Institute of Actuaries of Australia

Comparing return to work outcomes between vocational rehabilitation providers after adjusting for case mix using statistical models

Prepared by Jim Gaetjens

Presented to the Institute of Actuaries of Australia
12th Accident Compensation Seminar
22-24 November 2009
Melbourne

The Institute of Actuaries of Australia
Level 7 Challis House 4 Martin Place
Sydney NSW Australia 2000
Telephone: +61 2 9233 3466 Facsimile: +61 2 9233 3446
Email: actuaries@actuaries.asn.au Website: www.actuaries.asn.au

Abstract

Keywords: vocational rehabilitation; return to work; outcome measurement; provider outcomes; statistical modelling; case mix; CAPO

WorkCover South Australia has developed a measure of RTW outcomes achieved by vocational rehabilitation providers, adjusted for case mix. The measure is called the "Characteristic Adjusted Performance Outcome" (CAPO). A positive CAPO result means that the provider's RTW outcome is above average given their case mix, and is indicative of good performance. A negative CAPO result is indicative of poor performance.

The RTW measurement is based on reductions in income maintenance. Both full and partial RTW are encapsulated into a single measure, as well as partial RTW prior to redemptions of income maintenance.

RTW outcomes are a function of (i) case mix and (ii) provider performance. The impact of case mix has been modelled using statistical methods. Approximately two-thirds of the variation between providers in RTW outcomes is attributable to identifiable variations in case mix.

CAPO removes the effects of case mix insofar as they are identifiable from claims data. Thus, the CAPO outcome is driven by a combination of provider performance and the residual effects of case mix. WorkCover's claims manager uses the CAPO result together with other information to make a judgment about provider performance, which in turn informs future referral decisions.

1. Introduction

It is a difficult task to compare objectively the performance of providers. Outcomes can be difficult to measure objectively, and even when satisfactory measures exist the performance of providers is usually confounded with variations in case mix.

This paper describes a methodology used by WorkCover SA to compare return to work (RTW) outcomes between Vocational Rehabilitation providers. It describes how "return to work" is defined and measured, and how identifiable effects of case mix are removed from the measurements.

We have found that approximately 25 of our providers are of sufficient size to apply the methods described in this paper. These providers cover over 95% of in-scope referrals for vocational rehabilitation services.

Referrals are in-scope if the claim has at least some income maintenance at the time of referral. For in-scope claims, reductions in income maintenance following the referral provide the basis of RTW outcome measurement.

WorkCover SA has analysed the effect of case mix on income maintenance reductions using generalised linear modelling. We restricted the measurement of "case mix" to claim characteristics that are measured accurately and consistently on the administrative database. We found that approximately two-thirds of the variation between providers' RTW outcomes can be explained by variations in these claim characteristics.

Reductions in income maintenance together with adjustments for claim characteristics provide the basis of comparing providers' RTW outcomes. We have called these the "Characteristic Adjusted Performance Outcome" (CAPO) measurements.

The CAPO measurements provide objective and quantified outcome comparisons between providers. WorkCover SA uses these together with other data (including the cost of services) and qualitative information about each provider (including individual claim file reviews) to make a judgment about the comparative quality of services from each provider.

The judgments about service quality inform decisions about which providers to select for future referrals. The result is a system which rewards those providers who can maximise RTW outcomes given their particular case mix.

2. Definitions and measurement of RTW

2.1 Definitions

It is helpful to define some terms used in this paper with the aid of the claim timeline in Figure 1.

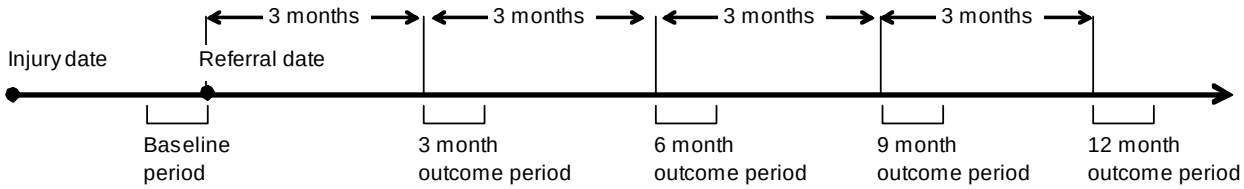


Figure 1: Timeline for a claim referred for vocational rehabilitation

The important definitions of dates and periods (refer to Figure 1) are:

- "Referral date" is the first date of service from the vocational rehabilitation provider.
- "Baseline period" is the two week period immediately prior to the referral date.
- "3 month outcome period", "6 month outcome period", etc. are the three week periods immediately following 3 months post referral, 6 months post referral, etc. respectively.

Other important definitions are:

- "Pre-injury hours" is the number of hours normally worked prior to the injury date.
- "Hours lost" is the number of hours not worked due to injury.
- "Pre-injury earnings" is the amount normally earned immediately prior to injury.
- "Income maintenance (IM)" is the amount paid by WorkCover to compensate for hours lost.
- "IM redemption" is a lump sum payment from WorkCover that redeems all future IM liability.
    - If a worker has been paid an IM redemption, then there is no further IM paid after the redemption date. However, for the RTW measurement we "impute" weekly IM at an amount equal to the weekly IM immediately prior to the redemption date.
- "Full IM entitlement" is the IM the worker would be entitled to if they were fully incapacitated, i.e. if their Hours lost per week were equal to their Pre-injury hours.
    - In SA, full IM entitlement equals Pre-injury earnings periodically increased by inflation (using the ABS Wage Cost Index) and periodically reduced by step-downs (*viz*. 10% reduction at 13 weeks post-injury and 20% reduction at 26 weeks post-injury).
- "Incapacity" is the *proportion* of pre-injury hours that the worker is not working.
    - Conceptually this equals Hours lost *divided by* Pre-injury hours.
    - In practice, for the measurements described in this report, "Incapacity" proportion is measured as IM paid *divided by* Full IM entitlement. These components are used (instead of Hours lost and Pre-injury hours) because they are measured more accurately and consistently on the database, and they give the same answer.
- "Baseline incapacity" is the incapacity during the Baseline period.
- "Outcome period incapacity" is the incapacity during a selected outcome period.

2.2 Measurement of RTW for a claim

The above leads to the definition of the four RTW measures (at 3, 6, 9 and 12 months respectively) for an individual claim:

> RTW at 3, 6, 9 and 12 months respectively
> = Baseline incapacity *minus* Outcome period incapacity at 3, 6, 9 and 12 months respectively

This definition encapsulates full RTW (or "IM discontinuance") and partial RTW into a single measure. It also recognises partial RTW that occurs prior to IM redemptions.

2.3 Measurement of RTW for a provider

RTW as defined above is calculated in aggregate for a vocational rehabilitation provider as follows:

- Select all claims with a referral date during a 12 month period (the "referral period").
- Restrict to claims with:
  - some IM paid during the baseline period
  - not reaching retirement age before the outcome period
  - not redeemed prior to the referral date
- Calculate incapacity as the sum (over all selected claims) of IM paid *divided by* the sum (over all selected claims) of Full IM entitlement, for the following four periods:
  - Baseline period. The result is "Baseline incapacity".
  - 3 month outcome period. The result is "3 month outcome incapacity"
  - 6 month outcome period. The result is "6 month outcome incapacity"
  - 9 month outcome period. The result is "9 month outcome incapacity"
  - 12 month outcome period. The result is "12 month outcome incapacity"
- Calculate the four RTW outcomes:
  - 3 month RTW = Baseline incapacity *minus* 3 month outcome incapacity
  - 6 month RTW = Baseline incapacity *minus* 6 month outcome incapacity
  - 9 month RTW = Baseline incapacity *minus* 9 month outcome incapacity
  - 12 month RTW = Baseline incapacity *minus* 12 month outcome incapacity

Thus, for each 12 month referral period, we have four RTW outcome measures for each vocational rehabilitation provider.

3. Assessing provider performance

Differences between providers in these measures are attributable to differences in (i) case mix, and (ii) the quality and effectiveness of vocational rehabilitation services (i.e. the provider's "performance").

In order to isolate the separate effect of the provider's performance, we attempt to remove the separate effect of case mix factors. In attempting this we recognise that even the best methodology cannot entirely remove them. The final result (CAPO) will still be driven in part by "residual" case mix factors as well as the provider's performance. (This is one reason why WorkCover's claims manager supplements CAPO measures with other information.) Nevertheless, the CAPO result is a better indicator of the quality of services than the unadjusted RTW outcome measures.

A positive CAPO result indicates a better than expected RTW. That is, the RTW outcomes exceed the average RTW expected for that provider's case mix. A negative CAPO result indicates a worse than expected RTW, given the particular case mix. If CAPO equals zero, the provider's RTW outcome is average, given their case mix.

3.1 The CAPO assessment

The CAPO result for each provider is calculated as follows:

- Calculate the unadjusted ("actual') RTW:
  - Actual RTW = Actual Baseline Incapacity *minus* Actual Outcome Incapacity
- Calculate the expected RTW, given the provider's case mix:
  - Expected RTW = Actual Baseline Incapacity *minus* Expected Outcome Incapacity
- CAPO result = Actual RTW *minus* Expected RTW

where "Expected Outcome Incapacity" is a function of the characteristics of the provider's claims and has been developed using the CAPO model.

3.2 The expected outcome model

The parameters for the expected outcome function, and the selection of claim characteristics to include in this function, were determined using generalised linear modelling, using the SAS procedure PROC GENMOD. The dependent variable is Outcome Period Incapacity (for each of the four periods). We used a logit link function (because the dependent variable is constrained between 0 and 1) and assumed normally distributed residuals.

We modelled the data using the experience of around 5,000 vocational rehabilitation referrals (the training data) and their associated RTW outcomes. We applied the selected models to an older set of referrals (the test data) that were not part of the training data, and tested that the variable selection and model fit were satisfactory for both the training data and the test data.

Our selection of claims mix factors is restricted to variables satisfying all of the following criteria:

- Recorded on the WorkCover SA administrative database.
- Measured accurately and consistently over time.
- Available for all in-scope claims.
- Measurable as at the referral date.

The number of variables satisfying the above is very large to the point of being unwieldy. Therefore we consulted widely with vocational rehabilitation professionals (both within and outside of WorkCover) to reduce the list to a smaller set of variables with the highest potential to impact on RTW outcomes. This led to the following "short list" of variables:

- Baseline incapacity
- Worker age
- Sex
- Occupation
- Claim duration
- Nature of injury and Body location
- RTW objective: Pre-injury employer vs Different employer
- Employer size
- Industry of employer
- Geography (metropolitan vs country)
- Expenditure to-date (by expenditure type)
- Expenditure last 6 months (by expenditure type)

Various recodes were also applied to the above. For example, Nature of injury and Body location codes were classified into around 30 groups following the advice of a medical consultant. Other categorical variables were recoded into smaller groups when necessary to avoid sparse data. Continuous numerical variables were transformed if partial residual plots suggested a non-linear relationship with the dependant variable.

Claims characteristics were selected from the short list if they met all of the following criteria:

- Statistically significant separate effect (based on SAS Type 3 p-values)
- Practical significance (based on variability and the size of parameter estimates)
- Significant separate effect indicated by visual inspection of partial residual plots
- Improved fit indicated by visual inspection of residual plots

3.3 Final model structure

The final claim characteristics selected for the models are:

- Baseline incapacity
- Worker age
- Claim duration (log transformation)
- Selected nature of injury and body locations: Hernia; Lumbar-dorsal; Hand-wrist; Ankle-foot; Stress; Lower leg; Knee sprain; Wrist fracture; and Multiple location
- IM paid during 6 months before referral, adjusted for pre-injury earnings
- Medical costs during 6 months before referral, adjusted for claim duration

There are separate models for different combinations of the following:

- four outcome periods (3, 6, 9 and 12 months after referral)
- two RTW objectives (pre-injury employer and different employer)
- two lengths of rehabilitation services (continuous services less than 12 months, and continuous services greater than 12 months)

4. Example: Pre-injury employer, 6 month outcome

This section shows the results for the model: 6 months outcome, pre-injury employer, and services less than 12 months. The purpose is to illustrate what the models "look like", the kind of results we have found and how they are applied.

The selected model to predict outcome incapacity for an individual claim is:

$$\text{Expected outcome incapacity} = \frac{e^{\eta}}{1+e^{\eta}}$$

where

$$\begin{aligned}\eta = &-4.01 + 1.12 \times \exp(\textit{BaselineIncapacity}) + 0.264 \times \ln(\textit{ClaimDurationWks}) \\ &+ 0.0109 \times \textit{Age} - 2.64 \times \textit{HerniaFlag} + 0.293 \times \textit{LumbarDorsalFlag} \\ &- 0.518 \times \textit{HandWristFlag} - 0.648 \times \textit{AnkleFootFlag} - 0.333 \times \textit{LowerLegFlag} \\ &- 1.17 \times \textit{WristFractureFlag} + 0.0579 \times \textit{IM}\,6\textit{mths}\end{aligned}$$

To illustrate the application of this formula, consider an individual claim with the following characteristics:

- Baseline incapacity = 80%
- Claim duration at referral = 30 weeks
- Age at referral = 45 years
- Body location / nature of injury = Wrist fracture
- IM paid in 6 mths prior to referral = 12 times pre injury weekly earnings

Such a claim would have expected outcome incapacity of 35%, corresponding to an expected RTW of 45%, calculated as follows:

η = -0.6042

$$\text{Expected outcome incapacity} = \frac{e^{-0.6042}}{1+e^{-0.6042}} = 0.35$$

Expected RTW = Baseline incapacity *minus* Expected outcome incapacity
= 0.8 – 0.35
= 45%

The variation for individual claims is very high, and therefore the prediction of RTW (45% in the example) is unreliable for any specific claim. However, when the expected RTW is calculated in aggregate for a reasonably large provider (we use a cut-off of at least 30 claims), the reliability is reasonable. This is discussed in more detail in section 5.2 "Overall model fit".

5. Results

5.1 Model fit – illustrative predictors

Figure 2 graphically illustrates the impact of selected claim characteristics in aggregate, and also shows the good model fit by these variables for the training dataset. (Aggregate model fit for the test dataset is illustrated by Figure 3 in the next section.)

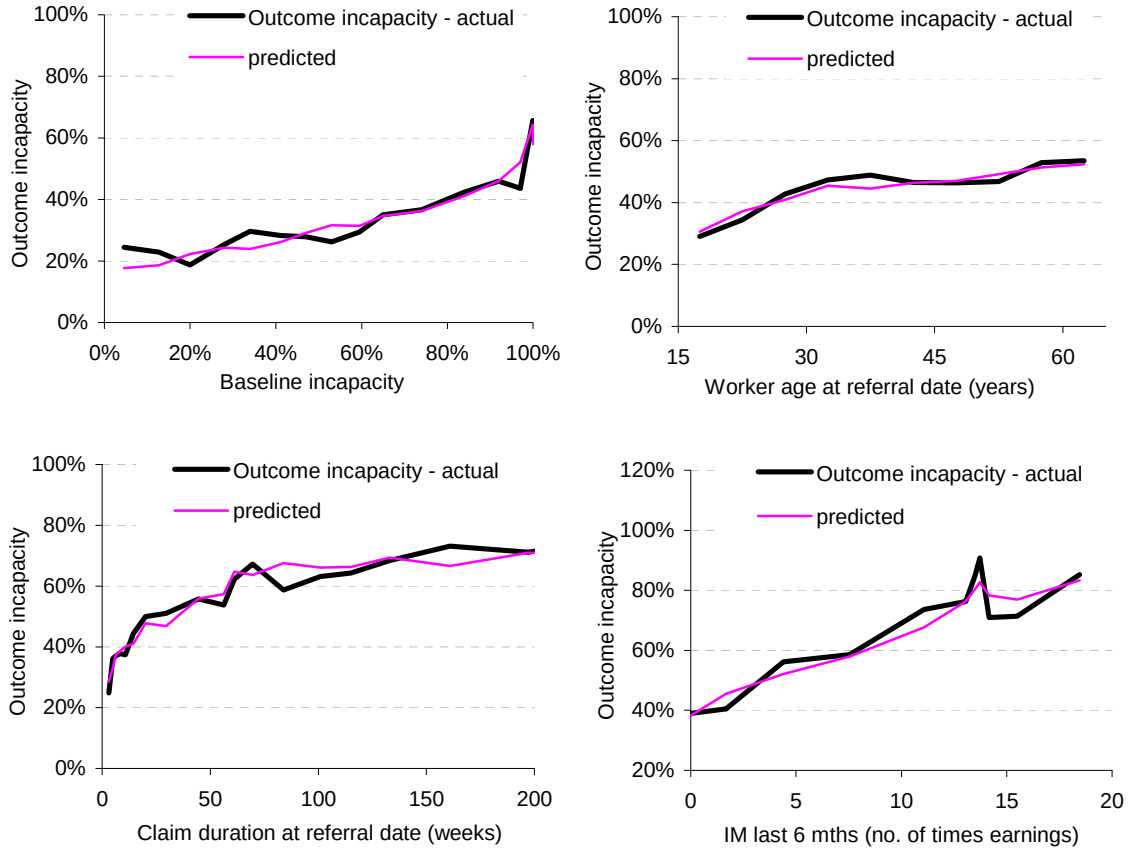
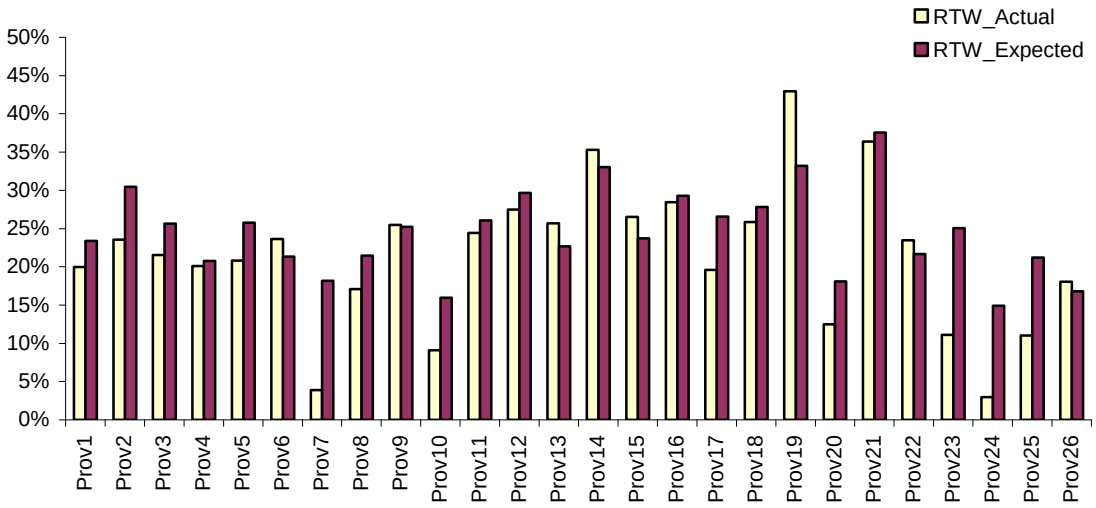


Figure 2: 6 month outcome incapacity, pre-injury employer objective, and service duration < 12 months. Relationship between selected predictors and outcome incapacity (training data). These plots are based on the average experience of claims summarised by the four respective predictors. The predicted values are not smooth lines because the values of several other predictors also vary as each of the four respective predictors vary.

5.2 Overall model fit

As a generalisation approximately two-thirds of the variation between providers in their RTW outcomes is attributable to the selected claims characteristics (based on test data). The remaining one-third is attributable to (i) variations in provider performance, and (ii) residual claims-mix factors.

This conclusion was reached by applying the model to a referral period prior to the modelling period. This should be a reasonable test of how the model will perform in the future, because none of the referrals or their outcomes in the test period was included in the modelling. Figure 3 illustrates the overall fit for one of the 6 month models. Other models are similar.



Variance removed by MODEL	0.56%
RESIDUAL variance	0.32%
TOTAL variance	0.88%
Proportion of TOTAL variance explained by MODEL	64%

Figure 3: RTW for 6 month Outcome Period, Actual vs Expected, for referrals 1/10/2006 to 30/9/2007 (test data). Training data were referrals 1/10/2007 to 30/9/2008.

The "total variance" is the variance of actual RTW. The "residual variance" is the variance of CAPO (actual *minus* expected RTW). The "variance removed by model" is the difference between these two.

5.3 Assessments of individual provider's performance

Figure 3 clearly shows the advantages of adjusting for case mix. The expected RTW follows the patterns of actual RTW reasonably well. We conclude that the providers with the best or worst unadjusted outcomes are not necessarily the best or worst performers respectively.

For example, consider Provider number 21. If we did not adjust for case mix, we would be very impressed by this provider because it has the second best RTW outcome, exceeded only by Provider number 19. However, our mathematical model suggests that Provider 21's result is driven by their case mix, because their actual RTW almost equals their expected RTW. (The difference is not statistically significant.) That is, their outcome is merely average.

When the claims manager applies this model, they will discover that none of the best performers are significantly better than expected. (The claims manager is provided with p-values.) Even Provider number 19, who's actual RTW is nearly 10 percentage points better than expected, is not significantly better than expected at the 5% level of significance (although they are at the 10% level). On the other hand, there are several poor performers with results significantly worse than expected, viz., Provider numbers 7, 17, 23, 24 and 25, for whom further investigation of their performance is indicated.

5.4 Model reliability

We supplement the CAPO result with the standard error for each provider and the corresponding p-value. The central limit theorem allows us to assume normality and to calculate reasonable p-values. Our claims manager uses this information to help determine whether a result might be merely random.

6. Tests of model assumptions and performance

We have tested the assumptions and performance of the models as follows:

- As discussed above, we have tested the performance of the model on separate unit records (test data) which were not included in the modelling dataset (training data). We found that the model performed well.
- Residual plots were produced and visually examined at all stages in the modelling. These provided a good guide that all potentially important predictors were included, that appropriate transformations were made, and that selected explanatory variables had a significant impact.
- We tested residuals for normality, and found they were sometimes at least "mound shaped" but other times departed from normality and were left skewed. Transformations did not fix this, and we could not identify an alternative distribution assumption that worked better overall. While this has the potential to affect parameter estimates and p-values, we have concluded that the problem is immaterial. First, we relied on a range of methods for variable selection and model fit (e.g. visual examination of residual plots), and usually only included variables with extremely low p-values. Finally, we checked the veracity of parameter estimates and the overall model by examining actual *vs* fitted data using the test datasets (see first dot point above).
- Excessively high correlations between predictors can make parameter estimates unstable. We applied a rule-of-thumb to investigate correlations higher than 0.7 (which means that more than half the variation in one variable is explained by the other), but found none of these. In general the correlations were very low, between 0.00 to 0.34. There was only one instance where the correlation between two variables in the same model was higher than 0.34, at r=0.63.
- We tested the link function by plotting z = g(mu) + (observed - mu) x g'(mu) (where g(.) is the link function) against the linear predictor. Strong curvature in these plots would suggest a possible problem with the link function (McCullagh & Nelder (1989) p. 401). There is no evidence of curvature for any of the models, and we conclude that the logit-link is satisfactory.
- The normal distribution when expressed in exponential family form requires a common dispersion parameter equal to the variance. To test this, we plotted (by model) the variance against (i) Predicted values, and (ii) Base incapacity. Many of these plots looked reasonable with no obvious trend, but for some there were slight trends.
- Outliers were identified, including records with extreme values of the linear predictor. Potentially influential records were excluded and the parameter estimates recalculated. In most cases parameter estimates were robust against these records. There were some exceptions, but they were not material. In particular, the ranking of providers were almost identical whether or not these records were included. Therefore, we decided to include all records for the modelling.

In summary, the link function is fine, and there are no excessive correlations between predictors. While there are problems with the distributional assumptions of normality (and possibly common variance), these do not materially affect the final results. We have minimised the effect of distributional problems by the modelling approach (in relation to the selection of variables, transformations and assessment of fit). We are confident in the selection of explanatory variables and their parameter estimates. Actual *vs* expected plots look good, and when applied to test data the selected models do a good job of prediction at the provider level. Finally, we decided to include all outliers in the modelling confident that this will not distort the final results.

7. Summary and Conclusions

WorkCover South Australia has developed a measurement system to compare return to work outcomes between vocational rehabilitation providers, after making adjustments for different case mixes. The result is the "Characteristic Adjusted Performance Outcome" (CAPO) measure.

Return to work is measured as the reduction in incapacity between the time of referral ("baseline incapacity") and respective outcome periods of 3, 6, 9 and 12 months after referral ("outcome incapacity"). The term "incapacity" refers to the ratio of income maintenance (IM) paid to the IM that would be paid if the claimant was fully incapacitated.

RTW measured in this way encapsulates full and partial RTW into a single measure. For redeemed claims adjustments are made so that partial RTW prior to the IM redemption is also included.

The effect of claim characteristics on the outcome incapacity has been modelled using generalised linear modelling. This allows the RTW outcome to be predicted for a given case mix. So we have:

CAPO = Actual RTW *minus* Expected RTW

A positive CAPO result is better than expected given the case mix. A negative result is worse than expected given the case mix.

Approximately two-thirds of the variation between providers in RTW outcomes is attributable to their case mix, to the extent that this is measurable. The CAPO measure removes this variation, and is therefore a superior indicator of relative provider performance than unadjusted RTW measures.

The CAPO scores for individual providers are used by WorkCover's claims manager together with other information such as the cost of services and individual claim file reviews, to reach conclusions about their relative performance. This in turn affects the case manager's decisions about which providers to select for future referrals. The result is a system which rewards those providers who can maximise RTW outcomes given their particular case mix.

8. References

McCullagh P & Nelder J A, 1989, *Generalized Linear Models*, 2nd Edition, Chapman & Hall, repr. 1999 CRC Press, Florida U.S.A.