**Article**

# Development of Cut-Scores for Feigning Spectrum Behavior on the Orebro Musculoskeletal Pain Screening Questionnaire and the Perceived Stress Scale: A Simulation Study

John Edward McMahon [*] , Ashley Craig , Ian Douglas Cameron

*Article*

# Development of Cut-Scores for Feigning Spectrum Behavior on the Orebro Musculoskeletal Pain Screening Questionnaire and the Perceived Stress Scale: A simulation study

**John E McMahon \*, Ashley Craig and Ian Cameron**

John Walsh Centre for Rehabilitation Research, Kolling Institute for Medical Research, School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, St. Leonards, NSW AUSTRALIA

\* Correspondence: drjohnmcmahon@icloud.com

**Abstract**

**Background/Objectives:** Feigning spectrum behavior (FSB) is the exaggeration, fabrication, or false imputation of symptoms. It occurs in compensable injury with great cost to society by way of loss of productivity and excessive costs. The aim of this study is to identify feigning by developing cut-scores on the long and short forms (SF) of the Orebro Musculoskeletal Pain Screening Questionnaire (OMPSQ and OMPSQ-SF) and the perceived Stress Scale (PSS and PSS-4). **Methods**: As part of pre-screening for a support program, 40-injured workers who had been certified unfit for work for more than 2-weeks were screened once with the OMPSQ and PSS by telephone by a mental health professional by telephone. A control sample comprised of 40-non-injured community members were screened by a mental health professional on four occasions under different aliases, twice responding genuinely and twice simulating an injury. **Results**: Differences between the workplace injured people and the community sample were compared using ANCOVA with age and gender as covariates, and then Receiver Operator Characteristics (ROC) were calculated. The OMPSQ and OMPSQ-SF discriminated ($\varrho < 0.001$) between all conditions. All measures discriminated between the simulation condition and workplace injured people ($\varrho < 0.001$). Intraclass correlation demonstrated the PSS, PSS-4, OMPSQ and OMPSQ-SF were reliable ($\varrho < 0.001$). Area Under the Curve (AUC) was 0.750 for OMPSQ and 0.835 for OMPSQ-SF for work-injured versus simulators. **Conclusions**: The measures discriminated between injured and non-injured people, and non-injured people instructed to simulate injury. Non-injured simulators produced similar scores when they had multiple exposures to the test materials, showing the uniformity of feigning spectrum behaviour on these measures. The OMPSQ-SF has adequate discriminant validity and sensitivity to feigning spectrum behaviour, making it optimal for telephone screening in clinical practice.

**Keywords:** orebro musculoskeletal pain screening questionnaire; perceived stress scale; simulation; malingering; feigning

## 1. Introduction

Central to the management of compensable injury is addressing psychosocial barriers to recovery [1–3]. Chronic pain and associated syndromes, including opioid misuse, represent a substantial cost to society and these conditions are predicted to increase in the future [4]. The successful management of chronic pain, if possible, would result in significant economic savings, reduction of personal suffering, and increase in productivity for the individual and society. Similarly, feigning spectrum behavior (FSB) for the pursuit of secondary gain, such as work avoidance and access to pharmaceuticals, represents a significant cost to society and is detectable in 29% of personal injury at formal psychological assessment [5]. As a medium of service delivery telehealth has significant benefits in terms of convenience, cost savings, and capacity to reach clients in remote locations bringing expertise otherwise unavailable due to distance. With the onset of Coronavirus,

there was increased interest in telemedicine and adoption of these media for service delivery, rather than face to face consultation [6], even among primary healthcare providers and there is increasing consumer acceptance of services delivered via telehealth platforms. Yet little is known about how feigning spectrum behaviour may manifest via telehealth. Inherent to compensation systems is the notion of assigning a dollar value to injury, illness, and distress. In such a system the report of financial duress could be seen as a way of assessing the motivation or intent behind feigning spectrum behaviour. Binder and Rohling conducted a meta-analysis showing a moderate effect size for financial incentives and reported abnormality and disability despite less severe injury in closed head injury [7]. Financial stress also plays a complex role in health, and structural equation modelling revealed that people reporting major financial stressors reported more interpersonal stress, psychological distress and lower levels of psychological well-being, and their psychological well-being rating was associated with elevated interleukin 6, an inflammatory cytokine marker of stress [8]. In compensable injury there appears to be an inextricable relationship between report of symptoms and secondary gains that may be exploited by some individuals.

The Orebro Musculoskeletal Pain Screening Questionnaire (OMPSQ) [9] is a widely used measure of the impact of injuries and this scale indexes psychosocial factors that complicate recovery. It has moderate predictive power to identify persisting pain and disability in patients with back pain [2], sciatic leg pain [3] and acute and non-acute pain [10]. The Perceived Stress Scale (PSS) [11] has been correlated with a variety of factors including health and smoking status, and help seeking behaviour [12], and has been used to investigate work stress and associated health behaviours [13]. Both the OMPSQ [14–18] and PSS [19–21] have been validated in several languages which is advantageous to the generalization of use in a multicultural society like Australia, although some qualitative research with culturally diverse people indicates that use of standardized questionnaires may have some unintended consequences such as negative impact on client-practitioner relationship [22]. The multilingual validation of these measures demonstrates the robustness of the constructs measured by these scales and the impact of psychosocial factors to recovery and health regardless of culture and language. The OMPSQ and PSS have short forms, the ten item OMPSQ-SF [23] and the four item PSS-4 [24,25] respectively, which have similar predictive validity in health. The PSS has been validated for telephone use and the OMPSQ-SF has been used telephonically, however, there has been no formal validation of the OMPSQ-SF for telephone administration. Unlike general health settings, the treatment of compensable injuries often involves complex interactions between the patient, treatment providers, agents of insurers, insurers, and legal service providers. Such complex systemic interactions can produce behaviours including malingering, factitious presentations, and false imputation of causes of conditions [26]. A comprehensive review of the literature across various search engines including MEDLINE via OVID, Scopus, PsychInfo via OVID, Web of Science, and Google produced no results for either the OMPSQ or PSS being validated with a simulation design to see if a feigned presentation can be distinguished from a genuine presentation.

The most successful method of detecting feigned mental disorder is the "Rare Symptoms" and "Quasi-rare symptoms" detection strategies [27]. These strategies detect feigning spectrum behaviour by indexing the overreporting of infrequent symptoms or the reporting of symptoms to an exaggerated extent. That is, on measures of clinical symptoms people engaging in feigning tend to endorse more pathology and rarer pathology than genuine respondents, producing inflated scores exceeding those generated by injured people reporting their pathology accurately. Usually, this strategy of detection generates a large to very large effect size [28]. Simulation studies involve the determination of cut-points for feigning based on the probability of genuine responding by comparing people with the condition to people without the condition who were asked to fake the condition. Criterion referenced or "known group" studies involve classifying people with the condition and those without the condition by some means, such as another measure or a known diagnosis, and then administering the subject measure. Research on the Minnesota Multiphasic Personality Inventory-Restructured Format (MMPI-RF), show that both simulation studies and criterion referenced studies generate large effect sizes [28,29]. Similar effect sizes were also found

regardless of what disorder or symptom was to be faked. However, the effect size reduced to the moderate range if the simulator was educated about the disorder or the measure, so called "sophisticated" simulators [29,30].

An alternative strategy to detecting feigning spectrum behaviour is forced choice testing, and this has been applied to cognitive and memory impairments, and malingered pain [31], using the Test of Memory Malingering (TOMM). However, the medium of telephone limits the stimuli to auditory and the TOMM is a visual task. A verbal forced choice test, akin to the 21-Item Test [32], a verbal list learning task, would offer a further detection strategy. This method identifying feigning spectrum behaviour were explored by the present study in a "simulation" design and a 10-item Brief Forced Choice Test (BFCT) of verbal learning was created to test this method against the cut-score approach.

The present study aims to develop cut scores to identify FSB on the OMPQ and PSS and their short-forms and to see if a novel forced choice strategy or financial rating can identify feigning. The design is a simulation study comparing the pattern of responding of injured workers (IW) with Community Sample (CS) of non-injured people instructed to simulate a workplace injury (CSS) and respond genuinely (CSG). A further aim was to see if multiple exposures to test materials influenced the strategy to engage in FSB, and if practice improved a non-injured person's ability to simulate an injury. The study also investigated the degree of reported financial duress to see if simulators equate financial distress with the intent to feign an injury, and if there is any discriminant value in enquiring about financial status. It was hypothesized that the CSG will score significantly lower than IW on all measures, and that CSS will score significantly higher on all symptom measures and significantly lower on the BFCT and financial rating than IW.

## 2. Materials and Methods

*Study Design and Participants*

Given the expected large effect size and comparisons to be made a sample size of 40 participants in each group would have sufficient statistical power [33] with a power estimate of 0.98 [34]. The IW were the first 40 referrals who consented to pre-screening for an independent support service for people who had suffered compensable workplace injuries. The pre-screening included assessing suitability for the program, current treatment, social support, relationship with employer and insurer and then the administration of the questionnaires. The CS was a sample of convenience of 40 non-injured people including employees from other business units and willing associates. Both groups were onboarded to the service by the first author and their permission to participate was acquired orally for all telephone calls to be recorded "for quality, training and research purposes". The administration of the questionnaires was explained as part of pre-screening and a convenient time was scheduled for administration of the questionnaires by an intern psychologist. Each CS participant was scheduled under an alias into the workflow of the intern psychologist and instructed on which calls to respond genuinely and when to simulate. Each CS participant responded to the assessment protocol four times, responding twice in the simulated condition (CSS) and twice genuinely (CSG). The order of simulation and genuine responding was counterbalanced to see if there were practice effects. In the simulation condition the community sample participants were instructed to "fake an injury that occurred in the course of your work".

*Measures*

The OMPSQ is a 25-item measure that is scored by summing the ratings of 21-items that are a Likert type scales which have various anchoring statements at the endpoints yielding a range of scores 4 to 210 [9] with scores over 105 predicting poor outcomes [35]. The OMPSQ-SF is a subset of 10-items of the parent scale, with scores greater than 50 predicting unfavourable outcomes [23]. These measures were administered as a set of verbal rankings using the anchoring statements and scaled out of 10 to yield totals as in the pen and paper measure. The PSS is a 10-item scale in which

statements are rated from "never" to "very often" and it was interpreted dimensionally using the printed norms [11]. The PSS-4 short form is a 4-item scale using items 4, 5, 7 and 8 of the parent scale and it was interpreted dimensionally using the printed norms. The short forms were calculated from the subset of items from the long forms of these scales. Consistent format with the OMPSQ items, participants were asked to rate their financial wellbeing from 0 to 10 anchored from 0 = "very poor" to 10 = "good". The BFCT was a 10-item list learning task with a free recall trial, a distraction trial of counting backward from 20 to 0, and then the forced choice trial [Appendix 1 lists the target & foil words].

### Data Collection

An intern psychologist with honours level qualifications, was trained in the telephonic administration of the measures, with the calls made within a period from November to December 2017 when the program was initiated. The dataset was then compiled and de-identified and provided to the researchers by the manager of the support program.

### Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics Version 26 (64-bit Edition). Shapiro-Wilk Tests and Levene's Tests were used to assess for normality. Differences between subsets of the community sample were investigated using Mann-Whitney U-test or T-tests depending on normality. Differences between the CS and IW were compared for age using independent groups T-test and for gender using a Bayesian Independent Groups analysis. Given the predicted directional relationship and large effect size, the IW, CSG and CSS were compared using one-way Analysis of Variance, or if marked difference between demographic variables was identified, then Analysis of Covariance (ANCOVA) would be conducted [36]. Given that the CSS and CSG was assessed on two occasions, the test re-test reliability was assessed using a two-way, single measures absolute intraclass correlation for each measure. With a 95% confidence interval and large expected intraclass correlation coefficient for the CSG and CSS a sample size of 40 (n>37) retained power (Power > 0.80) [37]. Receiver operator characteristics (ROC) and area under the curve (AUC) were calculated for the OMPSQ, OMPSQ-SF, PSS, and RSS-4. The Sensitivity (Sn), Specificity (Sp), Positive Predictive Power (PPP) and Negative Predictive Power (NPP) were calculated for viable cut scores.

## 3. Results

The study was conducted in Sydney Australia with the data deidentified and conducted to the researchers in December 2019 after ethics approval was received on 29th June 2019.

### 3.1. Demographic Differences

IW suffered non-trivial injuries and consisted of injury 15-upper limb, 13-lower limb, 3-shoulder, 9-back, two lung injuries, and one "internal organs other"; with three of the IW coded with two injuries. Table 1 outlines the demographic data describing the IW and the CS. An independent-measures t-test revealed that the IW were significantly older than the CS (t = 2.253, Sig (2-tailed) = 0.027). Bayesian Independent Group analysis of the gender ratios in the groups revealed the ratio of males to females was statistically significantly different (t = -6.637, Sig. (2-tailed) = 0.000) with more females than males in CS. Table 1 shows the differences between the IW and CS.

Shapiro-Wilks Test for normality in scores comparing males to females showed non-normality in CSG on five of the eight conditions comparing each gender on the PSS, PSS-SF, OMPSQ and OMPSQ-SF. Subsequent Mann-Whitney U-test showed no significant differences between genders and responses on these measures in the CSG. Shapiro-Wilks Test for normality comparing males to females in the CSS, showed normality for each condition and measure, and t-tests showed no significant differences on any of the measures. Shapiro-Wilks test of normality in scores comparing employed to not employed for the CSG showed non-normality in four of the eight conditions on PSS,

PSS-4, OMPSQ and OMPSQ-SF. Subsequent Mann-Whitney U-Test showed no significant differences between employed and not employed status in CSG and their test scores. Shapiro-Wilks test of normality in scores comparing employed to not employed in CSS showed non-normality in PSS-4 for not employed, but the other seven conditions had normality. The three t-tests and one Mann-Whitney U-Test showed no significant differences between employed and not employed within CSS and their test scores.

**Table 1.** Demographic Differences between Injured Workers (IW) and Community Sample (CS).

| Variable | Injured Workers | Community Sample |
|---|---|---|
| N | 40 | 40 |
| Age (SD, Median, IQR, Max & Min) | 41.575 (12.226, 39, 18 Min = 18 Max = 66 | 34.95 (14.016, 28, 24, Min = 19 Max = 72)* |
| Gender: Male/Female | 33 (82.5%) / 7 (17.5%) | 9 (22.5%) / 31 (77.75%)* |
| Professional with Degree | 5 (12.5%) | 14 (25%) |
| Qualified Trades (Mechanics, chefs, etc) | 14 (35%) | 11 (27.5%) |
| General Labour/Farm Worker/Machine Operator/Driver | 14 (35%) | 0 (0%) |
| Mining Industry | 5(12.5%) | 0 (0%) |
| Hospitality | 3(5%) | 2 (5%) |
| Students | | 8 (20%) |
| Mean days from injury to | 38.7 (13.2) | N/A |

*Significant differences (Sig. >0.05).

### 3.2. Analysis of Covariance Between Conditions

Table 2 contains the means for each condition and the results of the Analysis of Covariance with covariates of Gender and Age. The OMPSQ and OMPSQ-SF discriminated between the responses of IW, CSG and CSS. However, the PSS and PSS-4 and Finance Question only discriminated between the CSS and the groups responding genuinely. Gender and Age did not significantly contribute to score on OMPSQ (sig = 0.320 / 0.694), OMPSQ-SF (sig = 0.637 / 0.637), PSS (sig = 0.650 / 0.988), PSS-4 (sig = 0.547 / 0.707), or Forced Choice Total (sig. = 0.115 / 0.444), but gender did contribute to variance in the Finance Rating (sig. = 0.027).

**Table 2.** Means, Standard Deviations, Ranges of Scores and Analysis of Covariances For Age and Gender.

| Test Score | GSG Trial 1 Mean (SD) [range] | CSG Trial 2 Mean (SD) [range] | IW Mean (SD) [range] | CSS Trial 1 Mean (SD) [range] | CSS Trial 2 Mean (SD) [range] | Title 1 | Title 2 | Title 3 |
|---|---|---|---|---|---|---|---|---|
| OMPSQ | 38.67*C, D,E (21.87) [7-113] | 40.50*C, D,E (22.89) [4-112] | 84.15*A, B,C,D (28.83) [14-152] | 118.08* A,B,C (36.01) [46 – 192] | 112.25* ABC (35.15) [32-186] | entry 1 | data | data |
| OMPSQ-Short Form | 20.96*C DE (12.97) [1-63] | 21.88*C DE (13.05) [2-58] | 41.20*A BDE (13.87) [10-87] | 59.73*A BC (18.93) [18-91] | 56.18*A BC (18.32) [28-85] | entry 2 | data | data [1] |

| | | | | | |
|---|---|---|---|---|---|
| PSS Total | 13.11*[DE] (6.251) [1-23] | 12.51[DE] (6.76) [2-27] | 14.94*[DE] (7.23) [2-30] | 22.10*[ABC] (8.01) [0-39] | 22.16*[ABC] (6.28) [8-36] |
| PSS-4 Total | 4.55*[DE] (2.60) [0-9] | 4.35*[DE] (2.78) [0-11] | 5.45*[DE] (2.82) [0-12] | 8.53*[ABC] (3.41) [0-15] | 8.38*[ABC] (2.71) [3-15] |
| Finance Rating | 8.43*[DE] (2.03) [2-10] | 8.53*[DE] (1.96) [2 – 10] | 6.85*[DE] (3.45) [0-10] | 5.48*[ABC] (3.57) [0-10] | 4.78*[ABC] (3.29) [0-10] |
| Forced Choice Test | 9.35 (0.80) [7-10] | 9.05 (1.20) [5-10] | 9.30 (1.07) [6-10] | 9.00 (1.78) [1-10] | 8.65*[AC] (1.59) [3-10] |

* Significance at 0.05 level, A = Community Sample Genuine (CSG) Trial 1, B = CSG Trial 2, C = Injured Workers (IW), D = Community Sample Simulating (CSS) Trial 1, E = CSS Trial 2.

### 3.3. Short Term Test-Retest Reliability

The test-retest reliability for the CSG and CSS conditions were acceptable, revealing uniform behaviour across the different conditions with telephone administration and the results of this analysis are shown in Table 3

**Table 3.** Short-Term Test-Retest Reliabilities for the OMPSQ, OMPSQ-SF, PSS and PSS-4 of the Community Sample in two conditions administered twice.

| Measure | ICC (95% CI) | *P* value | Minimal Detectable Change |
|---|---|---|---|
| OMPSQ | 0.954 (0.943 – 0.964) | <0.001 | 21.719 |
| OMPSQ-SF | 0.905 (0.882 – 0.926) | <0.001 | 10.56 |
| PSS | 0.926 (0.908 – 0.942 | <0.001 | 13.509 |
| PSS-4 | 0.836 (0.791 – 0.874) | <0.001 | 6.114 |

### 3.4. Receiver Operator Characteristics

"Elbow" points are where the plot line of the ROC curve approaches 90-degrees which indicate that the test score clearly delineates between the two samples. Visual inspection of the ROC curves in Figure 1 shows that the OMPSQ has acceptable "elbow" points for identification of cut scores to differentiate between CSG and the IW and that the OMPSQ SF has some opportunities but tends towards a more dimensional interpretation. The other scales show little opportunity for cut scores. Figure 2 shows the OMPSQ and OMPSQ-SF have shapes that indicate likely cut-points for distinguishing CSS from IW.
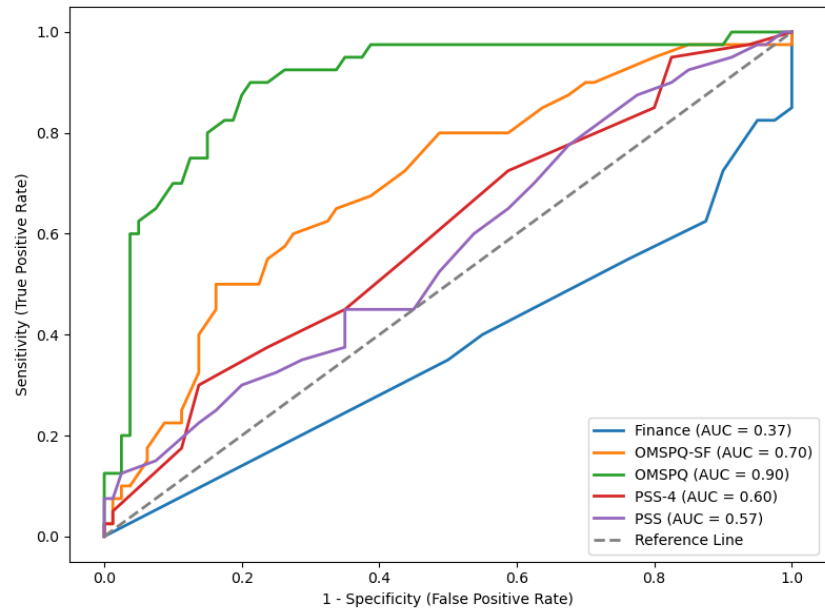
**Figure 1.** Receiver Operator Characteristics Curve or Injured Workers (IW) versus Community Sample Genuine (CSG).
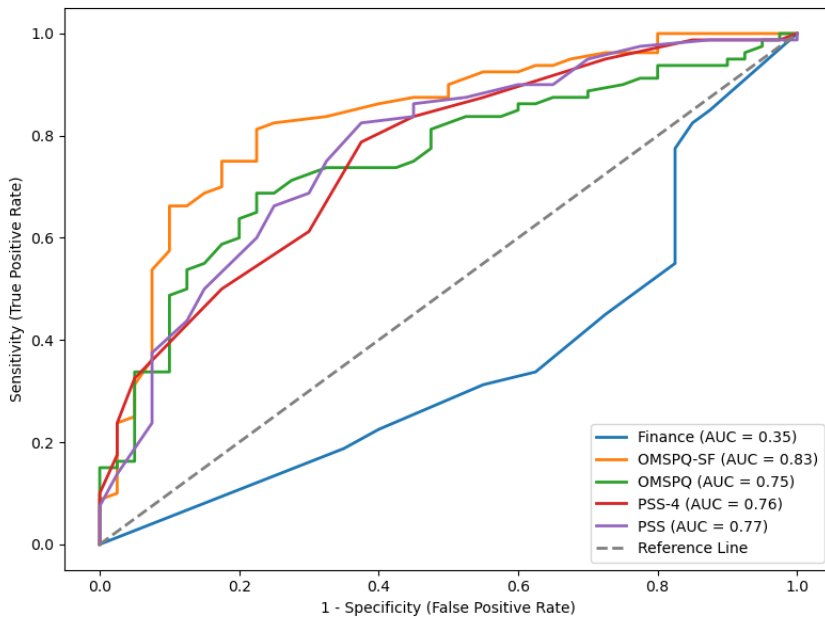


**Figure 2.** Receiver Operator Characteristics (ROC) Curve or Injured Workers (IW) versus Community Sample Simulating (CSS).

Table 4 contains the ROC of the CSG and CSS compared with the IW. Comparing the lower bound of the confidence interval of the estimated AUC as an assessment of the overall model quality, showed that the OMPSQ (0.83) and OMPSQ-SF (0.60) discriminated between the IW and CSG, but the other measures were not better than chance at classifying membership to this group. The PSS (0.68), PSS-4 (0.67) OMPSQ (0.66) and OMPSQ-SF (0.76) classified CSS from IW, with the OMPSQ-SF being the best for this determination.

**Table 4.** Receiver Operator Characteristics (ROC) comparing Community Sample Genuine (CSG) and Community Sample Simulation (**CSS) with Injured Workers (IW).**

| Measure | Comparison | Area Under Curve | Std. Error | Asymptotic Significance | 95% Confidence Interval Lower - Upper |
|---|---|---|---|---|---|
| OMPSQ | CSG v IW | 0.897 | 0.032 | < 0.000 | 0.834 – 0.959 |
| | CSS v IW | 0.751 | 0.461 | < 0.001 | 0.661 - 0.849 |
| OMPSQ-SF | CSG v IW | 0.700 | 0.51 | < 0.001 | 0.834 – 0.959 |
| | CSS v IW | 0.835 | 0.040 | < 0.001 | 0.757 - 0.913 |
| PSS | CSG v IW | 0.567 | 0.056 | 0.235 | 0.457 – 0.676 |
| | CSS v IW | 0.771 | 0.046 | < 0.001 | 0.681 – 0.862 |
| PSS-4 | CSG v IW | 0.587 | 0.055 | 0.078 | 0.489 – 0.704 |
| | CSS v IW | 0.760 | 0.046 | < 0.001 | 0.670 – 0.849 |
| Finance | CSG v IW | 0.370 | 0.055 | 0.180 | 0.263 – 0.478 |
| | CSS v IW | 0.354 | 0.540 | 0.007 | 0.247 – 0.460 |

A cut score for distinguishing CSG and IW on the OMPSQ of 15 had a Sn of 97.5% and SP of 90%, with a PPP of 2.5% and NPP of 100%. A cut score of 51 had 90% Sn, 23.8% Sp, PPP of 92.5%, and NPP of 78.75% which shows the broad range of scores in the IW sample. For discriminating between CSG and IW the OMPSQ-SF at a cut off at 20 had a Sn of 95%, 80% Sp, PPP of 97.5% and NPP of 21.25%.

Table 5 and Table 6 contain the prospective cut-scores for discriminating the IW sample from the CSS with Sn, Sp, NPP and PPP at each of the upper limit of the bands identified by Linton, Nicholas, and MacDonald [23] for the OMPSQ and OMPSQ-SF, respectively. Using a specificity of greater than 95% to minimize false positives, that the OMPSQ using a cut score of 140 identified 28.8% of the CSS and 5% of IW sample as engaging in FSB. The OMPSQ-SF using a cut score of 70, identifying 25% of CSS with a 5% identification of IW as engaging in FSB. Given base rates of FSB [38], it is probable that these are true positive identifications of feigning in this IW sample.

**Table 5.** Cut-scores, Sensitivity (Sn), Specificity (Sp), Negative Predictive Power (NPP) and Positive Predictive Power (PPP) form OMPSQ.

| OMPSQ Cut Score [Range] | Sn x100 | 1-Sp x100 | NPP % | PPP % |
|---|---|---|---|---|
| 49 [0-50] | 95.0 | 92.5 | 33 | 92.68 |
| 59 [51-60] | 93.8 | 82.5 | 36.84 | 86.21 |
| 69 [61-70] | 87.5 | 70.0 | 36.11 | 75.26 |
| 79 [71 – 80] | 83.8 | 50.5 | 36.73 | 68.36 |
| 90 [81 – 90] | 75.0 | 45.0 | 33.87 | 59.41 |
| 99 [91-100] | 68.8 | 25.0 | 35.36 | 48.18 |
| 110 [101-110] | 53.8 | 12.5 | 32.38 | 37.72 |
| 120 [111-120] | 47.5 | 10.0 | 31.85 | 33.04 |
| 126 [121-130] | 37.5 | 10.0 | 29.75 | 26.09 |
| 140 [131 – 140] | 28.8 | 5.0 | 28.57 | 18.80 |
| 150 [141 – 150] | 16.3 | 2.5 | 26.89 | 10.17 |

**Table 6.** Cut-scores, Sensitivity (Sn), Specificity (Sp), Negative Predictive Power and Positive Predictive power form OMPSQ.

| OMPSQ-SF Cut Score [Range] | Sn x100 | 1-Sp x100 | NPP % | PPP % |
|---|---|---|---|---|
| 10 [0-10] | 100 | 100.0 | 100 | 100 |
| 20 [11-20] | 100 | 95.0 | 50 | 98.76 |
| 30 21-30] | 98.8 | 80.0 | 47.05 | 89.655 |

| | | | | |
|---|---|---|---|---|
| 39 [31-40] | 90.0 | 50.0 | 41.67 | 71.71 |
| 50 [41-50] | 75.0 | 17.5 | 38.37 | 52.67 |
| 58 [51 – 60] | 52.5 | 7.5 | 33.63 | 36.21 |
| 70 [61 – 70] | 25.0 | 5 | 28.27 | 17.09 |
| 79 [71 – 80] | 8.8 | 2.5 | 5.04 | 5.04 |

## 4. Discussion

The present study compared methods for identifying feigning spectrum behaviour from genuine responding on variants of the OMPSQ and PSS, and other possible novel strategies including a forced choice auditory memory procedure and a question about finance delivered telephonically. The OMPSQ had sensitivity to discriminate between non-injured people, people with significant workplace injuries, and non-injured people asked to simulate an injury. The OMPSQ-SF appears to have a "Goldilocks" combination of questions that distinguishes between uninjured community sample members responding genuinely, workplace injured people, and community sample when asked to simulate an injury with the minimum number of questions. While the PSS and PSS-4 did statistically distinguish between the simulation conditions and people injured in the workplace these measures did not have good ROC for cut scores to identify feigning. Therefore, the PSS and PSS-4 scores should not be relied upon solely as indications of FSB, but may contribute in combination with other measures. Similarly, conclusions about high scores on the PSS and PSS-4 cannot be concluded to simply reflect stress, but may represent FSB, and furthermore the construct measured by these tests are falsifiable or able to be feigned without being distinguished from genuine responding.

Using the Linton et al. 2003 [7] and 2011 [23] cut scores of 105 and 90 respectively, 8 and 19 of IW were classified at risk of poor adjustment on the OMPSQ, and 51/80 and 62/80 of the CSS condition were classified as at risk for poor adjustment. The OMPSQ-SF cut-off of 50 classified 7 of the IW and 60/80 of the CSS as at risk for poor adjustment. This shows the differing sensitivities of these cut scores derived from the OMPSQ, with the OMPSQ-SF best resembling the original cut score sensitivity for poor adjustment. Scores above 70 on the OMPSQ-SF, identifying less than 5% of IW, likely reflects feigning spectrum behaviour, and 20 of the CSS would have been identified as simulating and 2 of the IW.

Examining the different demographic variables of the CS statistically showed no differences between variables of age, employment status on the pattern of responding in CSG and CSS conditions. Given the counterbalanced order of administration of test conditions to the CS, the strong Intraclass Correlation, and that there was no significant differences between Trial 1 or Trial to for CSS, this suggests people tend to engage in the same kind of feigning spectrum behaviour regardless of previous exposures to the test material. The data also suggests the OMPSQ and the OMPSQ-SF appear to retain sensitivity to feigning despite repeated administration and exposure to the test materials, and the opportunity to practice simulation. Other longer measures, such as the MMPI-2, have reduced sensitivity to identify feigning spectrum behaviour with coaching or repeated administrations [28,29]. The CSG inherently endorsed financial duress when simulating, and more so after previous exposure to the telephone protocol. This shows within simulated compensation scenarios people naturally equate distress with financial duress despite there being no inherent suggestion of such a relationship. Enquiring about financial wellbeing using a 10-point Likert Scale has some utility in identifying attitude to recovery in compensable samples, with scores below 5 a possible negative indicator. The CSS did not identify memory as a symptom to feign when asked to simulate a workplace injury. This may have been because the task seemed too easy, that the task seemed unrelated to the simulation of a work injury, or that it was easily identified as a possible detection strategy. Alternately, the forced choice task may have failed because the 10-items was not enough items to create the illusion that an injured person would reasonably perform poorly on it, so CSS did not try and fail this measure. At 50-items, the TOMM realistically creates the illusion of being fallible by being beyond free recall and requiring recognition, and at 10-items, the BFCT may not have created such a cognitive illusion.

The goodness of the ROC for identifying CSG on the OMPSQ and OMPSQ-SF was generally better than for the prediction of long-term recovery, with the AUC in studies of recovery ranging from 0.586 to 0.710 [7 & 23]. This likely reflects the increased accuracy of identifying a state rather than a trait in construct measurement, and it is possible that at times of measurement, these studies were confounded by low "trust" states resulting in feigning spectrum behaviour. The compensation status of the participants in these other studies was not identified and it is possible the presence of compensation can distort the predictive accuracy. For example, applying the cut scores identified above to Linton et al. [23] published results would identify 19% of those who had more than 14-days accumulated sick leave and 2.7% of those who did not have time off, as engaging in feigning spectrum behaviour.

Limitations exist in the current study. A confound of the study was that there was only one telephone assessor and the community sample were likely recognizable by voice such that a true blind could not be maintained across the experiment. A further limitation was the statistically significant gender and age differences between the IW and CS, although this did not generate statistically significant differences in responding. The ratio of males to females in the CS was roughly equivalent to the Primary Care Sample in the development of the OMPSQ-SF [23]. Research on gender and feigning spectrum behavior is somewhat limited, however recent research in both clinical and forensic settings found males tended to engage in malingering more than females [39,40] and of those deemed to be engaging in malingering as detected by an interview using the "rare" and "quasi rare" method, the Miller-Forensic Assessment of Symptoms Test, males and females did not differ significantly on their pattern of feigning behavior, except that female's report of hallucinations was more likely to be considered genuine rather than male's by expert examiners [39]. In a litigating Australasian sample, the Infrequency (F) scale on the MMPI-2 was used to compare the base rate of malingering in females and males [38], and despite more males comprising of the normative sample (57%) females were as likely to be identified as feigning as males. In the present study ANCOVA showed no significant effect for gender or age on the classification of simulating. The CS was significantly younger than the work injured sample, however, statistically age did not significantly contribute to variance. The research on age and feigning spectrum behavior indicates that young adults are more likely to admit to engaging in feigning spectrum behavior [41] and future research using the cut scores identified above could test "empathic confrontation" type interventions with young adults breaching the cut scores. Some research suggests that age matching and race matching of simulated malingerers to actual malingering populations is not necessary because the differences are not large enough to influence clinical interpretation of the pattern of responding [42] although on performance validity measures doing double duty to detect malingering are improved with age corrected scores [43]. On inspection, there were educational differences between the IW and CS. Research on FBS shows this behavior is detected more often in people with lower levels of education [4], and sophistication which may be assumed from more educated people reduced the effect size of simulation studies [29]. The effects of education status on ability to engage in FBS warrants further examination by well defined studies, and it is possible that this factor reduced the sensitivity of the cut-scores identified.

Future research with larger sample sizes could examine the relative utility of the cut scores developed above with more diverse samples. A further limitation of the current study was the same as in the development of the OMPSQ Short Form, that is, the short forms and statistics were calculated from the administration of the long form, and it is not known how completing only the short form version alone might affect responses and repeating the study with the short forms alone would contribute to the validity of their use. Given the utility and general acceptance of the OMPSQ by workers compensation systems further research into the characteristics of simulated injury is warranted with the present study representing an initiation of this process. Similar research would benefit from use of debriefing interviews to inform on simulator's strategies and sensitivity to the questions and methods. Similarly, empathic confrontation and debriefing of injured worker's identified as engaging in feigning spectrum behaviour and analysing the results would be

informative about the psychology and approaches of high stakes malingerers within compensation systems. Future research, using these or other cut-offs for feigning spectrum behaviour, could investigate tailoring interventions to address the motivational factors underlying the breaching of such cut-scores. By addressing FSB early in recovery, and assisting forensic examiners in retrospective evaluation of these factors in recovery, these and similar cut-scores can reduce unnecessary costs to society often left unchecked in our compensation systems.

**Author Contributions:** Conceptualization, methodology, statistical analysis, original draft preparation and rewriting was conducted by JM. AC and IC reviewed methods, statistics and edited and reviewed drafts.

**Institutional Review Board Statement:** The University of Sydney Research Integrity & Ethics Administration Human Research Ethics Committee 3 approved this study under cover letter dated 28th June 2019 (HREC3, 2019/136).

**Informed Consent Statement:** Oral informed consent was obtained for the use of responses to the telephone interview of injured persons for research purposes with the informed consent statement: "We are researching our support service. We do not use any personally identifying information in the research, just the ratings from the questions we ask. You can decline and it does not affect the service we will provide to support you, your inclusion in the study is voluntary. May we have your permission to use your responses to questions to improve the quality of our calls, train our staff, and for research purposes?". Oral informed consent was obtained from community research volunteers after a detailed description of the purpose of the research: "We are validating the questions that we use to assess injured workers to see how people without injuries respond and how people faking an injury may respond. We will not use any personally identifying information. You will be asked to respond genuinely and to fake an injury. There are no rewards for participating, it is voluntary, this is just to help us understand how people respond to the questions. Are you willing to participate in our research project?"

**Conflicts of Interest:** JM was the Chief Scientific Officer, a Director and Shareholder of Navigator Group Pty Ltd., the entity running the support program, during the period of study, and this may constitute a conflict of interest. The telephone assessor was a paid employee of Navigator Group Pty Ltd. and there was no other funding. IC and AC have no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| FSB | Feigning spectrum behaviour |
| SF | Short Form |
| OMPSQ | Orebro Musculoskeletal Pain Screening Questionnaire |
| PSS | Perceived Stress Scale |
| PSS-4 | Perceived Stress Sale – 4 Item |
| ANCOVA | Analysis of covariance |
| AUC | Area under the curve |
| ROC | Receiver Operator Characteristics |
| MMPI | Minnesota Multiphasic Personality Inventory |
| RF | Restructured Format |
| TOMM | Test of Memory Malingering |
| IW | Injured Workers |
| CS | Community Sample |
| CSS | Community Sample Simulating |
| CSG | Community Sample Genuine Responding |
| BFCT | Brief Forced Choice Test |
| Sn | Sensitivity |
| Sp | Specificity |
| PPP | Positive Predictive Power |
| NPP | Negative Predictive Power |

SD          Standard Deviation
Min         Minimum
Max         Maximum
N/A         Not Applicable
Sig.        Significance

## Appendix A

*Appendix A.1 Target and Foil Words for Brief Forced Choice Test*

| Targe Words | Foil |
|:---:|:---:|
| Cat | Rat |
| Boat | Coat |
| Rope | Hope |
| Van | Can |
| Book | Look |
| Hotel | Motel |
| Moose | Goose |
| Boot | Shoot |
| Pig | Jig |
| Tango | Mango |

## References

1.   WorkSafe Victoria. Clinical Framework for the Delivery of Health Services. 2009.

2.   Hockings, RL; McAuley, JH; Maher, CG. A Systematic Review of the Predictive Ability of the Orebro Musculoskeletal Pain Questionnaire. SPINE **2008**; 33(15), E494-E500

3.   Fjeld, O.; Grotle, M.; Siewers, V.; Pedersen, L.M.; Nilsen K.B.; Zwart J-A. Prognostic Factors for Persistent Leg-Pain in Patients Hospitalized with Acute Sciatica. Spine. **2017**, 42, E272-E279.

4.   Deloitte Access Economics. The Cost of Pain in Australia. **2019**: www2.deloitte.com/au/en/pages/economics/articles/cost-pain-australia.html (accessed on 30 June 2022)

5.   Mittenberg, W.; Patton, C.; Canyock, E.M.; Condit, D.C. Base Rates of Malingering and Symptom Exaggeration. J Clin Exp Neuropsychol. **2002,** 1094-1102.

6.   Malhorta, K.; Sivaraman, A.; Regunath, H. Coronavirus Disease 2019 Pandemic as catalyst for Telemedicine adoption: A single-center experience. **2020** Telemedicine Reports, 1(1), 16-21.

7.   Binder, L.M.; Rohlin, M.L. Money matters: A meta-analytic review of effects of financial incentives on recovery after closed-head injury. Am. J. Psychiatry.**1996**, 153(1), 7-10

8.   Sturgeon, J.A.; Arewasikporn, A.; Okun, M.A.; Davis, M.C.; Ong, A.; Zautra, A. The psychosocial context of financial stress: Implications for inflammation and psychological health. Psychosom Med. **2016**, 78(2), 134-143

9.   Linton, S.J.; Boersma, K.M.A. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire, Clin J Pain. **2003**, 80-86.

10.  Westman, A.; Linton, S.J.; Ohrvik, J.; Wahlen, P.; Leppert, J. Do psychosocial factors predict disability and health at a 3-year follow-up for patients with non-acute musculoskeletal pain?: A validation of the Orebro Musculoskeletal Pain Screening Questionnaire. Eur J Pain. **2008**, 12(5), 641-649.

11.  Cohen, S.; Karmarck, T.; Mermeistein, R. A global measure of perceived stress. J. Health Soc. Behav. **1983**, 24, 386-396.

12.  Cohen, S; Williamson, G. Perceived Stress in a Probability Sample of the United States. In *The Social Psychology of Health*, Spacapan, S.; Oskamp, S., Eds.; Newbury Park, CA: Sage 1988.

13.  Ng, D.M.; Jeffrey, R.W. Relationships Between Perceived Stress and Health Behaviours in a Sample of Working Adults. Health Psychol. **2003**, 22(6), 638-642

14. Cuesta-Vargas, A.I.; Gonzalez-Sanchez, M. Spanish version of the screening Orebro Musculoskeletal Pain Questionnaire: a cross-cultural adaptation and validation. Health Qual. Life Outcomes. **2014**, 12, 157-132

15. Fagundes, F.; Costa, L.; Fuhro, F.; Manzoni, A.; de Oliveira, N.; Cabral, C. Orebro Questionnaire: short and long forms of the Brazilian-Portuguese version. Qual. Life Res. **2015**, 24, 2777-88.

16. Shafeei, A.; Mokjtarinia, H.R.; Maleki-Ghahfarokhi, A.; Piri, L. Cross-Cultural Adaptation, Validity, and Reliability of the Persian Version of the Orebro Musculoskeletal Pain Screening Questionnaire. Asian Spine J. **2017**, 11(4), 520-530.

17. Gortle, M.; Vollestad, N.; Brox, J. Screening for yellow flag in first-time acute low back pain; reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. Clin J Pain. **2006,** 22, 458-67.

18. Chung, R.T.C.; Ying, J.L.S.; Fong, S.K.S.; Loi, E.S.M.; Yee, R.L.K.; Yan, T.W.F.; Chi, E.L.W. Reliability, Construct and Predictive Validity of the Hong Kong Chinese Orebro Musculoskeletal Pain Screening Questionnaire. J Occup. Rehabil. **2017**, 27, 584-592.

19. Klein, E.M.; Brahler, E.; Dreier, M.; Reinecke, L.; Muller, K.W.; Schmutzer, G.; Wolfing, K.; Beutel, M.E. The German Version of the Perceived Stress Scale-Psychometric characteristics in a representative German community sample. BMC Psychiatry. **2016**, 16, 156-164

20. Remor, E. Psychometric properties of a European Spanish Version of the Perceived Stress Scale. The Span. J. Psychol. **2006**, 9(1), 86-93.

21. Lu W.; Bian, Q.; Wang, W.; Wu, X.; Wang, Z.; Zhao, M. Chinese version of the Perceived Stress Scale 10: A psychometric study in Chinese university students. PLoS One. **2019**, 12 (12), 1-8

22. De Silva, T.; Hodges, P.W.; Costa, N.; Setchell, J. Potential unintended Effects of Standardized Pain Questionnaires: A Qualitative Study. Pain Med. **2020**, 2, e22-e33.

23. Linton, S.J.; Nicholas, M.; Macdonald, S. Development of a Short Form of the Orebro Musculoskeletal Pain Screening Questionnaire. Spine, **2011**, 36 (22), 1891-5.

24. Warttig, S.L.; Forshaw, M.J.; South, J.; White, A.K. New, normative, English-sample data for the Short Form Perceived Stress Scale (PSS-4). J Health Psychol. **2013**, 18 (12), 1617=28.

25. Cohen, S. Perceived Stress in a probability sample of the United States. In S. Spacapan 7 S. Oskamp. The social psychology of health. 1988:31-67.

26. Young, G. Malingering, Feigning, and Response Bias in Psychiatric/Psychological Injury: Implications for Practice and Court. International Library of Ethics, Law and The New Medicine, Springer, 2014, Vol. 56.

27. Rogers, R. Clinical Assessment of Malingering and Deception. 4th ed. New York: Guilford Publications; 2018.

28. Storm, J.; Graham, J.R. Detection of coached general malingering on the MMPI-2. Psychol Assess. **2000**, 12 (2), 158 – 165.

29. Wygant, D.B; Ben-Porath, Y.; Arbisi, P.A.; Berry, D.T.R.; Freeman, D.B.; Heilbronner, R.L. Examination of the MMPI-2 Restructured Form (MMPI-2-RF) Validity scales in civil forensic settings: Findings from simulation and Known Group Samples. Arch Clin Neuropsychol. **2009**, 24 (7), 671 – 680.

30. Bagby, R.M.; Nicholson, R.A.; Bacchiochi, J.R.; Ryder, A.G.; Bury, A.S. The predictive capacity of the MMPI-2 and PAI validity scales and indexes to detect coached and uncoached feigning. J Pers Assess. **2002**, 78 (1), 69-86.

31. Greve, K.W.; Etherton, J.L.; Ord, J.; Bieanchini, J.; Curtis, K.L. Detecting Malingered Pain-Related Disability: Classification Accuracy of the Test of Memory Malingering. Clin. Neuropsychol. **2009**, 23 (7), 1250 – 71.

32. Iverson, G.L.; Franzen, M.D. A brief assessment instrument designed to detect malingered memory deficits. The Behav Ther. **1993**, 6, 134-5.

33. Cohen J. A Power Primer. Psychol. Bull. **1992**, 112 (1), 155-159.

34. Faul, F.; Erdfelder, E.; Lan, A-G; Buchner, A. G*Power: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. Behav Res Methods. **2007,** 39 (2),175-191.

35. Linton, S.H.; Hallden, K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain. **1998**, 3, 209-215.

36. Miller, G.A.; Chapman, J.P. Misunderstanding Analysis of Covariance. J. Abnorm. Psychol. **2001**, 110 (1), 40-48.

37. Shoukri, M.M.; Asyali, M.H.; Donner, A. Sample size requirements for the design of reliability study: review and new results. Stat. Methods Med. Res. **2004**, 13, 251-271.

38. . Clifford, D.; Byrne, M.K.; Allan, C. Getting Caught in Court: Base Rates for Malingering in Australasian Litigants. Psychiatry Psychol Law. **2004**, 11 (2), 197-201.

39. Udoetuk, S.; Dongarwar, D.; Salihu, H.M. Racial and Gender Disparities in Diagnosis of Malingering in Clinical Settings. J Racial Ethn Health Disparities. **2020**, 7, 1117-1123.

40. Nesbit-Bartsch, A.E.; McDermott, B.E.; Warburton, K.D. Gender and Malingering in Defendants Deemed Incompetent to Stand Trial. J. Am. Acad. Psychiatry Law. **2021**, 49 (2), 211-218.

41. Sierles, FS. Correlates of Malingering. Behav Sci Law. **1984**, 2(1), 113-118.

42. Brennan, A.M. Are we Honestly Studying Malingering? A Profile and Comparison of Simulated Suspected Malingerers. Master of Arts, Louisianna State University, December 2004, LSU Digital Commons.

43. Ashendorf, L.A.; McCaffrey, R.J. Specificity of malingering detection strategies in older adults using the CVLT and WCST. Clin Neuropsychol. **2003**, 17 (2), 255-262

44. Svete, LJ, Tindell WW, McLouth CJ, Allen TS. A retrospective analysis of rates of malingering in a a forensic psychiatry practice. J Am Acad Psychiatry Law. **2025**, 54 (1), 1-9.