Yellow flag scores in a compensable New Zealand cohort suffering acute low back pain

Karen Grimmer-Somers[1]
Mathew Prior[1]
Jim Robertson[2]

[1]Centre for Allied Health Evidence, University of South Australia, City East Campus, North Tce, Adelaide, South Australia, Australia; [2]New Zealand Accident Compensation Corporation, Auckland, New Zealand

Background: Despite its high prevalence, most acute low back pain (ALBP) is nonspecific, self-limiting with no definable pathology. Recurrence is prevalent, as is resultant chronicity. Psychosocial factors (yellow flags comprising depression and anxiety, negative pain beliefs, job dissatisfaction) are associated with the development of chronic LBP.

Methods: A national insurer (Accident Compensation Corporation, New Zealand [NZ]), in conjunction with a NZ primary health organization, piloted a strategy for more effective management of patients with ALBP, by following the NZ ALBP Guideline. The guidelines recommend the use of a psychosocial screening instrument (Yellow Flags Screening Instrument, a derivative of Örebro Musculoskeletal Pain Questionnaire). This instrument was recommended for administration on the second visit to a general medical practitioner (GP). This paper tests whether published cut-points of yellow flag scores to predict LBP claims length and costs were valid in this cohort.

Results: Data was available for 902 claimants appropriately enrolled into the pilot. 25% claimants consulted the GP once only, and thus were not requested to provide a yellow flag score. Yellow flag scores were provided by 48% claimants who consumed two or more GP services. Approximately 60% LBP presentations resolved within five GP visits. Yellow flag scores were significantly and positively associated with treatment costs and service use, although the association was nonlinear. Claimants with moderate yellow flag scores were similarly likely to incur lengthy claims as claimants with at-risk scores.

Discussion: Capturing data on psychosocial factors for compensable patients with ALBP has merit in predicting lengthy claims. The validity of the published yellow flag cut-points requires further testing.

Keywords: acute low back pain, yellow flags, clinical guideline, lengthy claims

Background

Low back pain (LBP) is a significant public health problem in the Western world, with as many as 84% of the general population believed to experience LBP at some point in their lives (Walker 2000). Despite its high prevalence, most LBP is typically nonspecific, with almost all events occurring in the absence of clearly definable pathology (NHMRC 2003). Whilst acute LBP is typically self-limiting, recurrence is prevalent and chronicity not uncommon (Bekkering et al 2003; NHMRC 2003). LBP impacts on the individual, and on workplaces, families and the wider community.

The New Zealand Accident Compensation Corporation (ACC) is a national compensable body. Its role, broadly speaking, is to provide insurance for all New Zealanders, injured in road accidents, at work, and in the community (ACC 2007). LBP places a substantial burden on health care resources, with such injuries costing ACC an estimated NZ$130 million per annum. Despite this spending, ACC claimant outcomes from LBP are not always satisfactory.

The use of clinical guidelines has been proposed as a way of improving health care processes and outcomes (Grimshaw et al 2004; Barosi 2006). Supporting an

Correspondence: Grimmer-Somers
Centre for Allied Health Evidence,
University of South Australia, City East
Campus, North Tce, Adelaide, South
Australia
Tel +61 8 83022769
Fax +61 8 83022766
Email karen.grimmer-somers@
unisa.edu.au

15

increasing consumer expectation of evidence-based health care, clinical guidelines are being produced for large numbers of conditions. LBP is no exception, with at least ten clinical practice guidelines currently in use around the world. Perceived benefits of the use of clinical guidelines include a reduction in inappropriate practice variations, increased clinical efficiency and support, and better control of health care spending (Lenfant 2003; Grimshaw et al 2004). The research evidence is equivocal however, regarding the real cost and health outcomes of guideline-based care.

Recent advances in understanding of factors associated with LBP have recognised the influence of psychosocial factors on health outcomes, with current LBP guidelines reflecting this (Kendall et al 1997; Bekkering et al 2003; NHMRC 2003; Holohan et al 2006). Psychosocial risk factors, including depression and anxiety, negative pain beliefs, low mood and job dissatisfaction, have strong associations with poor health outcomes and chronicity of LBP (Bekkering et al 2003; NHMRC 2003; ACC 2004). Kendall (1999) proposes that many chronic pain behaviors are observable in the first few days of incurring ALBP. However, few studies have investigated the efficacy of strategies to prevent ALBP from developing into chronic pain (Newton-John et al 2001).

Most LBP clinical guidelines consistently advocate early identification of LBP sufferers with at-risk psychosocial symptoms and behaviors, so that early targeted interventions can be put in place (Kendall et al 1997; Bekkering et al 2003; NHMRC 2003; Holohan et al 2006). The Yellow Flags Screening Instrument (YFSI) (Kendall et al 1997) is an early version of the Örebro Musculoskeletal Pain Questionnaire (Linton and Boersma 2003). It is recommended as a mechanism to assess psychosocial factors associated with ALBP. Psychosocial factors include fear-avoidance behavior, low mood/withdrawal, expectation of passive treatment and negative pain beliefs (Kendall et al 1997; Turk 1997; Linton and Hallden 1998; Kendall 1999; Linton and Boersma 2003). The YFSI provides the following risk thresholds and descriptors (Kendall et al 1997, p 34–35):

- 'Scores >105: "…will enable you to identify at least three-quarters of the long-term cases."
- Scores 90–105: "This group will include a proportion who are at risk for medium-term problems."
- Scores < 90: "Most of these patients will exhibit recovery within the expected time period.'

The YFSI is reported to identify '75% cases correctly of those not needing modification to ongoing management, 86% correct identification of those who will have between 1 and 30 days off work, and 83% correct identification of those who will have more than 30 days off work' (Kendall et al 1997, p 39). There is scant literature which validates these claims in other ALBP populations.

This paper reports on the validity of the Kendall et al (1997) yellow flag score cut-points to predict LBP claims length (and higher claims costs) for ACC ALBP claimants during 2006.

Methods
Ethics
Ethical approval for the study was provided by the University of South Australia Human Research Ethics Committee (Australia), New Zealand Health and Disability Northern Y Regional Ethics Committee, and an internal ACC Ethics Committee.

Setting
Auckland region, New Zealand.

Study background
ACC, in conjunction with a primary health organization (PHO), devised and piloted a strategy to facilitate more streamlined and effective management of patients with acute LBP, by following a 'guidelines-based' treatment protocol. PHOs are local structures for delivering and co-ordinating primary health care services. PHOs bring together doctors, nurses and other health professionals (such as Maori health workers, health promotion workers, dieticians, pharmacists, physiotherapists, psychologists and midwives) in the community to serve the needs of their enrolled populations. This pilot strategy was based on the New Zealand Acute LBP Guide (ACC 2004), a recommended clinical practice guideline (pathway) framed on a review of the relevant scientific literature within which the YFSI is embedded (Kendall et al 1997). It was envisaged that this pilot strategy would result in better claimant health outcomes, with corresponding cost benefits to ACC. The pathway recommended five general medical practitioner (GP) occasions of service in the episode of care for ALBP, and the application of the YFSI at the second visit. An episode of care is a set of congruent health care visits (occasions of service) for the management of the one condition (Grimmer et al 2000). If claimants had at-risk yellow flag scores, and/ or if their episode of care necessitated more than five GP visits, then they should be referred to psychological services. A payment of over NZ$100 for the second visit was provided to GPs who assessed patients for yellow flags at the second visit, as per the ALBP protocol.

Subjects

ACC claimants presenting to the GPs registered with the affiliated GP organization between February and December 2006 were eligible to be enrolled into the guidelines pilot program. The time period over which claimant data was analyzed was March–December 2006, which accounted for approximately 98% of the enrolments into the pilot program. Approximately 85% GPs in the PHO participated in the pilot program. The guidelines recommended up to five visits to the participating GP as required, with the second visit being an extended consultation at which information on psychosocial factors (yellow flags) was recommended for capture, using a dedicated IT decision-making tool, linked to the GP practice management system. Claimants whose ALBP was not resolving after the 2nd GP visit were eligible, under the treatment guidelines, to be referred to specialist pain management services (Physiotherapy, Psychology). Repeat yellow flag scores were captured as required, at subsequent visits during the episode of care, if the claimant's symptoms were not resolving. Yellow flag assessment was however, not mandatory for claimants or GPs. Reasons for GPs not assessing claimants for yellow flags could be claimant refusal, GP disinterest, GP failure to follow the clinical pathway, or GP and claimant decision that sufficient improvement in symptoms had occurred such that the claimant would require few further treatments (making yellow flag assessment unnecessary).

Red flags

Red flags are symptoms or history indicating immediate referral to a specialist. Red flags for acute LBP can include fractures, infection, symptoms of cancer, and psychiatric and neurological disturbances (NHMRC 2003). The ACC clinical pathway recommended that claimants with suspected red flags be referred immediately for treatment elsewhere.

Yellow flags

The published classifications of psychosocial risks (measured as yellow flags) were applied to the claimant cohort in this pilot program (Linton and Hallden 1998). However this data was not derived on a NZ-born population and therefore its utility had not been validated. Following the yellow flag risk thresholds and descriptors (Kendall et al 1997), claimants with yellow flag scores were classified >105 (high risk), between 90 and 105 (moderate to high risk) and 50–89 (moderate risk) (Kendall et al 1997; Linton and Hallden 1998). Claimants with scores lower than 50 were not considered to be at-risk of chronic pain behaviors, and thus were designated as the comparison group for all analysis.

Claims data

Claims were analyzed as number of services within the episode of care, and the length of time over which the episode of care (open claim) extended (claims length). The number of days lost from work was also examined. ACC legislation allows for lost earnings compensation to be paid after seven days off work. Therefore, absences of less than seven days from work are not present in the data. Thus it is not known how many claimants took no days off work, or less than seven days. The number of claimants in this data set who took more than seven days off work because of their ALBP is therefore not representative of claimants who took no, or less than seven, days off work.

Data analysis

Deidentified amalgamated data was provided on ACC claimant data for the period of the pilot program. The dataset included information on number and type of health service consultations per event of acute LBP (episode of care), demographic information (age, gender, self-reported ethnicity), claim duration and total cost, the number and type of health contacts, 'red flag' status and 'yellow flag' scores (if provided). Raw claims data was tabulated using MS Office Excel (Microsoft Corp., Redmond, WA, USA), and analysed using SAS Version 8.2 (SAS Inc., Cary, NC, USA).

The investigations reported in this paper considered only those claimants who consumed more than one GP visit. They were thus eligible to provide a yellow flag score on their second visit, as recommended in the guidelines. Gender, age and ethnicity characteristics, and claims length and cost, were described for those claimants who provided, or not, yellow flag scores using median (25th, 75th range) for the equal interval data, and percentages for the categorical data. Analysis of Variance models (for equal interval data) and chi square tests (for categorical data) were used to establish significance (set at $p < 0.05$ for all statistical tests). Significant F and chi square values, and degrees of freedom (df) are reported.

Within the subgroup of claimants which provided yellow flag scores, the characteristics of the different yellow flag classifications (outlined by Linton et al 1999) were investigated in a similar manner. Univariate logistic regression models using independent categories of yellow flag scores were used to identify whether the yellow flag score classifications linearly predicted claims outcomes and costs. The comparison group was claimants with yellow flag scores 0–49 (considered to have low risk of chronicity). Significant odds ratios were identified when the

95% confidence intervals (CIs) did not encompass 1. The confounding effect of ethnic status, gender and age on this association was then investigated using step-wise multivariate logistic regression models. The confounding effect in the multivariate regression models was assessed during construction of the stepwise models by the significance of the amount of change in the Likelihood Ratio (df, *p value*).

Hypotheses underpinning testing

There is a positive linear relationship between yellow flag scores, claims length, and claims costs.

Results
Claimants

Figure 1 outlines the consort diagram for the pilot program, in which 902 eligible claimants were appropriately enrolled into this pilot program, 223 claimants consulted a GP once only and thus did not enter the clinical pathway, and of the remainder, 328 provided yellow flag scores.

Median age of the 902 eligible claimants was 38 years (25th% to 75th% 30–46 years) (range 16–72 years), with males predominating (58.8%). There was no significant gender difference in mean age. These claimants self-reported ethnicity, comprising European (42.8%), Asian (12.8%), Maori (12.2%), Middle Eastern/Latin American/African 1.4%, Pacific Islanders 18.7% and other (Indian, etc) (12.1%). The demographic profile of the ACC cohort did not differ from that of the greater Auckland city area, North Shore City, or Manukau City (Statistics New Zealand 2002), suggesting generalizable findings to other LBP sufferers.

Diagnosis was assigned by the treating GP using ACC codes. The predominant diagnosis was lumbar sprain (73.3%), followed by lumbar spine pain (7.8%), other (undefined) (7%), thoracic sprain (3.9%), lumbar disc prolapse with radiculopathy (3.1%), sacroiliac ligament sprain (2.2%), sciatica (1.8%), back sprain (0.6%), and lumbosacral strain (0.3%). There was no data available of inter-or intra-rater GP reliability when assigning these diagnostic codes, or the appropriateness of the diagnostic codes.

Claims information

The eligible claimants (N = 902) consumed 5976 health services during the pilot program (median 3 services per claimant, range 1–74) (25th% to 75th% 2–9). Claimants could not enter the ACC clinical pathway without an initial GP visit, which was the only contact for 223 claimants (24.7% total) (that is, they required no further treatment of any kind and thus did not proceed down the clinical path). These claimants were excluded from the analysis reported in this paper. A range of services were consumed by the remaining claimants (those recording two or more visits to health providers), namely physiotherapy (44.6%), GP (35%), acupuncture (8%), chiropractics (4.6%), diagnostic radiology (2.2%), osteopathy (1.9%), medical specialties (3.6%). Overall, this cohort cost ACC $275,783.54 (median cost $148.09 per claimant for an episode of care for acute LBP) (25th%–75th% $49.60 to $366.02).

Yellow flag scores

For the remaining claimants (N = 679), the clinical pathway recommended assessing psychosocial risk using the YFSI on the second GP visit. Yellow flags data was available for 328 claimants (48.8%) who consumed two or more GP (See Figure 1).

Considering the claimant cohort which attended the GP for two visits or more, there was no significant difference between claimants who provided yellow flag scores and those who did not (in age, gender proportions, diagnostic categories, occasions of service, length of episode of care, or the percentage of claimants who took more than seven days off work (and the number of days they took off work)). Occasions of service reflected any health service consumption, not just general medical practitioners. There was a significant difference in unadjusted claims costs however, between the two groups, with the yellow flag claimant group having high median costs which reflected the increased payment for assessing yellow flag scores on an extended second visit consultation. When this payment was removed from the total claims costs for this group, the difference between groups was no longer significant (adjusted median total cost for the yellow flag group $155.14 (25th% $78.40, 75th% $386.30). The comparison data is provided in Table 1.

To add weight to the similarity in service delivery consumption patterns (and adjusted total costs) of the claimants who provided, or not, yellow flag scores within this cohort, similar distribution of the numbers of occasions of service in the claimants' episodes of care for their ALBP is provided in Figure 1. Thus the findings from the claimants who consumed two or more GP visits and provided yellow flag scores are generalizable to those who also consumed two or more GP visits, but did not provide yellow flag scores.
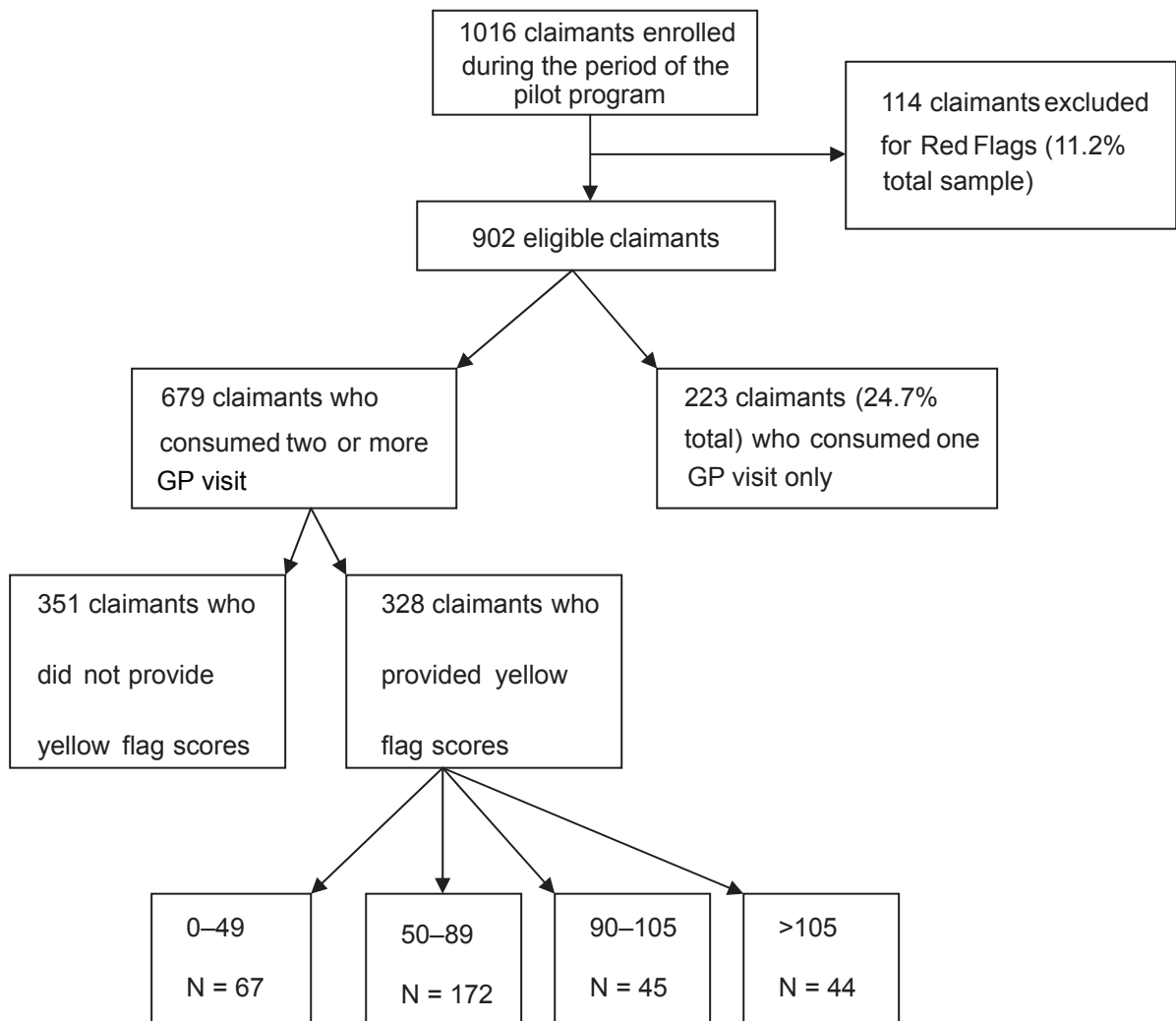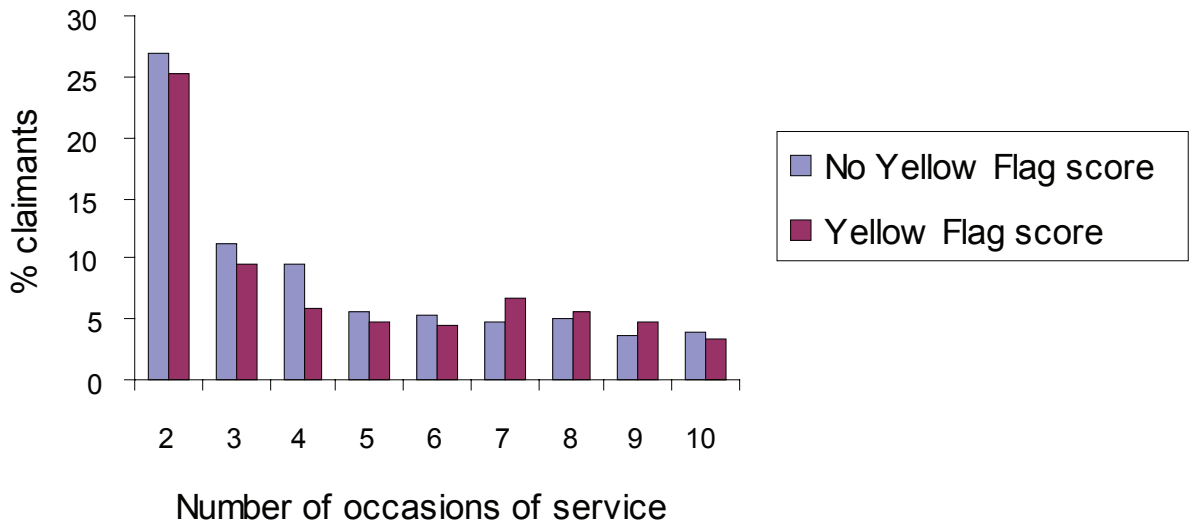
Figure 1 Distribution of occasions of service (2+) for claimants with and without yellow flag scores.

Table 1 Demographic profile of claimants consuming two or more GP visits, and providing, or not, yellow flag scores

	Claimants not providing yellow flag scores	Claimants providing yellow flag scores
Age (Median 25th%, 75th%) [range]	38 (30–46) [18–85]	39 (31–47) [18–75]
Claims length (Median 25th%, 75th%) [range]	27 (9–105) [3–312]	19 (10–66.5) [1–389]
Occasions of service (Median 25th%, 75th%) [range]	5 (2–10) [2–56]	6.5 (2–11.5) [2–74]
Days off work >7 (Median 25th%, 75th%) [range]	Incurred by 6.3% 22 (10–105) [4–312]	Incurred by 10.1% 27 (8–64) [1–389]
Total unadjusted cost of claim (Median 5th%, 75th%) [range]	**$154.04 ($77.33–$385.19) [$48.89–$3838.24]**	**$265.78 ($148.09–$541.55) [$58.32–$5476.81)]**
Gender	57.3% Male	55.6% Male
Most common diagnosis	Lumbar sprain (71.2%)	Lumbar sprain (77.6%)
Ethnicity		
• European	37.8%	51.3%
• Asian	15.1%	8.8%
• Maori	13.2%	10.3%
• Pacific Islander	19.3%	17.8%
• Middle Eastern/Latin American/African	1.6%	1.2%
• Other	13.0%	10.6%

Note: Significant differences identified by bold text.

Claimants with yellow flag scores

Claimant yellow flag scores ranged from 10 to 146. Of the claimants who provided yellow flag scores, N = 67 (20%) had scores less than 49 (deemed to be low risk of developing chronic LBP). N = 175 (52.9%) had yellow flag scores between 50–89 (some risk of chronic LBP, whilst N = 45 (13.6%) had scores between 90 and 105 (moderate to high risk of chronic LBP. Forty-four claimants (13.3%) had at-risk scores >105. This indicated, using the scoring criteria, that over 73% of the claimants who provided yellow flag scores had no, or minimal, risk of developing persistent pain behaviors.

Demographic and service consumption differences in yellow flag claimant groups

Demographic characteristics of the claimants with yellow flags are reported in Table 2. There was no need to adjust the cost of claims in this yellow flags subset as all GPs had the opportunity receive the extended visit payment for collecting yellow flag scores on the second visit. There was no significant difference in the frequency of gender, common diagnostic condition (lumbar sprain) and the number of days off work greater than seven, across the yellow flag groups. There was a significant difference across the groups in the frequency of ethnicity (chi square 35.6 18df) $p < 0.05$, as well as significant differences (all with 3df) in the median length of claims (F = 11.5), the number of occasions of service in the episode of care (F = 8.54), and total claims costs (F = 8.9) ($p < 0.05$). Occasions of service reflected any health service consumption, not just general medical practitioner visits. There was also a significant difference in the percentage of claimants in each yellow flag category who took more than seven days off work (F = 7.8) ($p < 0.05$).

Ethnic differences

The significant influence of ethnicity on psychosocial factors is evidenced by the distribution of yellow flag scores (Figure 2) in which the score distribution of the Middle Eastern/Latin American/African group scores was significantly higher than the 75th% of the other ethnic groups. Differences in service costs are illustrated in Figure 3 by the significantly higher distribution of scores for the Asian claimants compared with all others.

Treatment costs

Compared to claimants with yellow flag score less than 50 (not at-risk of chronic pain behaviors), all other claimants with yellow flag scores were significantly more likely to incur treatment costs in excess of the overall median cost per claimant (NZ$148.09). Significance was indicated as the 95% CI did not encompass 1 for any of the associations. For the low risk of LBP chronicity (yellow flag scores of 50–89, the unadjusted odds of incurring claims costs in excess of $148.09 was 6.8 (95% CI 4.5–10.3). For the moderate risk

Table 2 Characteristics of claimants in yellow flag groups

	Low yellow flag scores 0–49	Moderate yellow flag scores 50–89	Moderate to High yellow flag scores 90–105	At-risk yellow flag cores >105
Age (median, 25th%, 75th%)	40 (29–47) [18–69]	38.5 (31–47) [18–75]	39.5 (31.5–48.0) [20–57]	37 (31–46) [19–73]
Claims length (median, 25th%, 75th%) [range]	**12.0 (7–30) [0–254]**	**32 (10–86) [2–498]**	**46 (17.5–78.5) [4–370]**	**71 (23–238) [3–396]**
Days off work >7 (% yellow flag group) Median (25th%, 75th%), range	1.5% 4 (4–4), [4]	12.6% 27 (14-60) [7-287]	20% 12 (5-20) [3-339]	27.3% 49.5 (32.5-214.5) [1–389]
Occasions of service (median, 25th%, 75th%), [range]	3 (2–7) [2–37]	6 (2–12) [1–44]	8 (3–11) [1–46]	9.5 (5–16.5) [2–74]
Total cost of claim (median, 25th%, 75th%)	**$148.09 ($100.33–$325.26) [$58.32–$1433.16]**	**$258.85 ($148.09–$541.55) [$58.32–$3974.10]**	**$350.52 ($148.09–$485.29) [$58.32–$5421.28]**	**$740.75 ($265.65–$1560.38) [$58.32–$5476.81]**
Gender (Male)	55.2%	51.8%	61.4%	65.1%
Most common diagnosis	Lumbar sprain (77.6%)	Lumbar sprain (70.6%)	Lumbar sprain (79.6%)	Lumbar sprain (69.9%)
Ethnicity				
• Asian	14.9%	7.1%	6.8%	6.9%
• European	40.3%	56.5%	50.0%	48.8%
• Maori	4.5%	12.4%	11.4%	11.6%
• Pacific Islander	0.0%	0.0%	2.3%	6.9%
• Middle Eastern/Latin American/African	14.9%	14.1%	25.0%	13.9%
• Other	23.9%	10.0%	4.6%	12.5%

Note: Significant differences identified by bold text.
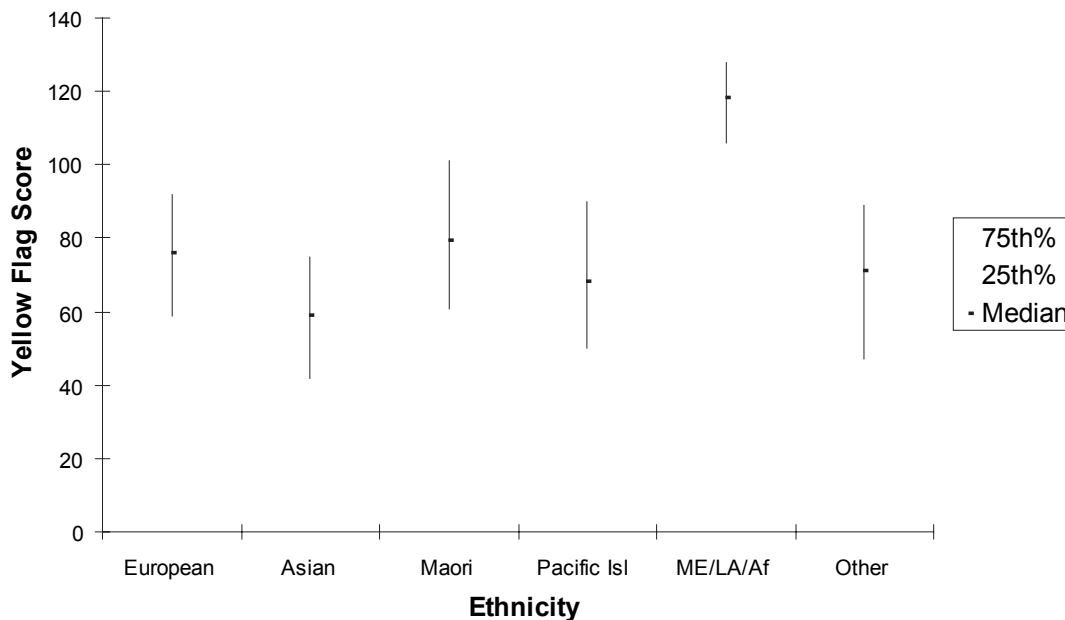

Figure 2 Yellow flag score by ethnicity.
Abbreviation: ME/LA/Af, Middle Eastern/Latin American/African.
Note: Vertical lines represent interquartile range (25th–75th percentile).
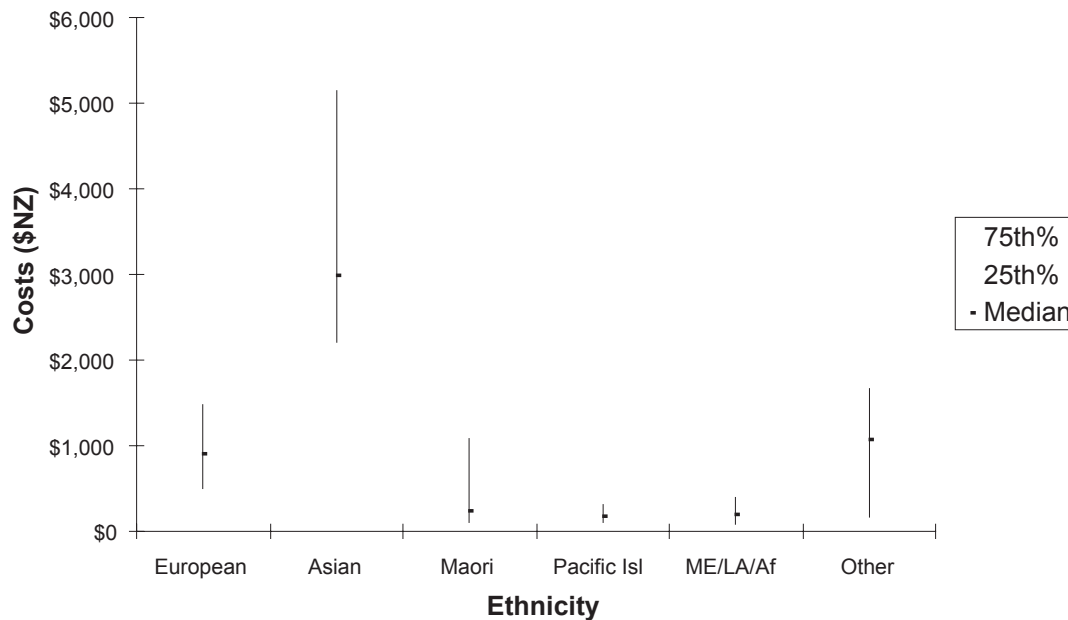
Figure 3 Per claimant treatment costs, by ethnic group ('High-risk' claimants only).
Abbreviations: ME/LA/Af, middle eastern/latin american/african.
Note: Vertical lines represent interquartile range (25th–75th percentile).

of LBP chronicity group (yellow flags scores 90–105), the unadjusted odds was 15.6 (95% CI 5.5–44.1), and for the high risk of LBP chronicity group (yellow flag score >105), the unadjusted odds was 11.9 (95% CI 4.6–30.5). This supports the general hypothesis that the higher yellow flag scores, the greater the likelihood of higher treatment costs.

Chronicity reflected by long claims: Considering claimants with yellow flag scores, and considering the median overall claims length (11 days), compared to claimants with low yellow flag scores (less than 50), 'high-risk' claimants (yellow flag scores >105) were significantly (seven times) more likely to have a claim lasting greater than the median claims length (unadjusted odds ratio [95%CI] 7.1 [2.7–19.1]). Claimants with yellow flag scores between 50–89 were also significantly more likely to become chronic (unadjusted odds ratio (95%CI) 2.5 (1.1–5.9), however claimants scoring between 90–105 did not have a significantly elevated risk compared with the not-at-risk claimants (unadjusted odds ratio (95%CI) 1.9 (0.6–5.5).

Adjusting for confounding

Significant change in the likelihood ratio (LR) in the stepwise multivariate logistic regression models indicated that ethnicity was a significant confounder on the association between yellow flags classifications and total cost (LR 20.2, 8df, p < 0.05) and yellow flags classifications and claims

length (LR 12.3, 8df, p < 0.05). The addition of gender or age did not significantly increase the LRs (LR 3.1, 9df; LR 1.9, 9df, respectively). The adjusted odds of claimants in the different yellow flags categories incurring costs higher than $148.09 or claims lasting longer than 11 days is reported in Table 2. The probability that higher yellow flag scores will predict high cost and claims length was increased in all yellow flag score categories, compared with the low risk category. There was similar risk of high cost claims occurring in the moderate, and moderate to high categories (in the range of 2–3 times the risk), with strongly elevated probability (nearly 8 times the risk) in the at-high risk group. Comparing the unadjusted and adjusted odds ratios for high cost, it appears that the confounding effect of ethnicity was greatest in the moderate to high risk yellow flag score category, as thus group showed the greatest change between unadjusted and adjusted odds. There was a different and linearly increasing risk pattern across the categories for long claims length, ranging from OR of three in the moderate risk group, six in the moderate to high risk group, and nine in the at-risk group). The moderate to high risk group again showed the greatest effect of deconfounding.

Claims lasting 90 days or more

The yellow flags classifications claim to be able to predict claimants who will have extended claims. ACC defines

extended claims as those lasting greater than three months (>90 days). Overall, of all claimants in the ACC pilot cohort (N = 902), 138 (15.3%) had claims which were not closed within 90 days. Sixty-three of this subset of claimants (45.7%) did not have a recorded yellow flag score, indicating that their care had not complied with the ACC clinical pathway.

Confounding by ethnicity but not by gender or age was the same for this analysis, as indicated by the change in LRs at each step (LRs 19.6 8df, 3.8 9df, 3.9 9df, respectively). Considering the adjusted odds ratios from the step wise regression models, there were significant and elevated probabilities that moderate, and at-risk yellow flag classifications would predict claims lasting longer than 90 days, compared to low risk yellow flag scores. However, despite deconfounding by ethnicity, the risk of claims >90 days in the moderate to high risk yellow flag group was no different to the low risk group, as outlined in Table 3. These findings suggest that the classifications of yellow flag scores require revising to more accurately predict lengthy claims.

To consider the sensitivity and specificity of the proposed yellow flag score classifications in this cohort of ALBP claimants, the number of claimants in the four yellow flag score categories with respect to extended (90 + days) claims length is presented in Table 4. The YFSI performed best when identifying claimants with low yellow flag scores who would not become chronic cases (specificity of 84.6%). Sensitivity was poorer (26.7%), as only 20 of the 75 claimants with a claim lasting in excess of 90 days, had yellow flag scores greater 105, and were correctly identified as at-high-risk for chronicity. Of the 44 claimants who had at-risk yellow flag scores (>105), less than half became chronic cases. Conversely just over one-fifth of low to moderate risk claimants (scores 50–89), and just under one-fifth of moderate to high risk claimants (scores 89–105) had extended claims, which refutes the notion that high yellow flag scores are sensitive predictors of lengthy claims. Considering claimants with yellow flag scores of 90+, only 45.6% were correctly identified as consuming claims longer than 90 days, which is considerably lower than the predicted percentage (89%) (Linton and Boersma 2003).

Referral
Less than a third (31.8%) of claimants deemed at 'high-risk' due to their yellow flag score were referred to psychological services, as recommended in the clinical pathway. Furthermore, of those who were referred, claimant utilization of these services was poor, with only three claimants completing treatment. Given the constraints on the data provided by ACC for this paper, it could not be determined why this low referral or completion rate occurred.

Discussion
This paper reports on a rare opportunity to examine psychosocial risk related to cost and claims length, in a large northern-New Zealand cohort of compensable claimants with ALBP. Claimant demography is representative of the wider northern New Zealand population. The clinical pathway underpinning the management of claimants with ALBP provided financial incentives for GPs to collect yellow flag scores from claimants who attended twice or more. Of the ACC claimants who had this opportunity to provide yellow flags scores, the characteristics of the claimants who provided yellow flag scores were no different from those who did not provide yellow flag scores. We believe that this sample is generalizable to other compensable claimants with ALBP.

Data limitations
Of those claimants who attended their GP at least twice and thus were eligible to provide a yellow flag score, less than half did so. This raises questions about yellow flag scores for the noncompliant subset. In health services research which evaluates processes and outcomes of service delivery programs applied in 'everyday' settings, it is not possible to predict how individual practitioners and claimants will behave, despite them agreeing to comply with established

Table 3 Adjusted odds of high costs and high claims length occurring in yellow flag groups, compared with the lowest risk group (AOR, 95% CI)

	Lowest risk yellow flags <50	Moderate risk yellow flags 50–89	Moderate to high risk yellow flags 90–105	At-risk yellow flags >105
Costs > $148.09	1 (comparison)	2.2 (1.2–4.1)	3.3 (1.4–8.1)	7.6 (2.6–21.7)
Claims lasting > 11 days	1 (comparison)	2.8 (1.5–5.1)	6.2 (2.3–16.4)	9.4 (3.0–29.0)
Claims lasting > 90 days	1 (comparison)	3.4 (1.3–8.9)	2.6 (0.8–8.3)	12.1 (4.0–36.2)

Note: Significant associations highlighted in bold.

Table 4 Distribution of claimants by yellow flag scores and claim length

	Yellow flag score			
	<50	50–89	90–105	>105
Claims ≤90 days	60	132	37	24
Claims >90 days	7	40	8	20
% all claims >90 days	9.3%	53.3%	10.7%	26.7%

program protocols. On reflection, at least 100 claimants who did not provide a yellow flag score can potentially be accounted for. Figure 1 reported that approximately 25% of the claimants who consumed two visits or more, and provided no yellow flag scores, only consumed two occasions of service within their episode of care. Furthermore approximately 10% only consumed three occasions of service in their episode of care. It is possible that, if the patient and GP predicted the imminent closure of the ALBP episode of care, they may have legitimately decided not to record a yellow flag score. This could be verified by the nonpayment of the second visit payment to these claimants' GPs. If this assumption is correct, the percentage of eligible yellow flag claimants who did not provide yellow flag scores, and, could be reduced to approximately 30% for whom no reasonable explanation could be found other than noncompliance.

There is little rationale in the ALBP clinical guidelines (ACC 2004) for collecting yellow flag scores on the second GP visit. While chronic pain behaviors are believed to be exhibited early in an episode of ALBP, it is not known whether capture of information on psychosocial risk at the first, second or subsequent visits to the GP is appropriate. Despite the increased second visit payment for assessing yellow flag scores. GPs may not have been sufficiently motivated or confident to capture this information from their patients, particularly if these practice behaviors were foreign to them. Further information needs to be obtained from GPs and their patients regarding their perspectives on the use of a standard protocol for the management of ALBP.

Crawford and colleagues (2007) suggests that for many GPs, identification and management of psychosocial factors depends on their world-view and orientation to the biopsychosocial model of pain. This may be reflected in the infrequent GP use of the YFSI in the management of patients with ALBP, despite the payment. Many GPs report treating patients purely according to biomedical principles initially, until a lack of progress was evident (Crawford et al 2007). Thus administering the YFSI on the second GP visit (as stipulated in the ACC clinical pathway) may not have

coincided with GPs' beliefs or usual management practices. It may also not have been acceptable to claimants. In the same vein, referral of claimants with high yellow flag scores to psychological services may not have coincided with GPs or patients' beliefs, which may have accounted for the low psychological counselling referral and treatment completion rate. These issues need to be better understood, as compliance by clinician and patient with any agreed clinical protocol is essential to its effectiveness (Grimshaw et al 2004).

Yellow flag scores

Yellow flag scores were significantly confounded by ethnicity but not gender or age. Claimants of middle Eastern ethnic background had significantly higher yellow flags than any other racial group. Considering the overall group median scores, increasing yellow flag scores were generally linearly related to higher claims costs and claims length. This concurs with the claims of the YFSI developers that the higher the yellow flag score, the more likely it is to predict long claims.

Claims longer than 90 days

When predicting claims lasting longer than 90 days however, despite deconfounding for ethnicity, claimants with scores between 50–89 (who traditionally represent a low to moderate risk (Linton and Hallden 1998; ACC 2004)), recorded a stronger probability of incurring a lengthy claim than claimants scoring 90–105. This finding challenges the notion that claimants with scores less than 90 are considered together as posing similarly and low level of risk for poor outcomes (Linton and Hallden 1998; ACC 2004).

The sensitivity of the YFSI high-risk classification was poor (less than 30%) and thus it needs to be reconsidered for use in this NZ cohort of ALBP claimants. Given the wide score range available for 'low-risk' claimants (50 points), in comparison to the available range of low to moderate risk (39 points) and moderate to high risk (15 points), it is perhaps not surprising that different risk profiles exist within each classification, particularly given the different ethnic profiles in each yellow flag score classification. The intention for the YFSI is to assist in identifying claimants to whom appropriate allocation of additional resources would be beneficial (Linton and Hallden 1998). From these results, there appears to be a considerable proportion of claimants who are being overlooked. Claimants previously deemed at low or moderate risk may still benefit from referral to other services, such as counseling. Further investigation of yellow flag score classifications with respect to ethnicity,

and increased claims cost, lengthy claims and development of chronic pain is warranted.

Conclusion

For those individuals whose ALBP threatens to take longer to resolve, there appears to be merit in embedding an assessment of psychosocial factors within a clinical pathway. This claim is supported by the generally linear association between yellow flag scores, treatment costs and claims length. However the YFSI requires testing in other ALBP populations to improve the sensitivity of its high score classifications in detecting individuals who are likely to have extended claims length (for instance greater than 90 days). Although the specificity of the YFSI was high, the high percentage of claimants who scored low on the YFSI yet consumed lengthy claims requires investigation with respect to yellow flag score range, and ethnicity.

Approximately 60% ALBP presentations in this sample resolved within five GP visits (as anticipated by the ACC clinical pathway). This describes the natural progression of, and expected compensable costs related to, ALBP. The standard ALBP protocol of applying the YFSI at the second GP visit needs to be revisited in light of the number of claimants who did not comply. These potentially included those who ceased care after the second or third GP visit, those claimants whose psychosocial risks were assessed as high, but did not consume lengthy episodes of care, and those claimants whose risks were assessed as low, and who had long claims.

Acknowledgments

References

[ACC] Accident Compensation Corporation. 2004. New Zealand Acute Low Back Pain Guide [online]. Accessed July 5, 2008. URL: http://www.acc.co.nz/wcm001/groups/external_ip/documents/internet/wcm002131.pdf.

[ACC] Accident Compensation Corporation. 2007. Annual Report [online]. Accessed January 10, 2008. URL: http://www.acc.co.nz/PRD_EXT_CSMP/groups/external_communications/documents/internet/dis_ctrb097219.pdf.

Barosi G. 2006. Strategies for dissemination and implementation of guidelines. *Neurol Sci*, 27(S3):s231–s234.

Crawford C, Ryan K, Shipton E. 2007. Exploring general practitioner identification and management of psychosocial Yellow Flags in acute low back pain. *N Z Med J*, 120:U2536.

Grimmer K, Bowman P, Roper J. 2000. Episodes of allied health outpatient care: an investigation of service delivery in acute public hospital settings. *Disab Rehab*, 22:80–7.

Grimshaw JM, Thomas RE, MacLennan G, et al. 2004. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess,* 8:iii–iv, 1–72.

Kendall NA. 1999. Psychosocial approaches to the prevention of chronic pain: the low back paradigm. *Best Pract Res Clin Rheumatol*, 13:545–54.

Kendall NAS, Linton SJ, Main CJ. 1997. Guide to assessing psychosocial yellow flags in acute low back pain: Risk factors for long-term disability and work loss. Wellington, NZ: ACC.

Lenfant C. 2003. Shattuck lecture: clinical research to clinical practice – lost in translation? *N Engl J Med*, 349:868–74.

Linton SJ, Hallden K. 1998. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain*, 14:209–15.

Linton S J, Boersma, K MA. 2003. Early identification of patients at risk of developing a persistent back problem: The predictive validity of the Orebro Musculoskeletal Pain Questionnaire. *Clin J Pain*, 19:80–6.

[NHMRC] National Health and Medical Research Council. 2003. Evidence-based management of acute musculoskeletal pain [online]. Accessed September 1, 2007. URL: http://www.nhmrc.gov.au/publications/synopses/_files/cp94.pdf.

Newton-John T, Ashmore J, McDowell M. 2001. Early intervention in acute back pain: problems with flying the yellow flag. *Physiotherapy*, 87:397–401.

Turk DC. 1997. The role of demographic and psychosocial factors in transition from acute to chronic pain. In: Jensen TS, Turner JA, Wiesenfeld-Hallin Z (eds). Proceedings of the Eighth World Congress on Pain. Seattle: IASP Press.

Walker BF. 2000. The prevalence of low back pain: a systematic review of the literature from 1966 to 1988. *J Spin Disord*, 13:205–17.