

*The Benefits and Challenges of Profiling
Providers in Workers Compensation*

Nancy R. Treitel, FCAS, MAAA, ARM, CPU,
Miriam Perkins, ACAS, MAAA, and
Bart Margoshes, M.D.

The Benefits and Challenges of Profiling Providers in Workers Compensation

Abstract

This paper provides a general overview of ways in which provider profiling can be used in developing, maintaining, and evaluating workers compensation managed care programs. It discusses some of the practical challenges that actuaries face in actually developing such profiles and using them. Specifics covered include determining the types of statistics one might want to review, creating the appropriate database needed to do the analysis, and adjusting and segmenting data so that differences in the types of claims providers handle are taken into account. Provider profiling in WC is relatively new to the managed care world. This intent of this paper is to introduce actuaries to provider profiling.

Authors: Nancy R. Treitel, FCAS, Liberty Mutual Group
Miriam Perkins, ACAS, Liberty Mutual Group
Dr. Bart Margoshes, M.D., Liberty Mutual Group

1. Introduction

Due to the significant rise in medical inflation in the late 1980's and early 1990's in workers compensation, cost containment strategies patterned after the group health model were developed¹. These strategies included utilization management, medical case management, and discounted provider networks². In theory, providers would be willing to accept a discounted fee for an increase in patient referral volume. Providers were chosen based on anecdotal information from the insurers, a willingness by the provider to treat workers compensation patients, and an agreement to accept a discount. Objective criteria based on treatment outcomes were not part of the selection process. As this model has matured and information systems have become more sophisticated, insurers and managed care companies are beginning to analyze provider performance and outcomes. The result of this capability is that insurers are beginning to select providers who objectively provide consistent high quality, cost effective care and are forming provider networks that are more limited in size but more focused on outcomes.

Profiling providers in the workers compensation system has some significant challenges and differs from the group health system. There are three major areas of divergence. In group health PPO models, a primary care provider is selected by the patient based on the group health contract. This primary care provider acts as a gatekeeper for care. In workers compensation, the selection of the provider is made by the employer (directed), the employee (nondirected), or a combination of the two. The issue of choice is based on the individual state's workers compensation laws. Another area of divergence is the fact that group health outcomes are generally based on discrete, time limited episodes of care. In contrast, in workers compensation, an episode lasts the life of the claim which can be years long. Finally, group health outcomes are limited to medical care only. Since both the medical and indemnity costs are the responsibility of the carrier in workers compensation, a provider's performance (effectiveness) must be based on both these costs.

This paper provides a general overview of ways in which provider profiling can be used in developing, maintaining, and evaluating workers compensation managed care networks. It will then discuss some of the practical challenges that actuaries face in developing these profiles. Specifics covered include determining the types of outcomes one might want to review, creating the appropriate database needed to do the analysis, and adjusting and segmenting data so that differences in the types of claims providers handle are taken into account. Since provider profiling in workers compensation is relatively new to the managed care world, this paper is intended to introduce actuaries to provider profiling and to stimulate discussion as to ways to enhance techniques in place today.

The benefits of provider profiling are plentiful. These include fine tuning provider network composition, providing a feedback tool for quality improvement, training and education, and improving the quality and cost effectiveness of the care delivered to the

injured worker. The challenges, however, one faces in building a provider profiling system are substantial.

2. The Concept of Provider Profiling

The basic concept is to use an insurer's existing data repositories to rank individual providers or provider groups relative to one another. This concept is akin to the idea of experience rating, with insurers trying to determine which subset of providers are better than average by reviewing historical experience to the extent credible. Insurers also may look at other risk characteristics (e.g. location of providers to customers, quality of provider's staff) to supplement experience based information similar to what is done in the risk selection process when deciding which employers to insure. Experience rating and risk selection across employers enables insurers to rank employers and enables the insurance company to charge a price commensurate with expected cost. In the case of provider profiling, the process includes:

- selecting and maintaining superior networks and panel lists of providers (analogous to determining the quality of an insurer's book of business);
- rewarding better than average providers (analogous to charging insureds different prices based on experience);
- using the findings in quality assurance and provider education (analogous to risk management and loss prevention);
- using the findings with customers to encourage channeling and to differentiate one managed care program from another; and
- maintaining a competitive edge.

3. Data Considerations

A large practical problem in the process of provider profiling is securing relevant data for the task. In the group health environment, studies done are largely based on clinical information¹¹. Typically, a single type of illness is reviewed (e.g. myocardial infarctions), and clinical data covering single episodes of treatment are gathered. Because medical insurance covers calendar year periods of time and because people may change health insurers and health care providers over time, group health studies find it difficult to track patients and their corresponding medical treatment over long periods of time.

In workers compensation, medical coverage is provided for the life of the injury. Indemnity benefits are also often tied to the duration of the injury. Workers compensation, therefore, has the benefit of having information that tracks results by claimant (e.g. patient) over time, something not available in the group health world. All is not ideal, however. While information is captured over time in workers compensation, it is often difficult to determine exactly when an episode ended. Claim closure date is often used as a proxy but that date depends on more than just the medical condition of a claimant. In addition, clinical data at the provider or medical bill level is not something

that has historically been tracked by claimant in workers compensation. In order to profile providers in workers compensation, specialized databases need to be created.

The primary databases available for workers compensation are those developed in conjunction with statistical agents/rating organizations to support the ratemaking process, i.e. those based on unit report data, financial aggregate data, or the sample of detail claim information (a.k.a. DCI data) collected by statistical agents/rating organizations from insurers. In addition, companies typically have reserving databases with workers compensation data used in analyses done to determine ultimate liabilities for the company and estimates of incurred but not reported losses (IBNR).

The above mentioned databases are designed for ratemaking and reserving and do not contain information pertaining to medical case management. Instead, medical management information may reside on a variety of systems not traditionally accessed by actuaries. Two of the first challenges actuaries face when trying to profile providers are 1) developing the list of desired items to measure (see next section of paper) and 2) determining what data exists at the provider level so the list can be narrowed down to those items that can realistically be measured. In some instances, determining what data is available may mean going to the managed care vendor used by the insurance company (e.g. PPO network administrator or medical bill review company). In other cases, a company may already retain the detailed medical information in internal systems. Once finding out the data that is available, the actuary will likely need to request that a special database be built for provider profiling: one that contains provider transaction level detail, as well as medical and indemnity loss outcomes on each claim.

4. Performance Measurements

Given the complexity of the medical encounter in workers compensation, it is a challenge to select those provider performance outcomes that have relevance to quality and cost effectiveness. In order to narrow the list of potential outcomes it is important to determine at a high level what the insurer values as indicating cost effective, quality care. In this paper, we focus on treatment quality, return to work outcomes, medical outcomes and patient satisfaction. Another factor in determining what outcomes to measure is the ability to obtain the data and the data integrity. Since several disparate databases contain the relevant information, this decision process can be formidable.

Treatment Quality

There is no single statistic that perfectly measures the quality of treatment provided by providers. Possible measurements for treatment quality include relapse rates (reopened claims), litigation rates, closure ratios, the percentage of cases referred to nurse case managers, referral rates to a specialist, use of diagnostic studies, average number of visits, use of prescriptions, and degree of documentation of the medical record. A brief discussion of some of these indicators follows.

The provider often plays a pivotal role in determining whether or not a claimant returns to work. High levels of claims reopening may be indicative of claimants returning to work too soon. This is not a desirable outcome. Since the provider is not solely responsible for the return to work outcome, this measure, like most, is not a perfect measure of quality. Additionally, carriers may not always track reopenings in such a way that the information is usable in developing provider profiles.

Managed care programs which encourage providers to get involved early in a claim and to play an active role in facilitating a patient's return to work are thought to have an influence in an injured worker's decision to litigate or not. The theory is that injured workers receiving high quality medical care will be less apt to litigate. Thus high levels of litigation would reflect poorly on a provider while low levels of litigation would reflect well. How highly correlated litigation rates are to provider quality is open to debate. Some may argue that litigation is driven more by employer-employee relations. Even if one reaches agreement that providers play a role in litigation, companies may find it a challenge to accurately measure litigation rate. Insurers often have statistics on their own use of attorneys but information on claimant's defense attorneys is often incomplete.

Higher closure ratios are better than lower closure ratios when profiling providers. High closure ratios indicate earlier return to work and efficient medical care. As long as they are not accompanied by higher relapse rates, high closure rates are considered to be a favorable measure of quality. Of course, in addition to the provider, the injured worker, employer, and insurer play a role in the injured worker's return to work so again, it is not a perfect measure. On the positive side, this statistic is often one that an insurer can easily measure.

Some indicators more consistently represent positive behavior than others. Other indicators may be hard to interpret. For example, high levels of nurse case management on claims handled by a provider may be a good thing or a bad thing depending on the reason for the nurse's involvement. Higher levels of this indicator could mean that the insurer was compelled to bring the nurse in due to the claim not being resolved as anticipated, or in contrast, it could reflect a proactive doctor who involves the nurse by choice to facilitate the injured worker's return to work. Referral to a specialist could be a positive result if it brings experience and knowledge into the process at an early point, or a negative result if the case has deteriorated and higher intensity care is needed. Well documented provider records are an indicator of quality but may present some practical problems in incorporating into a formal ranking procedure.

Return to Work Outcomes

Possible measurements for return to work outcomes include the number of temporary total disability days (TTD days), the ratio of temporary partial disability dollars to temporary total disability dollars, the indemnity cost per claim, the total cost per claim, the percentage of return to modified duty claims, the percent of all claims that are lost time, and the percent of claims where return to work is within published guidelines.

These measures all relate to the success in getting the injured worker back to work. Lower TTD days are desirable. Light duty is encouraged and a higher ratio of temporary partial disability days to temporary total disability is a sign that the provider is making use of light duty programs (again, many others also play a role in returning a worker to light duty). The percentage of claimants returning to modified duty is another measure along these lines. Companies typically track different types of payments being made (e.g. temporary total, temporary partial) and the length of disability for which these payments are being made within their claim systems. However, some programming may be needed to turn the information into usable output. Adjustments may also be needed to account for waiting periods. How to deal with lump sum settlements when trying to estimate the duration of disability is yet another challenge.

Total indemnity costs (possibly excluding fatal and permanent total claims, due to their low frequency and high severity) are a sign of dollars spent but may be heavily influenced by factors other than provider treatment. Assuming each provider has a similar profile of claims, low levels of claims that are lost time may indicate that providers are getting claimants back to work within the waiting period (state statutes typically have a 3 or 7 day waiting period before indemnity benefits get paid). Both these statistics are generally easily accessible within company systems.

Measuring the success of providers in returning injured workers back to work within specified guidelines is desirable but may not be practical. First one needs to have guidelines in place. There are several sets of guidelines commercially available for a price (e.g. Presley Reed, Milliman & Robertson). Then one needs to consistently track how a provider does relative to the guideline on a claim by claim basis. This requires keeping sufficient detail to know that a claim's injury type is consistent with that of the guidelines. The guidelines also often give a range of disability days rather than a single point estimate so one needs to decide how to deal with the ranges.

Medical Outcomes

Possible measurements for medical outcomes include medical cost per lost time claim, medical on medical only claims, the percent of cases referred to physical therapy, and the average number of office visits.

Potential concerns with medical outcome indicators include the fact that high dollars on a medical only claim could underlie a situation where a lost time claim was avoided, perhaps by the physician authorizing an employee's return to work under a modified duty program. The avoidance of an indemnity loss is a positive outcome, yet it would increase the physician's average severity on medical only claims. An offsetting result could be a decrease in claim severity on the indemnity side, if the physician's behavior reduces time lost from work in general. Thus, the measurements are not independent and a positive behavior would not necessarily result in a positive result across all measurements. The calculation of an overall score should take these interdependencies into account, through

the weights assigned to the measurements. Also of concern when measuring medical outcomes is the difference in medical treatment costs between communities. A state may need to be subdivided into regions reflecting medical cost differences, and the results normalized across regions before being combined. This type of procedure is often used in pricing health insurance.

Detailed clinical outcomes that are diagnostic specific, such as pain relief after medication administration, are not practical at this time due to data related issues (e.g. much of information needed exists in doctor's reports but is not captured electronically today) but are an area of significant potential.

Satisfaction

Satisfaction can be measured using patient satisfaction surveys, customer (employer) satisfaction surveys, and retention of the patient in the network for the first 30 days. The response rate on patient surveys may be low unless the patient is comfortable that their anonymity will be preserved.

Overall

The measurements can use closed claim data only or data on open as well as closed claims. Measurements involving dollars can use paid or incurred loss amounts. There are benefits and drawbacks to each. Closed claim data is the final result on a claim, and as such is not subject to question or manipulation. However, depending on the length of the experience period used, closed claims may include only non-serious cases. This would mean the exclusion of results on the more difficult claims. Paid loss data does not include the subjective element of case reserves over which the doctor is not in control. However, paid amounts to date may be minor and do not contain information on the best estimate of the ultimate claim amount.

Dealing with outliers also needs to be addressed. Averages can be heavily influenced by a small number of very large claims. In some cases, it may be more appropriate to use medians or to only include data that is within a certain percentile (perhaps from the 10th to 90th percentile) when computing the average.

Once measurements have been selected and the underlying data to be used is determined, weights need to be selected in order to combine the measurements for an overall score. Weights are needed within each of the four areas and across the areas. At the beginning of a provider profiling program, selection of the weights may be heavily judgmental, with the greatest weight assigned to measurements with the most direct link to ultimate costs. For example, under the return to work category, more weight would be given to the median indemnity cost than to the percentage of claims with lost time. In the overall rating, the most weight might be given to medical outcomes and return to work outcomes since these are the ones most directly tied to costs. Over time as experience becomes available, the correlation of the measurements with aggregate ultimate loss cost outcomes

should be tested. For example, did a group of doctors that scored well on temporary total disability days also have low average claim severity once the claim matured? Did doctors with a high number of office visits have higher average claim severity? The relative weights assigned to the measurements should be adjusted based on the results. Measurements that do not prove to be good predictors of loss costs should be eliminated.

5. Data Considerations Revisited

Once the statistics to be measured are determined, the actual data needs to be obtained. Provider information and claimant data can typically be linked by a common claim number. Medical bill detail and provider information can typically be linked together by a common provider number which can then be linked to claim level data (bill level data may not always contain claim numbers).

Some preliminary questions still need to be addressed when developing and using the database before provider results can be compared. First, one must decide how to uniquely identify providers or provider groups. This may be a challenge. Names and addresses can be used but if the spelling is not the same from one bill to the next, it may be hard to link all claims together. Using provider tax identification number (TIN) is an option, but there may be multiple providers paid under the same TIN. Depending on company systems, one may find it necessary to manually review provider lists to figure out what approach works best.

Second, one needs to decide how to identify the primary care physician on a claim. A single doctor or clinic needs to be designated as primarily "responsible" for the case's outcome. This is analogous to the gatekeeper in the group health system. Should this be a specific type of provider such as an occupational physician. Should this be the doctor with the greatest billed amount? Should it relate to the number of office visits? Should it be the specialist, if the case was referred?

Third, one must decide which doctors or clinics will be among those to be evaluated: those with very few claims will not have credible experience. Should there be a minimum number of claims above which the doctor's experience will be used in full? Given that random fluctuations can affect conclusions, should partial credibility be introduced? Obtaining a sufficient sample size may be a stumbling block even for large writers of workers compensation given the number of providers that exist in a state.

Fourth, one must decide how to deal with claims at different maturities across providers (or how to ensure that the data is all at a similar age). At a minimum, a check should be done to see if there is wide variation in the age of claims. If there is, one can limit the data used to common accident periods and valuation periods. Alternatively, one can look at individual claims a specific number of days after the claim has occurred (similar to data reporting for NCCI's detail claim information). Development factors might also be used though their use is not without problems. The volume of data available may play a role in deciding how best to address this issue if it exists.

6. Adjusting for Case Mix Differences

An important consideration when comparing average costs between doctors is the type of claims being handled. Case mix differences will be present due to physician specialties, the types of occupational injuries that predominate in an area, and random variation in the severity of injuries being treated. Many of the performance measurements will be distorted unless an adjustment is made for case mix. The provider's results must be normalized for case mix differences. This is necessary even if we are only profiling a subset of providers (e.g. physical therapists, occupational health clinics).

The process to normalize the data begins with defining the injury categories. This can be done using any number of methods and degree of detail. Using body part and injury type codes is one way to categorize claims. Other options include using broader injury groupings or using ICD9 codes.

The relative severity of a body part/injury type (or other chosen segmentation) then needs to be determined. One possible methodology is the Bailey Simon method^{iv}, an interactive technique that determines the proper relativities when a general category can be grouped in more than one manner. This method has been used in auto ratemaking to produce the proper relativities by territory and class of driver. In provider profiling we may use it to do the same for body part and injury type, allowing us to assign high, medium, or low severity to body part/injury type combinations. Some body part/injury type combinations are extremely severe and unusual, and as "outliers" are unsuitable to include in the measurement of provider results. These should be excluded from the calculation of the severity index.

Once injury categories have been determined, the data can be normalized for case mix differences. One may determine if an individual provider's case mix is significantly different from the overall population using the chi square test, which tests the hypothesis that the sample has the same probability density function as the overall population. In health insurance, adjustments are made for case mix (differing levels of patient risk) using one of two techniques. In the first, indirect standardization, expected results are determined from the overall population and then applied to the provider's case mix. These expected results using the provider case mix are then compared to actual results. In the second, direct standardization, provider results are applied to a standard case mix. The latter approach produces results that best compare relative performance. Adjusting the provider profiling data for case mix differences removes the penalty that would otherwise be given to a provider who sees a disproportionate number of high severity injury types. Appendix A contains a simplified example of a normalization technique.

7. Ranking Providers

The actual ranking of providers is the easy part of provider profiling, once one has gotten the required data and computed the desired statistics with adjustments for case mix as needed. Provider performance relative to their peer group may be determined by seeing where the provider falls in the distribution of results. This can be done for individual statistics with the four categories discussed above. If scores are assigned to each statistic and weights given to each, the ranking can also be done by category (quality, return to work outcomes, medical outcomes, and satisfaction), or it can be done overall. One may determine if an individual provider's results are statistically different from the overall population using the t test, which tests the hypothesis that the sample has the same mean as the overall population. The number of standard deviations from the mean measures the degree of departure in the results. Using the standard deviation also allows for consistency of results from one measurement period to the next.

Once providers have been ranked, we can use the information in a variety of ways:

- selecting and maintaining superior networks and panel lists
- rewarding better than average providers
- use in quality assurance and provider education
- use with customers to encourage channeling and to differentiate one managed care program from another
- maintaining a competitive edge

8. Summary

Workers compensation managed care programs have grown extensively in the 1990s. The use of preferred provider networks is now commonplace as a cost containment strategy. How effective are these networks and which doctors deliver the best outcomes are often asked questions. Provider profiling can be used to answer these important questions and to give companies a competitive edge. Techniques to profile providers in workers compensation are in their infancy. As highlighted in this paper, the benefits of provider profiling are plentiful, but there are many challenges to overcome before implementing a provider profiling system.

Appendix A

Normalization Technique Adjusting for Case Mix Differences

Data as Reported:

Provider	Simple Claims		Complex Claims		All Claims	
	Number	Avg \$	Number	Avg \$	Number	Avg \$
A	90	600	10	6000	100	1140
B	70	500	30	5000	100	1850
Total	160	556	40	5250	200	1495
% Total	80%		20%		100%	

If we look at the average cost per case for All Claims in the above chart, provider A appears to outperform provider B since provider A's average cost is \$1,140 versus provider B's average cost of \$1,850. However, when we look by type of claim, provider A's average costs are actually higher than provider B's average costs. The results for All Claims reflects differences in the mix of simple and complex claims between providers.

How should we adjust overall results to remove distortions due to differences in case mix? One method is to recompute the averages using the distribution of claims for all providers combined. This is done in the chart below.

Information Normalized to Reflect the Same Mix of Claims for All Providers:

Provider	Simple Claims		Complex Claims		All Claims	
	%	Avg \$	%	Avg \$	%	Avg \$
A	80%	600	20%	6000	100%	1680
B	80%	500	20%	5000	100%	1400
% Total	80%	556	20%	5250	100%	1495

After adjusting for case mix, provider A looks worse than provider B in all cases.

¹ Vincenzino, J.V., "Health care cost: Market forces and reform." Statistical Bulletin Metropolitan Insurance Company 76(1):29-35, 1995.

" Margoshes, Bart. " Disability Management and Occupational Health", OCCUPATIONAL MEDICINE: State of the Art Reviews, MANAGED CARE, Volume 13, Number 4, October-December, 1998, pp. 693-704.

³ Iezzoni, Lisa I., "Risk Adjustment for Measuring Healthcare Outcomes," Second Edition, Health Administrator Press, 1997.

⁴ Bailey, Robert A., "Insurance Rates with Minimum Bias," PCAS L, 1963, pp. 4-14.

