doing poorly, and it is not doing well. The executive can use this information to make the decision to focus organizational resources on other more pressing performance metrics.

Once again, the overall indicator calculations can be applied to every period prior to time t, as well as the period at time t. This gives us the picture of the evolving performance, and in a continuously improving system, the light will always be green. The next subsection details how to chart this information, and what it looks like for the current example.

4.2.7 Charting Performance Indicator Values

Taking the example series from Figure 8c, we can chart the evolving percentile for each measure of each model from t=200 back as far as is measurable. These charts are useful when determining the cut-offs, as the effect of choosing different cut-off levels can easily be seen. In addition to charting the changing percentile, we can also chart the changing lights. This is particularly useful once cut-offs have been chosen since the executives may want to know how the metric's performance has changed over time. (For the Chapter V examples see Appendix D)

Charts can also be created for the overall performance indicator in the same way. Figures 9 a, b, c below display the evolving aggregate score against the metric's values. Model A shows that as the metric increases between t=100 and t=130, the aggregate score increases, then the model begins to get used to the trend and the aggregate score levels off and decreases. At t=160 when the metric begins its decline, the aggregate score drops, and as the KPI increases again at t=180, so does the score.

For Model B, the aggregate score reacts similarly, but more significantly than Model A. Model C, however, does something different. It is significantly more sensitive than the other two models to changes in the trend at t=100 and t=180, but within 12 months, if the trend remains, the score drops back to more "normal" levels.
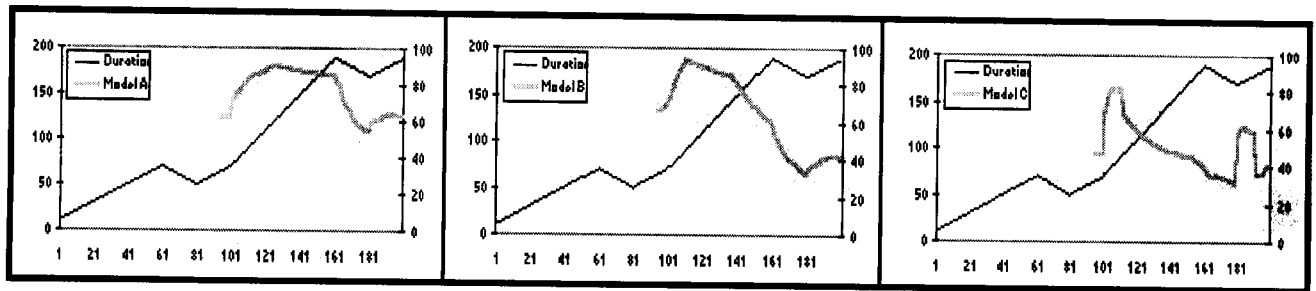

Figure 9a,b,c – Model A, B, C Actual Values and Aggregate Scores

Depending on where the cut-offs are chosen for the aggregate score, the proportion of red, yellow and green lights will change. By examining the charts in Figure 9, the effect of the cut-offs can easily be determined. For this example we have chosen the cut-offs (35,65), and the resulting charts are shown in Figure 10 a, b, c on the following page.

Model A begins with a yellow light for 2 months at period 96. It then turns into a red light and stays red until period 165, as the series dramatically increases. For the remaining 35 periods, the light is displayed as a yellow, as the actual values improve and then begin to deteriorate again. Model B shows a similar pattern, but the light actually turns green for 3 periods during the downward trend between periods 179 and 181. Model C behaves very differently from the other two in this example. The overall indicator quickly becomes desensitized to the increasing trend and the red light turns yellow at period 111. This result may or may not be acceptable to the owner, but given that action has not been taken to improve the performance after 14 periods, it might not be unreasonable to see the light turn yellow. The green light in Model C is also displayed for a longer period than Model B, lasting 16 periods.
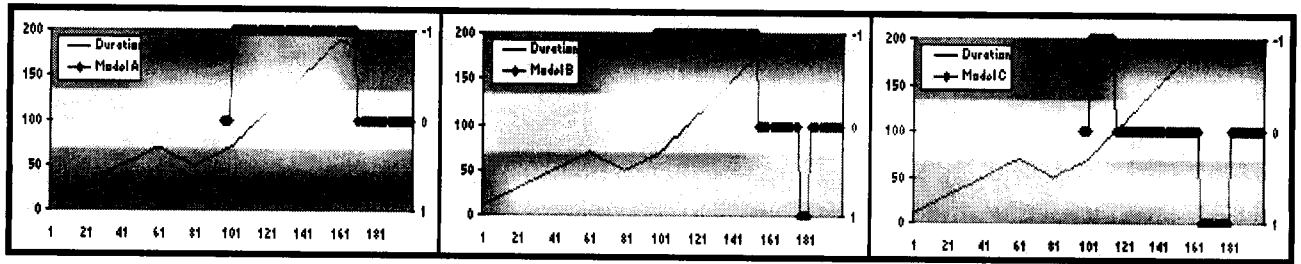
Figure 10a,b,c –Model A, B, C Historical Indicator Lights

What we are trying to identify is changes in behavior within the metric, and, in particular, changes in the trend. All three of these models attempt to do that, and, depending on the metric, any one of the models may be more appropriate than any other. The next section outlines methods for selecting the "best" model and its parameters.

4.3 Selecting a Model and Its Parameters

Selecting the appropriate model and the appropriate parameters (weights and cut-offs) for each metric is not a trivial task. There are many combinations of models and parameters, and it is not convenient to test every one of them. Even before we begin testing these combinations, we need to know what we are going to compare the results against. How do we know if one model formulation is better than another?

In the ideal situation, hindsight would be perfect and we would know what the overall indicator light should have been at every historical time period; performance would be known to be good, poor or average at any given prior month. Having this knowledge would allow the model builder to optimize the effectiveness of the model against the known performance. An optimization formulation could be used to minimize the difference between the model output and the actual known performance. Given that this is effective, the resulting model would likely be the best and most credible model for assessing new results over the short term, which is the primary purpose of our efforts in this thesis.

Unfortunately, in the business world perfect information is illusive, and perfect performance knowledge is not available for most business metrics. There do exist, however, opinions about performance, and these opinions can be both useful and misleading. In the case where there is no perfect assessment of historical performance available, we are forced to rely on the subjective

assessment of "true past performance". People with "expert knowledge" of the history of the metrics will define the "true past performance."

Given that the information we are going to use is not perfect, it does not make sense to create a formal process to minimize the error. Instead, it is more appropriate to choose the model and its parameters through ad-hoc testing, given general criteria about how the overall performance indicator should look. This is the process that is used later for the Chapter V results.

If perfect information is available or a more formal process is desired, however, then some decisions are required regarding the definition of optimal performance. If the model shows a red and it should be green, or vice versa, then there is definitely a problem. But, what if the model suggests performance is yellow and it is actually red, or if the model displays green and performance is actually yellow? And, what percentage of accuracy is considered effective?

If we refer back to the executive's interpretation of the red, yellow and green lights, we can note that in most cases red and yellow trigger an action in the organization, while green, however, does not. If the model signals that performance is acceptable (green) but in actual fact performance is poor (red), the executive will ignore a real problem and the poor performance will go unrecognized, potentially getting worse. This is a Type II error, and it is the most harmful error to the organization.

Intuitively then, it should be preferable for the model's performance assessment to experience a Type I error (rather than Type II), signaling that action is required (red light) when performance is actually satisfactory (green light). But, consideration must also be given to the credibility of the model and the efficient use of the organization's resources. If action is signaled too frequently or is required on too many KPIs, then the indicator light will be ineffective; it does not allow the executive to focus the organization's resources on what is truly important. In addition, without signaling credibility the organization will quickly become desensitized to its recommendations.

So, how do we measure this error? Well, to start with, the effectiveness of the model at assessing past performance can be calculated. If we assume that red and yellow lights result in an action in the organization, then we want to count the number of times that the model displays a green light when

performance is suspected to be otherwise. In addition, we want to calculate the converse problem, when performance is displayed as non-green when it should be green.

By adjusting the parameters in each of the three models (A, B, C) we either want to minimize the total errors (Type I plus Type II) or just the failures to signal the need for action (Type II). The latter choice is better if the organization would prefer to be extremely cautious. This resulting error measurement can then be used to choose the model that maximizes effectiveness, by minimizing the error. We leave this here as an opportunity for future research.

However, as mentioned earlier, although a mathematical process is possible, we decided at this stage of development to just adjust the parameters until a pattern was achieved that matched the owner's expectations of past performance. This calibration was ad-hoc, but quite effective. In Chapter V, we examine 4 typical WCB KPIs, and all of them were judged to be appropriate by the business.

By extrapolating from the advantages and disadvantages of each of the 3 models, a recommendation can be made as to which model is appropriate given specific conditions. Model B is useful when there is an established global trend that the business accepts as the norm. This model can be used to detect deviations from the accepted trend. Model C, on the other hand, appears to be the most sensitive of the three models. Because the MRV measures the deviation of the most recent month from the expected value of the 12-month linear trend projection, the model reacts more quickly to short-term results. This model is most appropriate when Model B is not, and the KPI is very stable, thus requiring sensitive measures. Model A, however, is the simplest model for the business to understand. Given that it appears to be as effective as the other two methods, it should be the default model whenever possible.

The caveat to these guidelines, though, is that the priority is for the pattern of historical lights to match the expectations of the "expert". If acceptable results cannot be achieved for a particular model, then it should not be used. If all of the models fail to meet the expectations of the business, it might be more appropriate to use only the target setting approach, as detailed in Chapter III.

Finally, the parameters and model choice should be reviewed annually. Part of this review should include a validation that the model was effective over the prior year. Specifically, the lights should have

turned the color they were supposed to, and they should have been neither too stable (never changing) nor too sensitive (changing every month). In addition, the lights for the individual measures should be reasonable and useful.

4.4 Limitations

Now that the models have been developed and the KPI's current and historical performance has been ranked and assessed, we need to recognize the limitations of this approach and of the models used in this analysis.

4.4.1 Limitations of the Approach

There are a few limitations to this performance assessment method that must be noted. Historical data may have interventions, be too volatile, or contain cycles, all of which result in significantly biased model results.

Interventions: Using an historical approach to assess performance requires that we assume that there are no interventions. In other words, we are assuming that there have been no dramatic changes caused by uncontrollable external influences during the measurement time frame.

An intervention can take several forms. It can be a one-month "blip" caused by a special circumstance, such as the effect on Sick-Leave Hours if the staff members in a call centre come down with food poisoning on "treat day" (all the employees bring in baked goods to share). Typically, this is considered an outlier and it could influence a 12-month moving-average. A more common intervention is a change in the natural level of the time series that occurs either suddenly from one month to the next or gradually over a number of months (an example might be a legislation change, such as an increase or decrease in coverage that results in a dramatic shift in average claim cost). The importance in recognizing an intervention is that it is a "one time event" that is not likely to occur again. Interventions are not part of the natural history of the metric and are caused by something outside the control of the organization.

If something has occurred in the time series, it is likely more valid to exclude the data prior to the change. However, this is only effective if enough data still remains for the analysis. Another option is to adjust the raw data for the change, but this is only realistic if the intervention can be accurately

quantified. If the data cannot be truncated and cannot be adjusted, it is possible that our analysis will simply not be appropriate.

Finally, it should be noted that although interventions interfere with the effectiveness of the performance assessment technique developed in this chapter, interventions themselves still contain valuable information for the organization. Many changes occur in the business environment, such as regulatory changes or changes in business practices, and recognizing the impacts that these interventions have had on the organization can still be valuable during decision-making. Intervention information must be communicated through the dashboard's analysis pages.

High volatility: When the smoothed time series is still too volatile, regression may be inappropriate. The best indicator of the effectiveness of the slope calculations is the Mean Square Error. The R^2 can be calculated easily for every trend measure at every historical month. Because we are ranking the most recent month's measures against every prior month, it is important that the appropriateness of all regression calculations is considered.

The variability of each of the metrics used in this thesis was not calculated; however, by visual inspection it can be seen that each of the smoothed data series have low volatility. In the case where a series remains highly volatile, linear regression would not be appropriate unless additional smoothing is performed first.

Cycles: If there are cycles in the data set, the ranking method may provide misleading results. On the upswing of the cycle, the trends will be highly ranked, while on the downswing the opposite would be true. These cycles can significantly bias the performance indicator, resulting in incorrect assessments.

For some WCB metrics, cycles are a natural response to cycles in the external business environment. If a cycle can be clearly identified within a business metric, it is better to remove it from the data series before applying the above modeling process. However, this is extremely complicated and is another opportunity for further research, as it is not a common problem in the WCB's KPIs.

Seasonality: In addition to considering cycles, there is also seasonality. Extreme seasonality can significantly impact a businesses operation. Because we are using a 12-month moving-average in our

analysis, the impact of seasonality has been removed. However, for some metrics, seasonality contains important information that can be used to make organizational decisions. There are various approaches that can be used to analyze and evaluate the deviations of the actual values from established seasonal patterns. Seasonality is another opportunity for further study.

4.4.2 Limitations of the Models

There are several limitations related specifically to the modeling. First, the Model B measures become unbounded as the global slope approaches zero, while the Model C measures become unbounded if any of the projected values approach zero. Second, for series with less than 10 years of data, the LTT becomes significantly less effective in all three models.

Unbounded At Zero: Model A is a simple model; it is never limited by values approaching zero, unlike Models B and C. Model B compares the slopes of the short term and long term trends with the slope of the entire data set, the global trend. The LTT and STT measures become unbounded as the global slope approaches zero, and this can happen for any month used in the Historical Ranking Process. Many of the WCB metrics face this issue early in the series development, but become more stable as the number of months used in the calculation increases.

Model C measures the deviation of the projected measurements. It captures the impact of both the slope and the intercept of each linear regression calculation. Unlike Model B where instability is caused by slopes approaching zero, this model becomes unstable when projections approach zero. Fortunately, this is not an issue for any of the current WCB Key Performance Indicators.

Limited Data: Given that for some business metrics fewer than 10 years' worth of data is available, in this case it is best to drop the LTT measure. The three models would then contain only the three measures: GT, STT, and MRV. This new model format can be used with as little as 3 years worth of data, although, once again, at least 5 years is strongly recommended.

V. APPLYING THE HISTORICAL RANKING PROCESS

We begin this chapter with a very brief review of the Historical Ranking Process developed in Chapter IV, and follow this up with the results of our model testing. To test the sensitivity we use the simplified STD Duration series (also from Chapter IV) and add various degrees of noise. Subsequent to this testing, we discuss the application of the Historical Ranking Process to several of the WCB's actual KPIs. We explore four different KPIs in this thesis: two operational metrics, a financial metric, and a human resources metric. We conclude this chapter with a brief discussion on the implications of our results.

5.1 Process Review and Model Sensitivity

5.1.1 Historical Ranking Process Revisited

First, let us begin with a brief review of the process developed in Chapter IV:

1) Calculate the GT, LTT, STT, MRV measures for each of the 3 models at time t=n, where n is the current period.

2) Calculate the 12 measures at every *possible* time t less than "n".

3) Calculate the percentile of each measure at time t=n against all values prior to n.

4) Apply cutoffs (x, 100-x) to these percentiles to get red/yellow/green lights for each measure.

5) Calculate a weighted aggregate score using the 4 measure percentiles in each of the three models:

(e.g. For Model A: the aggregate score equals $a_A{}^*G_A + b_A{}^*L_A + c_A{}^*S_A + d_A{}^*M_A$, where the weights $a_A + b_A + c_A + d_A = 1$ and G_A, L_A, S_A, M_A are the percentiles corresponding to the GT, LTT, STT, and MRV measures for Model A).

6) Apply cutoffs (y, 100-y) to the aggregate score to get the overall red/yellow/green light.

For each of the three models we have created there are six parameters (a, b, c, d, x, y). These can be set either before the analysis and based on a subjective opinion of the relative importance of each measure, or afterwards to calibrate the model output (the overall performance) to an expert's assessment of past performance. The latter choice allows for the creation of a model that is effective at assessing past performance; it is therefore assumed to reliably assess performance into the near future while using the same parameters. (This method is obviously limited by the expert's subjective evaluation of the "actual performance" of the KPI over the history of its measurement, and as mentioned previously, this introduces a bias into the modeling that will at some point have to be

examined more closely by the WCB. In the absence of a perfect assessment of what historical performance should look like, calibrating the model using an expert's assessment is the next best thing.)

Since we still have 3 models at this point, we need to select one. In some cases it is obvious even before choosing the parameters that either Model B or Model C is inappropriate. This can occur when the Global Trend slope or its projection approaches zero in any given prior month. If more than one model remains valid, it may be that one or more of the remaining models can be disregarded because they cannot effectively be calibrated using the expert's assessment of the KPIs performance.

Of the remaining models, only one should be chosen for the dashboard application. If more than one is still valid, the KPI owner should make the selection of the final model based on the ease of use of the measures. For example, the basic model (Model A) might be easier to interpret and explain. On the other hand, if the global trend (increasing or decreasing) is acceptable to the organization, then Model B might be more useful. In some cases it makes sense to choose the model even before the calculations are made. In this thesis, however, we present the results from all three models, and then leave the choice of the appropriate model to the executive or owner to be made at a later date.

5.1.2 The Impact of Noise on the Simplified STD Duration Series

Figure 11 below reviews the data series used in Chapter IV. We started with the STD Duration raw data - a monthly measurement. Smoothing the raw data using the business-standard 12-month moving average to reduce the noise and remove the seasonality, we present this time series as the actual KPI. However, in order to demonstrate our modeling process we arbitrarily simplified the smoothed series to create the easier-to-work-with "simplified series".
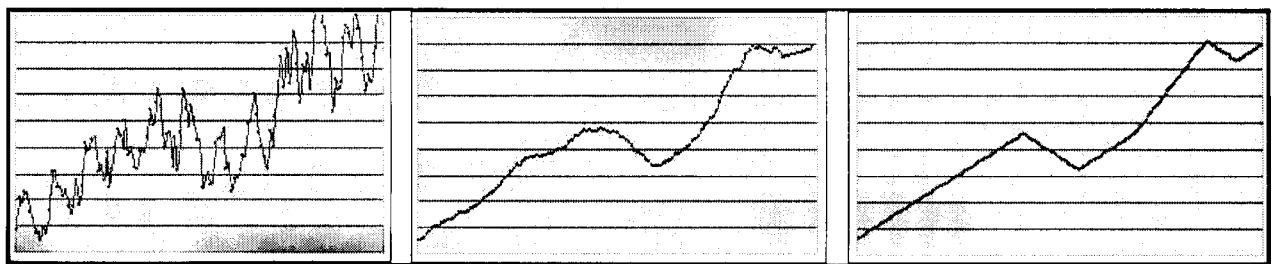
Figure 11 – "STD Duration" Raw Data, 12-Month Smoothed Series, and Simplified Series

Now, the simplified series is devoid of noise, and so this leads to an artificial stability of the Chapter IV modeling results. So, the next question is, Just how sensitive is this performance evaluation technique? How robust is the methodology? In order to test this, we added a random noise component at increasing levels to the simplified STD Duration series and then repeatedly simulated the changing series' results (Figure 12).

In adding noise to the three models, we used the random numbers γ, δ, and ε, where we defined γ as our "low noise level", δ as our "moderate noise level", and ε as our "high noise level". In the STD Duration Simplified Series the ranges chosen were -2.5 to 2.5, -5 to 5, and -10 to 10, respectively. We elected to generate our random noise using a Uniform rather than a Normal distribution for simplicity and to cap the range of possible values. If anything, using the uniform distribution is the more conservative choice, as the distribution of the deviations is not clustered around zero. So, to summarize, the noise we added to the simplified series was: $\gamma \sim U(-2.5, 2.5)$, $\delta \sim U(-5, 5)$, and $\varepsilon \sim U(-10, 10)$. Figure 12 on the next page illustrates a typical example of each new series.

The fifth graph in the figure (the "extreme noise level") was tested to gain insight into the effect of using the Historical Ranking Process on much more extreme data, such as a non-smoothed time series. It was created by multiplying the original data by ζ, where $\zeta \sim U(0.75, 1.25)$. Using a multiplicative factor allowed the level of noise to increase as the time series increased, and combined with the size of the parameter, it created an increasingly volatile series.
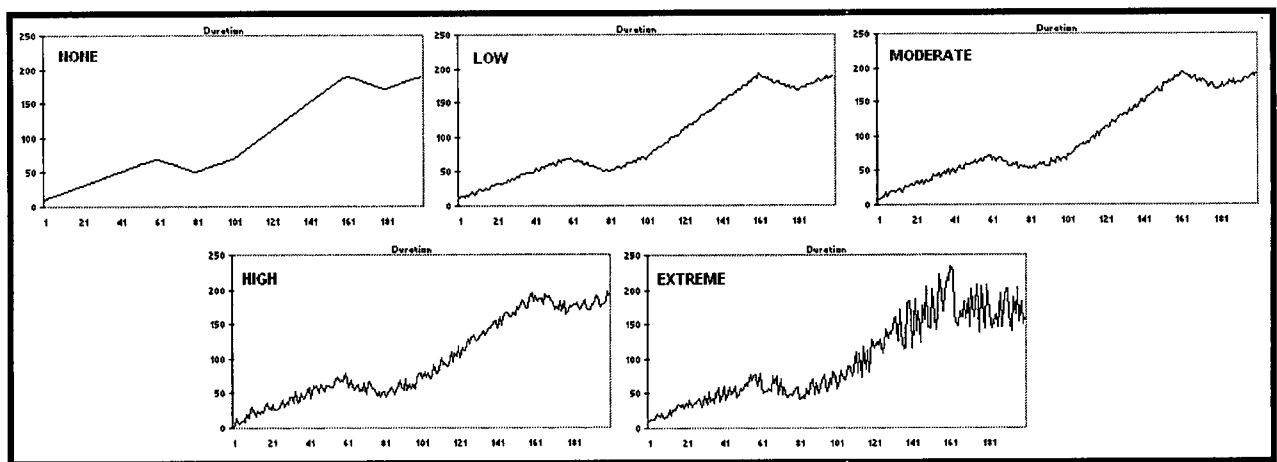

Figure 12 – Simplified "STD Duration" Time Series with Varying Levels of Noise

Testing the impact of the four levels of noise on the Simplified STD Duration weighted aggregate score series and overall performance lights yielded the following typical results, illustrated in Figures 13a through 13e. For simplicity, the weights and measure values are not included in this thesis, but were chosen to approximate the expected historical performance pattern. The weights are not the same across the three models.
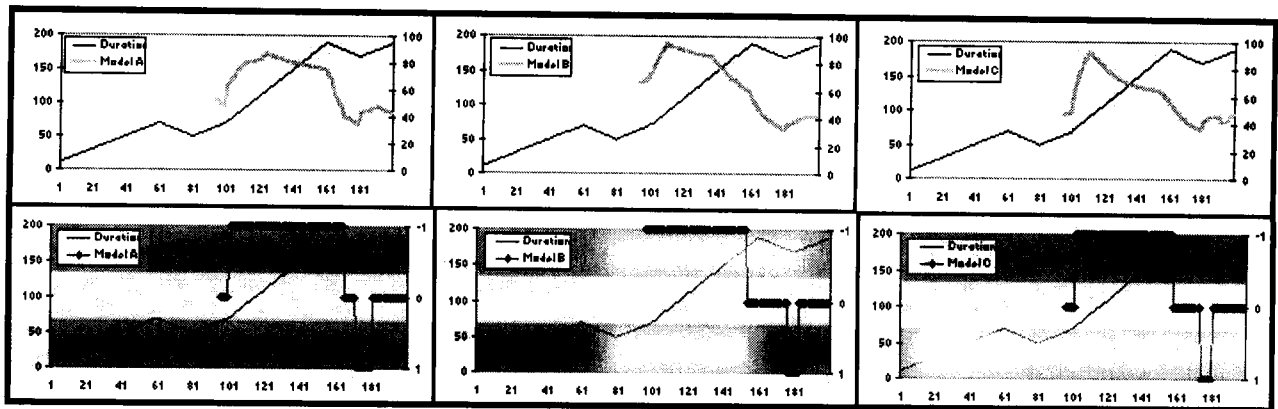

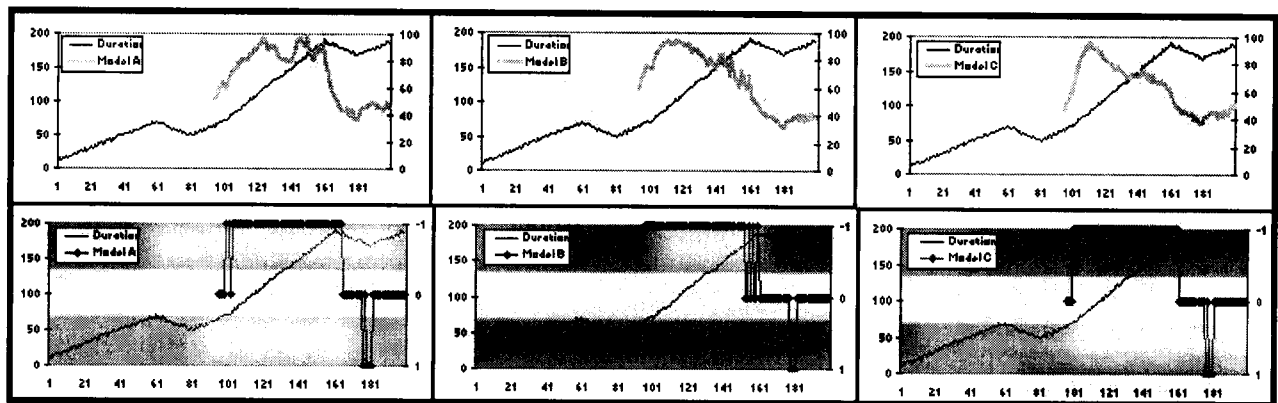**Figure 13a – Aggregate Percentile - Zero Noise Level, Same Parameters Simplified Series**

Figure 13b – Aggregate Percentile - Low Noise Level, Same Parameters Simplified Series
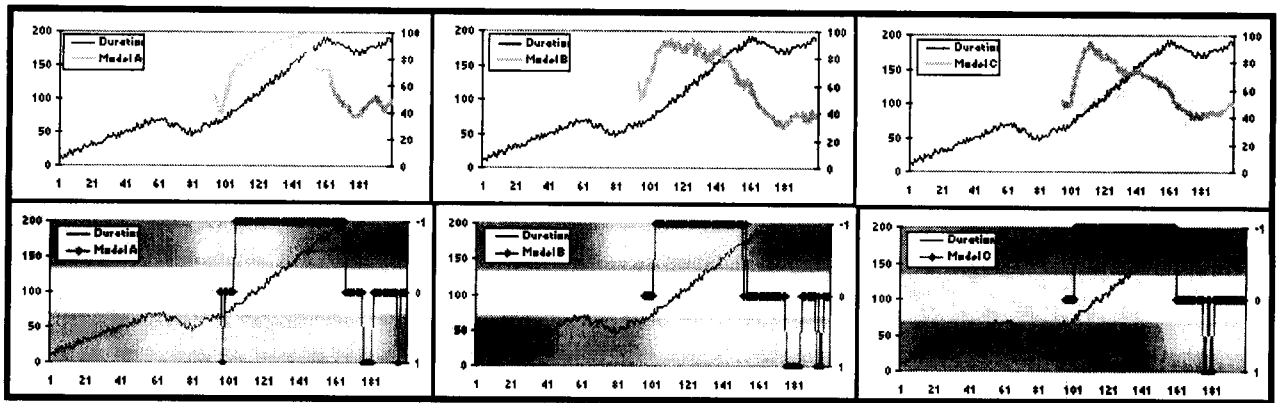
Figure 13c – Aggregate Percentile - Moderate Noise Level, Same Parameters Simplified Series
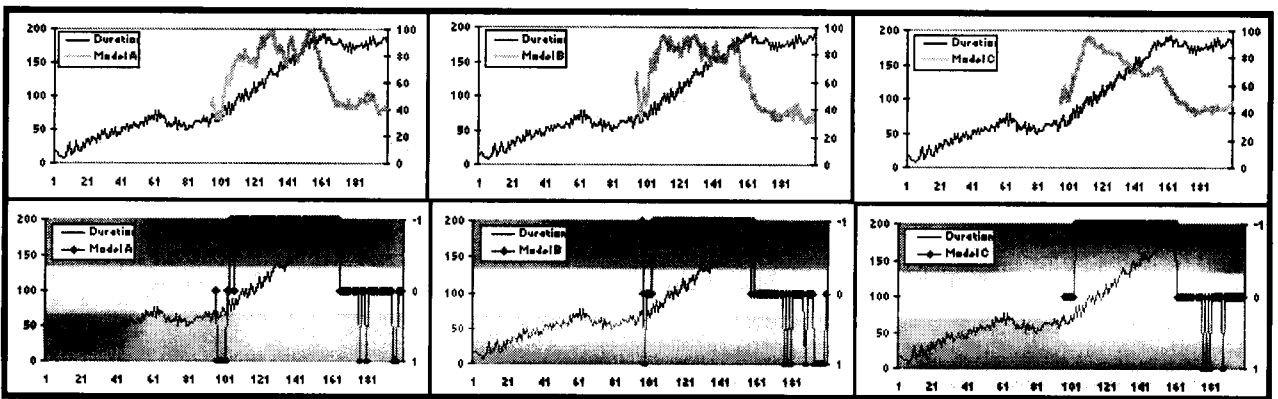

Figure 13d – Aggregate Percentile - High Noise Level, Same Parameters Simplified Series
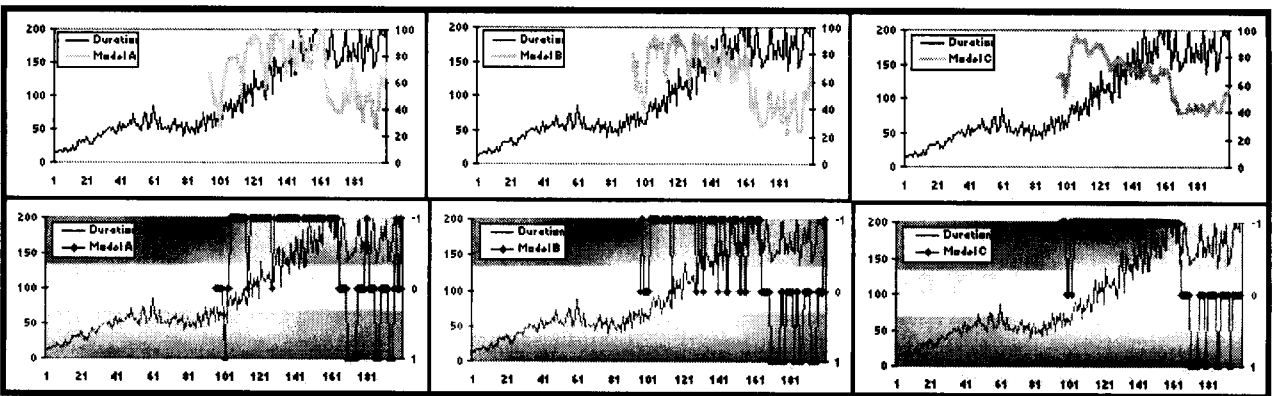

Figure 13e – Aggregate Percentile - Extreme Noise Level, Same Parameters Simplified Series

From these charts, it can be seen that the performance indicator lights for the three models becomes more volatile as the noise in the series increases, but that even up to the high level of noise the

resulting progression of lights does not look unreasonable. In fact, the general pattern of red, yellow, green, yellow does not change at any but the Extreme Noise Level.

However, it must be noted that two of these models (A and C) use only a 10% weight on the MRV measure, and that as this weight is increased, the volatility of the aggregate performance lights becomes much less acceptable. In general, the more noise a time series has the less useful the MRV measure becomes, but, that being said, the Historical Ranking Process itself will remain effective. In addition, given that we are using a 12-month smoothed series for all of our business metrics, we did not come across a case where the noise has exceeded that of even the moderate level (the four metrics in the next section are representative of this).

Note that in the above discussion we have focused on the impact of the noise on the aggregate scores and the overall performance lights; however, it is also possible to examine the effect that noise has on the individual measures: Global Trend, Long Term Trend, and Short Term Trend. One possible approach is to use the coefficient of variation. We do not explore this in any depth in this thesis, other than to note that our conclusion is the same as for the sensitivity testing of the aggregate score. Our conclusion, then, is that the Historical Ranking Process is effective and reasonably robust, but as the noise grows, the MRV measure becomes less useful and its impact (weight) on the aggregate performance measure should be reduced.

5.2 Historical Ranking Process Applied to Other WCB KPIs

So, let us now take a look at several of the WCB metrics, beginning with the actual "STD Duration" series. The subsequent examples exhibit characteristics different to that of the nicely increasing STD Duration metric. The KPIs chosen for this chapter are from the operational, human resources, and financial perspectives of the WCB Balanced Scorecard. At the time of this thesis all of the WCB's Customer Service KPIs were only being tracked on an annual basis (something that the WCB was looking to amend in the future) and thus were not appropriate.

5.2.1 "Average Short Term Disability Claim Duration" (Operational)

The "Average Short Term Disability Claim Duration" (STD Duration) can be defined as the average number of workdays lost due to a work-related temporary total disability (injury or illness) experienced by BC workers per claim. A number of factors can influence the performance of this metric, such as

the administration time of processing claims, the effectiveness of rehabilitation activities, and the changing attitudes of workers to immediately returning to work (perhaps influenced by a poor economy).

In addition, due to the complexity of this metric, changes in the distribution (or mix) of underlying factors can also significantly impact the overall result. Specifically, it has been shown that the significant reduction of less-severe claims (short duration) over the past 12 years, combined with a less dramatic reduction in more-severe claims (high duration) has meant that the average duration has steadily increased. Although the results have appeared poor for many years, the organizational strategy to address the administratively costly shorter duration claims means that at least a portion of the dramatic and steady increase in the average duration is acceptable. Since we don't know exactly how much of the increasing trend is attributable to acceptable causes, we still expect that the indicator light will indicate poor performance for much of the historical measurement of this KPI.

The figures in Appendix D capture the results from the application of the Historical Ranking Process to the WCB's STD Duration data. The evolving performance of each of the 4 measures (the Global Trend, Long Term Trend, Short Term Trend and Most Recent Value) is illustrated as a new time series for each of the three models. It can be seen, for example, that the Model A GT starts at zero and progresses up to around 70. This means that the Global Trend has gradually deteriorated over the past 4 years moving from a definite green light to a yellow, and may even be a red in the most recent months, depending on where the measure cut-offs are chosen. However, we will not go into any further details regarding individual measure performances here, but we include the charts in the Appendix for those interested.

Given that we now have the historical percentiles for each measure, we next applied weights to these percentiles to get the aggregate score time series. When we apply the aggregate level cut-offs to this new series we get the representative historical performance lights, the ultimate outcome of the Historical Ranking Process (Figure 14).
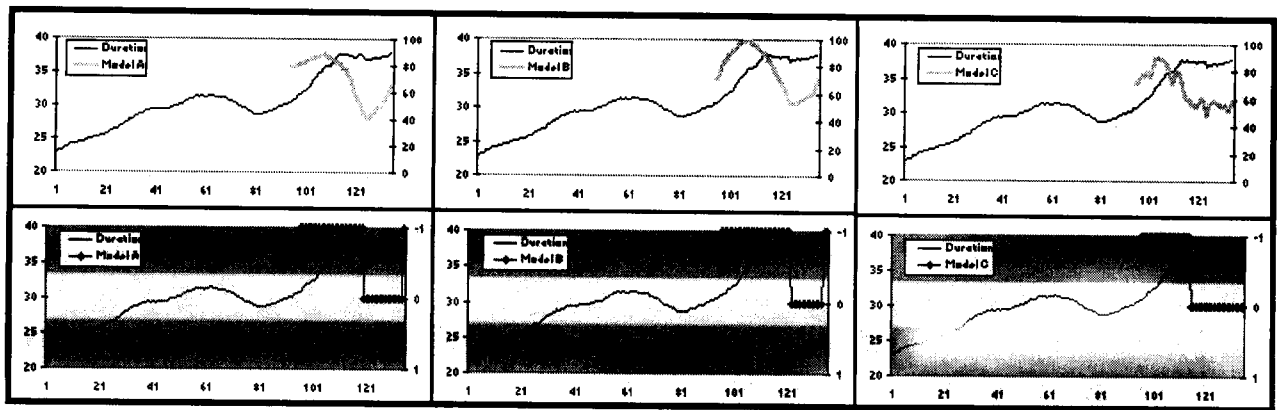
Figure 14 – Actual "Average STD Claim Duration" Series Results

It was by adjusting these parameters that we were able to duplicate the expert's assessment of performance (in this case, just red during the increasing trend and then a yellow once the series stabilizes). It could be said that this set of performance lights is too stable, but it is by choice and can be made more sensitive by increasing the weight on the two short-term measures.

Now that all three models have been calibrated to produce a similar outcome, one must be chosen. For this particular KPI the increasing trend is generally acknowledged, and so Model B might be the better model to select. Each of Model B's measures is calculated relative to the Global Trend. And so, Model B with the selected parameters can now be used to monitor performance for the next year, at which time the new data should be considered and the model and its parameters revised. The performance light for the current month, then, would be red.

5.2.2 "12-Month Moving STD Injury Rate" (Operational)

The WCB's "12-Month Moving STD Injury Rate" KPI tracks the total number of BC claims that received a first short-term disability payment during a 12-month period, per 100 person years. Most of the monthly values are based on a rolling weighted-average estimate of exposure (the total number of people employed in BC in any given month), while the December value each year is calculated using the actual measured exposure. Injury Rate is one of the most important KPIs used by the board; it is a key indicator of the effectiveness of their prevention activities – the heart of the board's mandate.
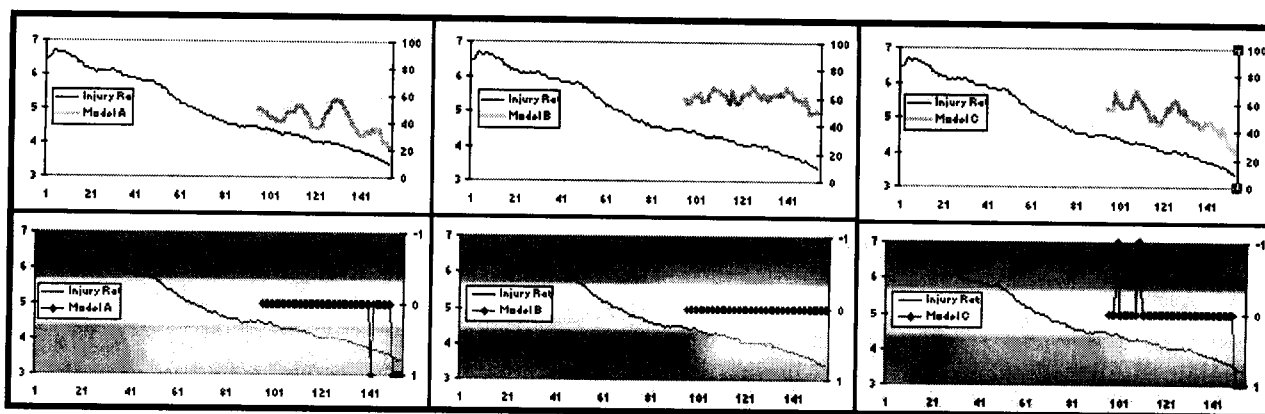
Figure 15 – Actual "12-Month Moving STD Injury Rate" Series Results

This time series (Figure 15) is an example of a very stable decreasing time series. Because of this, the Injury Rate KPI is a potentially controversial metric to apply the Historical Ranking Process to. For this series, many people would suggest that, because the metric continues to "get better", the aggregate light here should always be green, and certainly never red. However, the light will actually be yellow for most of the measurable series due to the rapid decrease early on. In fact, the light may even turn red as the series begins to decrease at a less rapid rate or levels off. At such a time as this occurs, the performance assessment methodology may need to change. It is obvious that this series cannot continue on its decreasing trend forever, and this is a limitation of the process developed. Ideally we want "continuous improvement", but with some metrics that may not be a practical requirement, and this KPI is one such extreme case.

Despite the above, we still apply the method as the most recent year is still decreasing at a rapid rate, and there is no sign of a "leveling off" yet. In this series, yellow lights would be displayed for many months, and most recently, the light would turn green as the Injury Rate decreases faster, and this fit with the "expert's" conception of past performance given our methodology. During the parameter selection, it was determined that Model B was not sensitive enough and would probably be impractical for this KPI despite the acceptance of the long-term trend. Model C can be used here due to its sensitivity, and so the current month performance indicator light would be green.

5.2.3 "Relief of Costs Ratio" (Financial)

The "Relief of Costs Ratio" is defined as the percentage of claim costs that were left unallocated to a specific company or industry. For the majority of claims, the business that the injured party works for is known; however, for some claims (like late developing occupational disease claims) a company may

have gone out of business. This KPI is important because it is an indicator of the effectiveness of the "correct premium for the given exposure" principal of insurance.

This particular time series is not as stable as the previous two, and it exhibits dramatic shifts in the short-term trend over time. The "expert's" expectation of historical performance for this KPI was that the first decrease of the time series should show a green light. As the values begin to climb higher again, the light should progress to red, then back to yellow during the decrease, and red again during the last 12 months of increasing values. We chose the parameters for each of the three models accordingly, and the results are displayed below (Figure 16). Model A can be used for this KPI, and the display for the current month would be a red light.
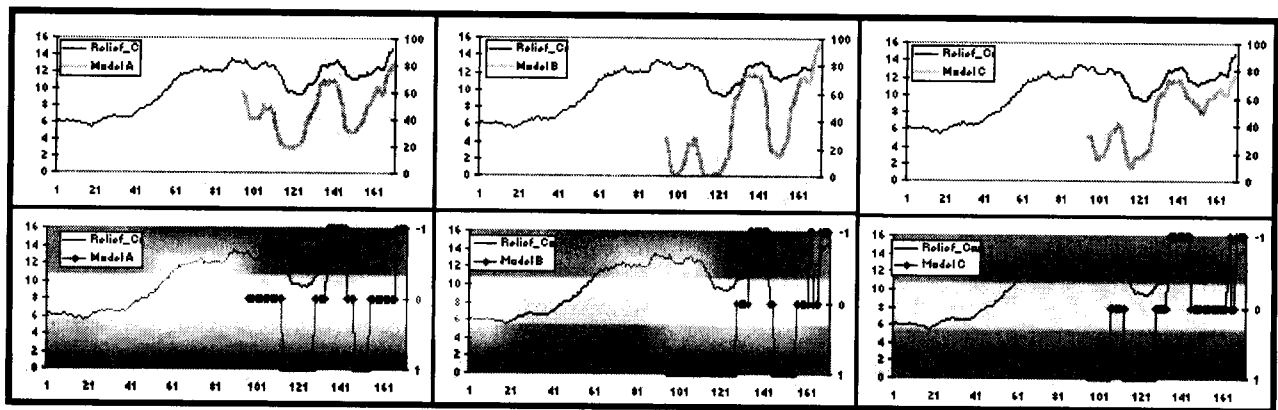

Figure 16 – Actual "Relief of Costs Ratio" Series Results

5.2.4 "Sick Leave Hours" (Human Resources)

"Sick Leave Hours" is a human resources KPI and is defined as the average number of hours per WCB employee per pay period taken off for sickness, medical, dental, or family. This KPI is important to the organization because it is an indirect measure of employee productivity.

The data available for most of the human resource metrics is limited by the fact that a new information system was introduced in 1999. There is significantly less data available from this system than is available from the data warehouse. The other complication with this metric is that all of the human resource data at the WCB is tracked on a pay period basis rather than monthly, and there are 26 pay periods in a year. The smoothing of the series becomes a 26-pay-period moving-average function, but removes seasonality and reduces noise in the same manner as the 12-month moving-average.
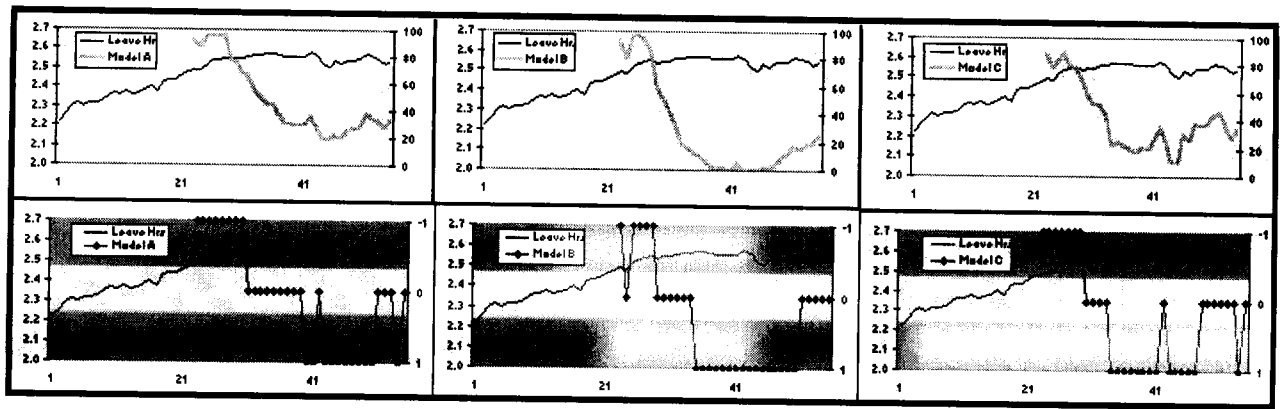
Figure 17 – Actual "Sick Leave Hours" Series Results

Because of the reduced volume of data, we did not use the Long Term Trend Measure (LTT), as it requires at least 6-years of data. The GT, STT, and MRV performance charts, meanwhile, can be found in Appendix D, just as with the past three KPIs.

Using weights with the three measures produced the aggregate score chart above. The pattern is similar to the expert's assessment: red during the increase, yellow as the series levels off, green as the performance is maintained, and then yellow as performance becomes expected. There are some "blips" in the data series, and this causes a pay period to temporarily display another color. These "blips" could be the result of noise, but should be investigated by the KPI owner, and explained within the dashboard analysis. Given that Model A is effective, the current month would be displayed as a yellow light.

5.3 Results

So, what does all this mean and, how and why is it useful to the executive?

What we have created is a performance assessment method that only requires a KPI's historical data. We generate a useful assessment of how today ranks against yesterday, and it uses not only the most recent month's performance but also the short and long-term trends. In addition, it is automated.

The overall performance light is the trigger for the executive's attention. If the light is red, the executive must deal with it; if it's yellow, time permitting, it is dealt with; if it is green, the executive may choose to only review it every couple of months or so. Given that a red light exists, the executive will want to know why it is red. Is the poor performance driven by a single month's performance, a

71

developing short-term trend, or a deteriorating long-term trend? The measure lights effectively communicate the areas of concern.

The key to this method is to have an explanation for performance, whether good or bad. It is possible that performance can be explained by non-controllable factors, in which case this should be noted in the analysis. However, if there is no external explanation for deteriorating results, corporate initiatives are needed for results to improve. If no initiatives exist for a KPI or the existing ones are judged to be insufficient, additional work may be needed by members of the organization to improve the performance of the KPI.

Given an aggressive, continuous improvement strategy, the overall light should remain green at all times as both trends and results get better and better. In addition, in a well-established and well-managed system there will never be any red lights, as performance issues would be corrected during the yellow light – the pre-emptive strike.

So, let's take a look at the specific examples from this chapter. Table 6 below summarizes the lights displayed in the executive dashboard. The overall light is shown first, as it is the initial trigger for the executive. The four measure lights are used to define exactly where any problems or successes lie.

	STD Duration	STD Injury Rate	Relief of Costs	Sick Leave
Model	Model B	Model C	Model A	Model A
Aggregate	Red	Green	Red	Yellow
GT	Yellow	Green	Green	Green
LTT	Red	Yellow	Yellow	N/A
STT	Yellow	Green	Red	Yellow
MRV	Yellow	Green	Red	Yellow

Table 6 – Summary of KPI Performance Lights for the Current Month – Four KPIs

The interpretation here then is that the executives would first be drawn to the STD Duration and Relief of Costs KPIs, as they are both red. If there is time after exploring these, the Sick Leave Hours should be investigated. The Injury Rate is green and would probably not attract the executive's attention.

Given that the executive first examines the STD Duration, he can immediately see that the Long Term Trend (5 year trend) is very poor relative to prior 5-year trends in this metric. The MRV, STT and GT are less of a concern, but as can be seen from the series they are still not doing well. There must be a reason for this (there is, as previously mentioned), and it should be evident in this KPI's summary. For this KPI, there should also be projects under way to improve performance, and this information should also available from the dashboard application.

The Relief of Costs KPI also has a red overall performance light for the current month. In this case, the overall light is being driven by the MRV and STT. Given that the GT is green, and the LTT is yellow, this is a short-term problem and should be fixed immediately, before it develops into a long-term trend. This KPI must have initiatives put into place to improve performance.

Sick Leave Hours is yellow. This KPI is of a moderate priority, and, if examined, it can be seen that there are no red lights for the individual measures. Finally, the Injury Rate indicator is green, and it also has no measures with red lights. These last two KPIs may or may not have initiatives in place to improve their performance; but focusing on the improvement of the STD duration and the Relief of Costs would be a more effective use of the organization's resources.

VI. CONCLUSION

6.1 SUMMARY

Committed to the continuous improvement of their business decisions, the executives at the Workers' Compensation Board of British Columbia initiated a project in 2002 to improve the use and communication of their existing Balanced Scorecard. A significant portion of this improvement was the result of transferring the existing paper system to the WCB's local Internet environment while concurrently redesigning the business processes used to generate and represent the information.

One such improvement was the introduction of performance indicator lights (red, yellow, and green) to communicate the performance of each Balanced Scorecard metric (or "Key Performance Indicator") and whether executive involvement is required. Two approaches were explored. First, a methodology was developed to aid in the setting of the targets required to trigger these lights. Second, a process was created to evaluate the latest month's value using only the available historical data.

The purpose of using targets is to measure the organization's KPIs against an expected level of performance, or planned value. This expectation may be simply based on the projected trend or on an aggressive improvement or benchmarked goal. In some cases, even deterioration relative to past values might be expected (e.g. the result of poor economic conditions). If the organization is performing better than expected, the indicator light would be green; if not, it will be either yellow or red, depending on the severity of the deviation from plan. Choosing the expected performance level should use a combination of approaches: forecasting, improvement goals, and benchmarking.

Setting targets is effective when the organization wants to evaluate the most recent performance of a metric against where the organization wants to be. However, this does not answer how the metric has performed relative to its past values, nor does it acknowledge nor evaluate the trends within the data. To address this we introduced the Historical Ranking Process. Utilizing short-term and long-term trends, we capture trend information at every possible historical point and then rank the most current measurement against every prior measurement. This trend evaluation can then be combined with an assessment of the most recent month to give us an aggregate score for performance. Finally,

by applying an upper and lower cut-off we can assess the performance of the metric using the three levels or zones: red, yellow, and green. This complicated but intuitive approach can easily be automated, and it is a useful assessment of where the organization is relative to its past.

There is value in using the information from both the Target Setting Process and Historical Ranking Process to evaluate the performance of a Key Performance Indicator. As a quick example, a KPI may be performing well (green) relative to the organizations targets, but poorly (red) due to adverse trends. The opposite might be true too, if, for example, the targets are set too aggressively.

This project will significantly improve strategic decision-making and organizational performance review at the WCB. At a quick glance an executive can now determine if strategies are working or not. The information is from a central and single source, and several layers of detail and analysis allow the executive to explore only as deeply as desired.

6.2 FINAL THOUGHTS

The creation of the Balanced Scorecard in the late 1990's was a big step toward a better understanding of the effectiveness of the WCB's strategic goals. The creation of the dashboard application and the development of a performance assessment methodology was the next, and these will continue to evolve. The concepts and ideas put forth in this thesis reflect the prototype methodology, and it is expected that as the WCB begins to fully utilize the dashboard application it will learn and build on the foundation we have created. KPIs will be added, removed, and modified; performance targets and Historical Ranking Process parameters will be adjusted; even the structure of the application itself will adapt as the business environment changes.

Over the past 10 years effective and efficient performance assessment has been a popular topic in both the for-profit and non-profit business worlds, and its importance is expected to continue. The dashboard application is a significant investment by the WCB and it is seen as a vital tool that will be used by all members of the organization for years to come. Balanced Scorecards are being implemented every year in large businesses. They are common because they take the ever-increasing complexity and volume of business data and convert it all into manageable information that can be used to make both strategic and tactical decisions.

Going forward there are many opportunities for the WCB. With the continued development of their data warehouse, the effectiveness of the dashboard tool can only improve. As the volume and scope of the data increases this tool will become more powerful. But, the challenges will become how to remain focussed on only what is critical to the organization's success; how to avoid information overload; and how to optimize the effectiveness of the tool.

Once all members of the organization are comfortable with the dashboard, departmental Scorecards can be created and automated, using metrics that are tied into the executive's Balanced Scorecard KPIs. Performance targets and historical ranking similar to that of the executive dashboard can be used to qualify results as good or poor. The benefits of consistency, accountability, improved turnaround, and clear performance communication can be taken down to the next levels of the organization.

In addition to departmental scorecards, there are also cross-functional teams working on complex projects vital to the organization's strategic goals. A set of metrics chosen using the Balanced Scorecard methodology can be tied into the KPIs and can be included in the dashboard as project scorecards. Eventually, it will be transparent to every person how their efforts influence the organization, and how they each contribute to the success of the business as a whole. The goals of the individuals can be made synergistic with the goals of the organization.

Inspired by the need to know "how are we doing?" the design and implementation of a dashboard application to contain and communicate the organization's Balanced Scorecard has opened the door to an exciting new field of research and employment at the WCB. The quantitative representation of the WCB's strategic direction, and the qualitative communication of performance are areas that will continue to be explored and developed in future years. As this knowledge is discovered, shared and accumulated, the organization can only get stronger.

BIBLIOGRAPHY

Adam, Frederic and Jean-Charles Pomerol 1996; "Critical Factors in the Development of Executive Systems – Leveraging the Dashboard Approach"; University College Cork, Ireland.

AWCBC (Association of Workers' Compensation Boards of Canada), May 15, 2002; <http://www.awcbc.org/ >

Balanced Scorecard Collaborative May 5, 2000; "Balanced Scorecard Functional Standards" Release 1.0a; Balanced Scorecard Collaborative, Inc.

Balanced Scorecard Collaborative May 15 2002; <http://bscol.com>

Ballard, Tanya N. February 15, 2002; "Bush management scorecard gets a green light"; Government Executive.

BC Ferries 2001; "2001 Annual Report"; <http://www.bcferries.bc.ca/ >

Henrich, Karen May 23 2000; "Seagate Software Helps Insurance Corporation of British Columbia Drive Its Business Strategies"; Crystal Decisions Press Release; <http://www.crystaldecisions.com/about/press/releases/2000/052301.asp>

Housing Corporation (The) May 15 2002; <http://www.housingcorp.gov.uk>

Housing Forum (The) May 15 2002; <http://www.thehousingforum.org.uk/hf/about/kpis.asp>

Hunt, H. Allan 2002; "Why Not the Best? Service Delivery Core Review Report": W.E. Upjohn Institute For Employment Research.

Hunt, H. Allan, Peter S. Barth, Michael J. Leahy 1995; "The Workers' Compensation System of British Columbia: Still in Transition": W.E. Upjohn Institute For Employment Research.

Kaplan, Robert S. and David P. Norton 1992; "The Balanced Scorecard – Measures that Drive Performance": Harvard Business Review 71.

Kaplan, Robert S. and David P. Norton 1996; "Translating Strategy Into Action – The Balanced Scorecard": Harvard Business School, Boston, Mass.

Kaplan, Robert S. and David P. Norton 2001; "The Strategy Focused Organization"; Harvard Business School, Boston, Mass.

Kersnar, Janet July 1999; "Re-Inventing The Budget"; CFO Asia; <http://www.cfoasia.com/archives/9907-34.htm>

Macfarlane, Victoria and Angela Weltz 1998; "Royal Commission on Workers' Compensation in BC – Performance Indicators Final Report" <http://www.qp.gov.bc.ca/rcwc/research/macfarlane-weltz-indicators.pdf>

Makridakis, S., S.C. Wheelwright and R.J. Hyndman 1998; "Forecasting Methods and Applications (third edition)"; John Wiley & Sons Inc.: New York.

Marr, Bernard and Andy Neely May 15, 2002; "Balanced Scorecard Software Report"; <http://www.som.cranfield.ac.uk/som/cbp/BScorecard.html>

Mayor, Tracy October 1, 2001; "Red Light, Green Light"; CIO.

Minnosota DOT 2002; "Target Setting Framework, Performance Measures, Targets and Policy Guidance"; Moving Minnesota 2003.

Miyake, Dylan July 26, 2002; "Beyond the Numbers"; Intelligent Enterprise.

Neely, Andy, Michael Sutcliff and Herman Heyns May 15, 2002; "Driving Value Through Strategic Planning and Budgeting"; Cranfield School of Management and Accenture; <http://www.som.cranfield.ac.uk/som/cbp/Research_Report.pdf>

Niven, Paul R. May 15 2002; "Cascading the Balanced Scorecard: A Case Study on Nova Scotia Power, Inc."; <http://www.primerusconsulting.com/pdf/Cascading.pdf>

Qayoumi, Mohammad H. November 2000; "Balanced Scorecard"; California State University; <http://www.calstate.edu/qi/Agenda/Nov-00/QIPresentationPP.pdf>

QPR, QPR ScoreCard, May 15, 2002; <http://www.qprsoftware.com/>

Senior Executive Committee May 1999; "Advancing the Plan: A Review and Restatement of the WCB Strategic Plan".

Sklarew, Arthur 1982; "Techniques of a Professional Commodity Chart Analyst": Windsor Books.

TickQuest May 15 2002; NeoTicker EOD; <http://www.tickquest.com/NeoTickerEOD/>

Workerplace Safety & Insurance Board Ontario, May 15, 2002; <http://www.wsib.on.ca/wsib/wsibsite.nsf/public/home_e>

WCB of Alberta, May 15, 2002; <http://www.wcb.ab.ca/>

WCB of British Columbia, May 15, 2002; <http://www.worksafebc.com/>

WCB of Saskatchewan, May 15, 2002; <http://www.wcbsask.com/>

WizeTrade May 15 2002; WizeTrade; <http://www.wizetrade.com/index.html>

Wordnet Dictionary 1997; <http://www.dictionary.com/>

GLOSSARY OF TERMS

Balanced Scorecard – A methodology proposed by Kaplan and Norton that recommends organizations take a more balanced view of their performance. Reducing their emphasis on backward-looking financial metrics, and balancing them with 3 forward-looking perspectives: Customer, Learning and Growth, and Operational.

Dashboard – Usually a computer application or intranet tool, a dashboard is an automated performance tracking, reporting, and communication medium. Commonly, dashboards utilize data from data warehouses and summarize the information in charts, graphs, and performance lights.

Executive – A senior director of the organization, usually the President or a Vice President. For each KPI in the Balanced Scorecard, there is an executive that is ultimately responsible for its performance.

Initiatives – Projects initiated by the organization to improve performance.

Intranet – An organization's internal Internet.

Key Performance Indicator (KPI) – Metrics that are used in the Balanced Scorecard. Only a select number of business metrics are chosen for the Balanced Scorecard. All KPIs are metrics, but only some metrics are KPIs.

Measure – Defined for the purpose of this thesis as a calculation performed on the data used to generate a business metric. An example is a linear regression slope calculation.

Metric – Defined for the purposes of this thesis as any business measurement used to monitor the performance of some process in the business. An example might be the number of injuries that occur in a given year per 100 employees.
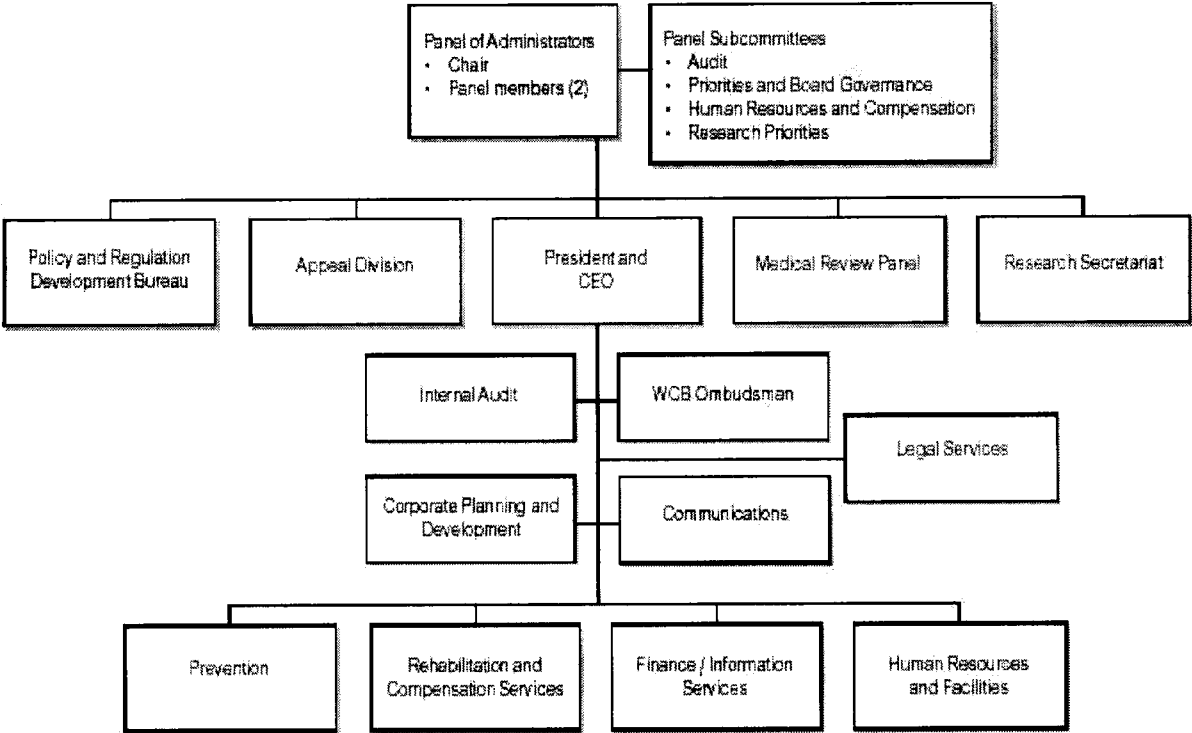
Owner – The Balanced Scorecard methodology requires that each KPI have an owner that is accountable for performance. The owner is also responsible for updating the analysis and the initiatives recorded on the dashboard. In some cases the owner will be an executive.

Performance Light – A circular indicator used in the dashboard application to represent the performance of the KPI. The light is red, yellow or green, which represents poor, average and good performance levels respectively.

Perspective – Part of the Balanced Scorecard Methodology, a perspective is a grouping of KPIs. There are usually 4-5 perspectives in an organizational Balanced Scorecard, the most common being: financial, operational, customer, and learning and growth (or human resources).

Target – A target is a value that is set by the organization for a particular metric. Targets can be set using historical trends, improvement goals and/or benchmarks against other organizations. Targets are tied closely to the strategic objectives of the business. The goal is for actual values to meet or exceed the targets.

APPENDIX A: WCB of BC Organizational Structure


```
┌─────────────────────────┐  ┌──────────────────────────────────┐
│ Panel of Administrators │  │ Panel Subcommittees               │
│  • Chair                │  │  • Audit                          │
│  • Panel members (2)    │  │  • Priorities and Board Governance│
│                         │  │  • Human Resources and Compensation│
│                         │  │  • Research Priorities            │
└─────────────────────────┘  └──────────────────────────────────┘
```

| Policy and Regulation Development Bureau | Appeal Division | President and CEO | Medical Review Panel | Research Secretariat |

| Internal Audit | WCB Ombudsman | Legal Services |

| Corporate Planning and Development | Communications |

| Prevention | Rehabilitation and Compensation Services | Finance / Information Services | Human Resources and Facilities |

APPENDIX B: Metrics from the WCB Monthly KPI Report

#	KPI Title	Perspective
1	Injury Rate	Operational
2	STD Timeliness & Income Continuity	Operational
3	VR Timeliness	Operational
4	Timeliness of New Loss of Earnings	Operational
5	Active STD Claims Inventory	Operational
6	Total Duration by Month	Operational
7	VR Return to Work Outcomes	Operational
8	Disallow Rate	Operational
9	Appeals Received	Operational
10	Review Board Allow Rate %	Operational
11	Injured Worker Satisfaction Survey	Customer
12	Ombudsman Cases	Customer
13	Ombudsman Cases by Referral Source	Customer
14	Public Contribution Index	Customer
15	Funding Status	Financial
16	Placeholder - Percent Funded by RG	Financial
17	Average Premium Rate	Financial
18	Statement of Operations	Financial
19	STD Payments Distributed by Days Paid	Operational
20	HC Payments by Major Category	Operational
21	VR Payments by Major Category	Operational
22	LTD Pension Liability (New Discount Basis)	Operational
23	Balance Sheet	Financial
24	Cash Flow Analysis	Financial
25	Administration Expenses by Function	Financial
26	Permanent & Temporary Full Time Equivalents	HR
27	Income Components	Financial
28	Claims Costs	Financial
29	Total Compensations Index for Unfinalized Claims (FRCC)	Financial
30	Unfinalled Claims Liability Indices	Financial
31	FSTD Payment Index	Financial
32	Claims Costs Index	Financial
33	FRCC Index	Financial
34	Value of Investment Portfolio	Financial
35	Improve the Work Climate - Non Management	HR

#	KPI Title	Perspective
36	Improve the Work Climate - Management Divisional	HR
37	Improve the Work Climate - Employee Assistance Usage	HR
38	Relationship with the CEU	HR
39	YTD Distribution of Paid Hours	HR
40	Diversity of Workforce	HR
41	Injury Rate	Operational
42	STD Claims	Operational
43	STD Claims - Construction	Operational
44	STD Claims - Health Care	Operational
45	STD Claims - Wood & Paper Products	Operational
46	STD Claims - Metal & Non-Metallic Mineral Products	Operational
47	STD Claims - Forestry	Operational
48	STD Claims - Hospitality	Operational
49	Traumatic Fatalities: Forestry & Construction	Operational
50	Fatal Accident Investigation Report Details	Operational
51	Prosecutions	Operational
52	Administrative Penalties Imposed	Operational
53	Financial Reports - Balance Sheets, etc.	Operational

APPENDIX C: KPIs from the 2001 Annual Report

Performance Highlights

	2001 Target	2001 Actual	2002 Plan
For those we serve			
Decrease injury rate	3.9	3.7	3.6
Reduce average total claim duration	49.5 days	50.8	49.8
Improve return-to-work outcomes for all workers with a permanent functional impairment	New measure	75%	82%
Improve timeliness of initial claim acceptance	14 days	16 days	16 days
Raise injured workers' rating of service (10-point scale)	8.0	7.8	8.0
Raise public contribution index	70%	67%	65%

	2001 Target	2001 Actual	2002 Plan
In our finances			
Achieve an accident fund in the range of 110–115%	106%	105%	99%
Rate groups between 95–105% funded	85%	45%[1]	To be determined by actual results
Attain an aggregate premium rate between $1.75 and $2.25 (per $100 of rateable payroll)[2]	$2.01 ($1.77 after transition and surplus)	$1.98 ($1.77 after transition and surplus)	To be determined by actual results

	2001 Target	2001 Actual	2002 Plan
With our staff			
Improve the work climate every year (5-point scale)	Staff: 3.5 Managers: 3.7	Staff: 3.4 Managers: 3.7	No survey during collective bargaining year
Reflect within our workforce the diversity in our community	Complete Phase 2 (Recruitment, Training, and Mobility) of the Employment Systems Review	Phase 2 completed December 31; outcomes to be studied	Development of implementation strategy

## APPENDIX D: Performance Time Series' for GT, LTT, STT, MRV

Each of the following charts tracks the changing percentile rank of the individual measures for each given model. There are 12 charts for each Key Performance Indicator or metric; there are three models (A, B, C), and then four measures for each model (Global Trend, Long Term Trend, Short Term Trend, Most Recent Value). As a series approaches the top of the graph (100) performance is considered to be deteriorating (poor), while a series approaching the bottom of the graph (0) is improving. If we were to apply cutoffs to these charts, the top portion would trigger a red light, the middle a yellow light, and the bottom a green light.

5.2.1 "Average Short Term Disability Claim Duration" (Operational):

Model A
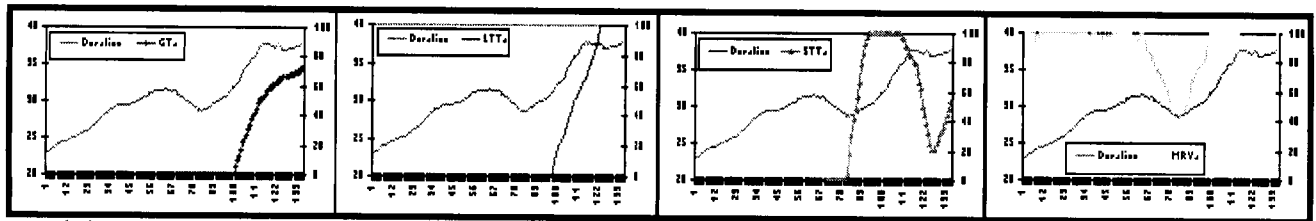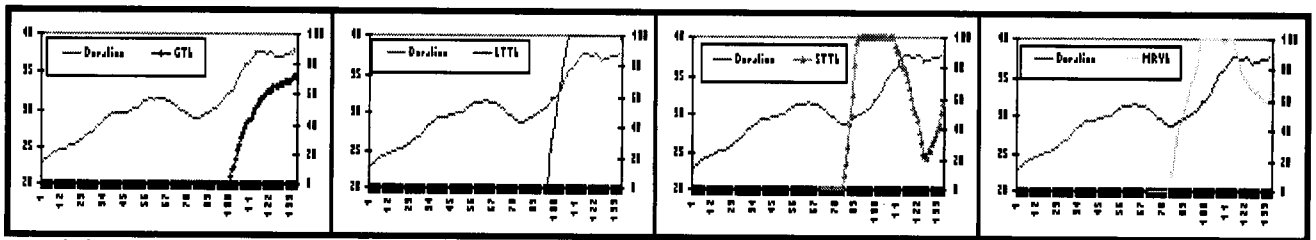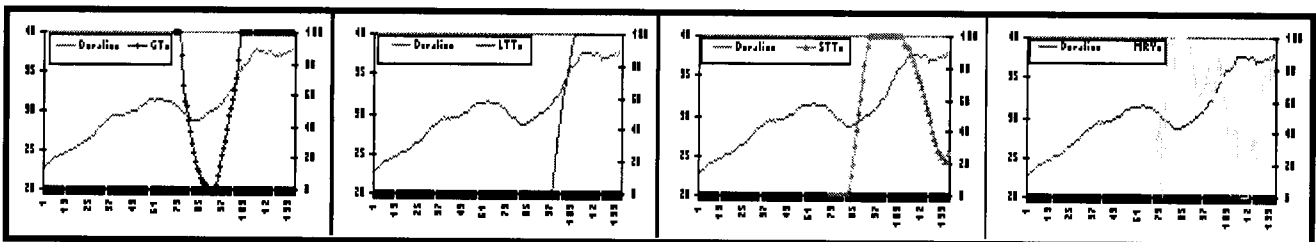
Model B

Model C

5.2.2 "12-Month Moving STD Injury Rate" (Operational):
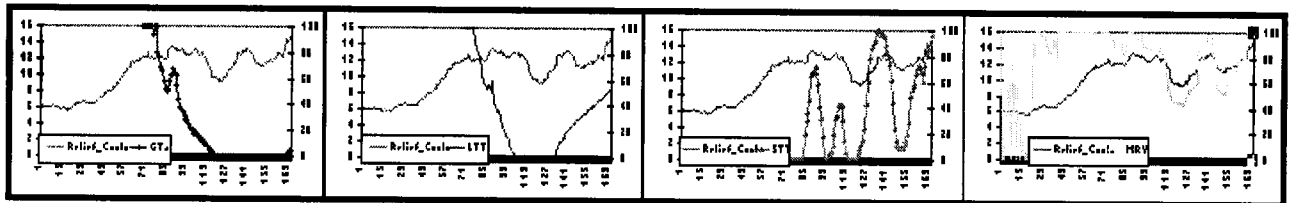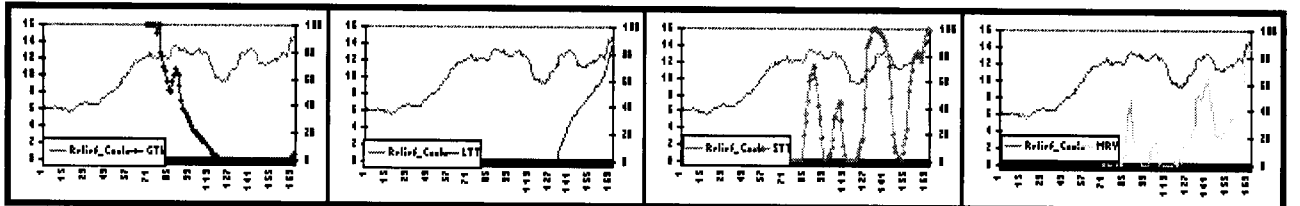
Model A
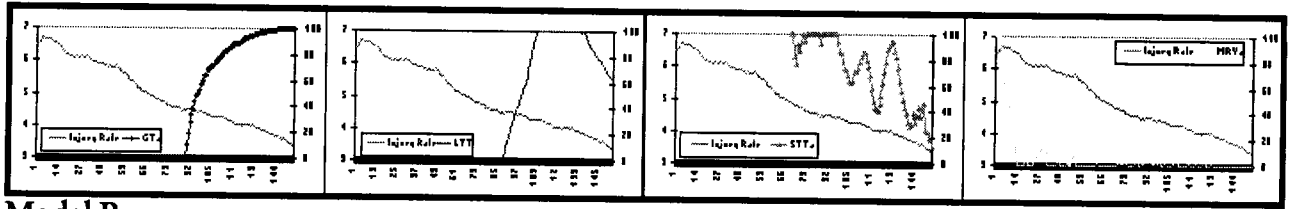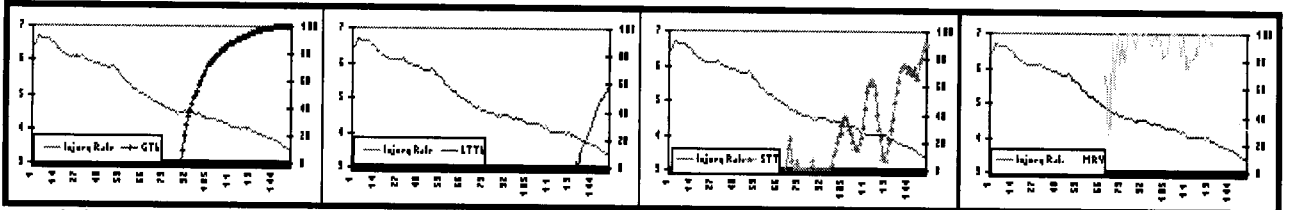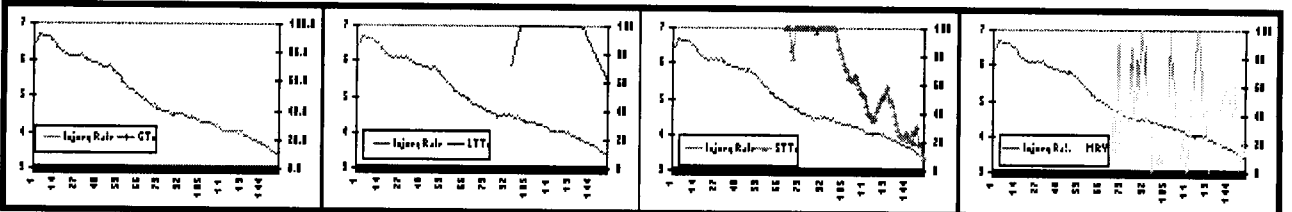
Model B

Model C

5.2.3 "Relief of Costs Ratio" (Financial):

Model A

Model B
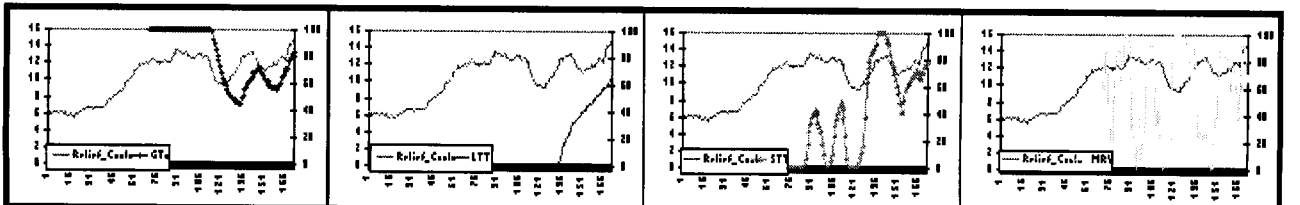

Model C

85

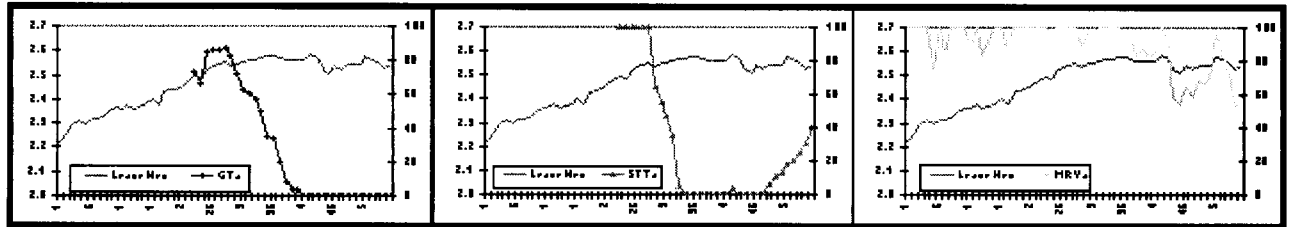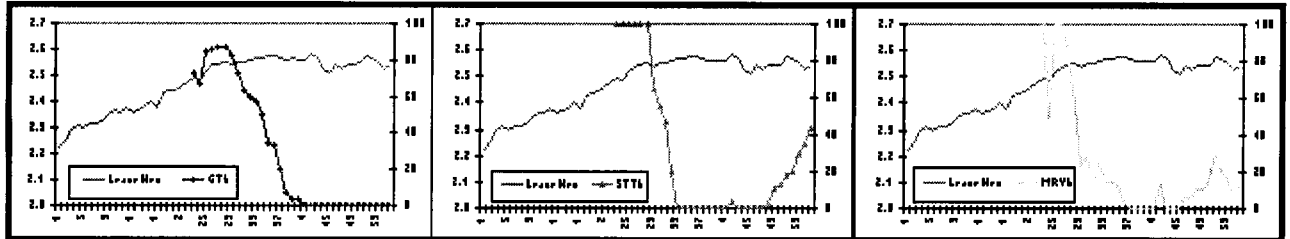## 5.2.4 "Sick Leave Hours" (Human Resources):

Model A

Model B

Model C