ABSTRACT

Whiplash injuries account for the vast majority of casualties in road traffic crashes, leading to long-term consequences. The majority occur in rear-ends crashes. Consumer crash tests play an important role in promoting effective concepts to reduce the problem. The current Euro NCAP whiplash test protocol includes three sled tests at varying impact speeds and pulse shapes using a BioRID test dummy and 8 measures to assess whiplash potential based on previous best practice. Given the complexity of the test and with more experience, a real-world evaluation of the current protocol was undertaken. Three analyses were undertaken comprising an analysis of test outcome data, a logistic regression analysis, and a correlation analysis comparing crash and injury outcome. 13,389 drivers reporting whiplash injury symptoms to Folksam Insurance in Sweden were studied, of which 1,266 occurred in cars tested by Euro NCAP. For all occupants reporting initial symptoms, the risk of permanent medical impairment was followed up according to the procedure used by Swedish insurance companies. Test scores according to Euro NCAP, JNCAP and IIWPG protocols were calculated, as well as combinations of the three Euro NCAP pulses. For each combination or protocol, the test score was compared with the real-world outcome. A correlation analysis of the included injury criteria was also performed for the three crash pulses included.

The results showed that overall Euro NCAP, JNCAP and IIWPG all predict real-world whiplash injury outcome in terms of Permanent Medical Impairment (PMI). Based on limited data available, there was no statistical evidence using logistic regression and ROC analyses that any of the three tests performed better than any other. Correlations between the test scenarios of each of the three protocols, as well as the outcome associations with crash outcomes, suggested consistent improvements in the risk of permanent medical impairment. The main strength of the analyses conducted here was to show the validity of Euro NCAP, JNCAP and IIWPG whiplash test protocols when measured against real-world crash outcomes, which are the most important criteria showing that the tests are appropriately designed to help prevent injuries among the community. Some caution needs to be taken with these findings as many were not statistically significant because of the limited number of cases available. Further evaluation when additional data are available is warranted.
Since the late 1990s more advanced concepts aimed at reducing the risk of WAD have been introduced on the market (Jakobsson 1998, Wiklund and Larsson 1997, Lundell et al. 1998, Sekizuka 1998). Better protection is achieved through improved geometry and dynamic properties of the head restraint or by active devices that move in a crash as the body loads the seat. The main ways to lower the risk of WAD in rear impacts are to minimize the relative motion between head and torso, to control energy transfer between the seat and the human body and to absorb energy in the seatback. Studies have been presenting showing the effect of the some seat concepts indicating a reducing effect in WAD of approximately 20-50% (Viano and Olsén 2001, Farmer et al. 2008, Jakobsson 2004, Jakobsson et al. 2008, Krafft et al. 2003, Kullgren et al. 2013).

In 2004 consumer test programmes were introduced (IIWPG and Folksam/SRA) (Thatcham 2013, Krafft et al. 2004). And since 2009 the European New Car Assessment Program (Euro NCAP) has introduced a test protocol for assessing the Whiplash potential of vehicle seats as part of its vehicle safety rating system in Europe. The protocol specifies 3-sled tests at varying impacts using a BioRID test dummy and 8-measures to assess whiplash potential based on previous best practice. To date, more than 200 make and model vehicles seats have been evaluated using the Euro NCAP protocol. JNCAP first conducted similar assessments of vehicles in Japan using other injury criteria and pulses around 2010 (JNCAP, 2014).

Given the complexity of the test and with greater knowledge, Euro NCAP decided to evaluate the current protocol to see if it was still appropriate and whether it could be improved and simplified without reducing its effectiveness. The objectives of the study was to evaluate the effectiveness of the current Euro NCAP whiplash assessment procedure and current assessment criteria to determining the real-world injury outcome and whether reduced tests and output measures would equally predict real-world performance.

METHOD

Three analyses were undertaken comprising an analysis of Euro NCAP test outcome data, a logistic regression analysis, a ROC analysis, and a correlation analysis of whiplash injuries reported to Folksam Insurance, comparing crash test results and real-world injury outcome in crashes with car models tested by Euro NCAP.

Receiver Operating Characteristic curves (ROC) are a plot showing the comparative performance of systems across a varying threshold. Plotting the sensitivity and specificity of varying outcomes, in terms of true and false positives, the cumulative distribution function can then be found for each of the comparative relationships (the true positive rate against the false positive rate for the different possible cut-points of a diagnostic test). The area under the curve can then be used as a measure of the performance of the systems under test. Should one of the systems under examination score a larger area under the curve, it can be assumed to out-perform the others if the difference is statistically significant. ROC analysis is commonly used in association with a logistic regression analysis.

The Folksam material used in the correlation analysis consisted of 13,389 drivers (at least 18 years old) reporting symptoms of WAD to Folksam Insurance in Sweden between 1998 and 2013, of which 1,266 (55% females) occurred in cars tested by Euro NCAP (see Table 1). For all occupants reporting initial symptoms, the risk of permanent medical impairment was followed up according to a procedure used by Swedish insurance companies. In case an injured occupant is not recovered after approximately one year, a medical assessment is made by medical specialists to predict the impairment degree. The injured occupant is classified with a degree of medical impairment, between 1-100% depending on the injury type according to the Swedish manual for “Grading Medical Impairment” (Försäkringsförbundet 1996). All Swedish insurance companies use this manual. The manual consists of instructions of how to set the degree of medical impairment and table works for all injury types and their consequences. WAD often results in between 3-18% of medical impairment degree. The symptoms are regarded as permanent when no additional improvement in the injured patient’s mental or physical status has taken place, normally a maximum of three years after the crash.

It has been found that medical expertise in Sweden gradually has been classifying long-term consequences from whiplash associated disorders more restrictively (Kullgren et al. 2013). Therefore adjustments were made by weighting the number of occupants with long-term symptoms according to the year of impact. A reduction factor of 11% per accident year was used. In order not to change the total number of occupants with long-term symptoms the weighting was made based on accident year 2006, which is the mean accident year in the accident sample. All occupants with long-term symptoms in crashes occurring before 2006 were weighted lower and all after 2006 were weighted higher (Equation 1). By making an adjustment for accident year for each driver, the outcomes from all groups of cars under study could be compared with each other. In total 52 drivers sustained a permanent medical impairment and two out of three were females. The numbers are presented in Table 1.

Kullgren et al. 2
\[ x_{\text{pmi, adjusted}} = \frac{x_{\text{pmi}}}{1.1^{(2006 - \text{year accident})}} \text{, } x_{\text{pmi}} = \text{occupants with pmi} \]  

Eq (1)

**Table 1: Number of drivers reporting whiplash symptoms and those with pmi in cars tested by Euro NCAP.**

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>n pmi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>547</td>
<td>17</td>
</tr>
<tr>
<td>Female</td>
<td>677</td>
<td>35</td>
</tr>
<tr>
<td>Unknown</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1266</strong></td>
<td><strong>52</strong></td>
</tr>
</tbody>
</table>

Test scores according to Euro NCAP, JNCAP and IIWPG protocols were calculated based on the test performed by Euro NCAP. To increase the data available test results were also calculated based on non-official Euro NCAP test data from a test series performed before 2009. Also combinations of the three Euro NCAP pulses were calculated. For each combination or protocol, the test score was compared with the real-world outcome. The following comparisons were studied:

- Low, mid and high severity tests in Euro NCAP
- Combinations of tests (pulses) in Euro NCAP
- Individual injury criteria in each test
- Combinations of reduced criteria and reduced number of pulses

Test score expressed as per cent of maximum score versus proportion of occupants with pmi was plotted for the various combinations. For the injury criteria the absolute measured value was plotted against proportion of occupants with pmi. In each plot a linear line fit was added. The line fits were weighted for number of crashes in each point. Each point represent an average outcome in an interval. R-square values showing how well the line fits the points in each plot are calculated for each plot and p-values and 95% CI were added to each line fit.

**RESULTS**

The results are presented in two forms. First, a statistical analysis was performed using logistic regression analysis and Receiver Operating Characteristics (ROC) curves. This was intended to provide a rigorous statistical analysis of the overall benefit of the three existing test protocols (Euro NCAP, IIWPG, and J-NCAP) as well as show if there were differences (improvements) between the individual findings. Second, a correlational analysis was also conducted both overall as well as on components of the test protocols to indicate opportunities for future improvement.

**Regression and ROC Analysis**

An analysis was undertaken of the relationship between sensitivity and specificity of variations of the three test protocols that were plotted on ROC Curves. From earlier work, it was apparent that the sex and age of the victim had a substantial impact on the results in terms of risk and long term outcome, hence a regression analysis was also undertaken to control for these characteristics. The three test protocol combinations shown below were:

1. Euro NCAP comprising the median crash test (16km/h, 5.5g, with triangular pulse) and the seven test criteria (NIC, Nkm, Rebound velocity, Fx, Fz, T1 and THRC);
2. IIWPG with quantified scores, comprising a single crash pulse of 16km/h (5.5g) with triangular pulse, and four test criteria namely Time to head restraint (\( \leq 70 \) msec), T1 acceleration (\( \leq 9.5g \)), Fx and Fz; and
3. J-NCAP comprising a single triangular crash test pulse of 17.6km/h, and 7-injury criteria including NIC, Upper Fx (backward shear), Upper FZ (tension), Upper My (flexion), Upper My (extension), Lower Fx (backward shear), and Lower Fz (tension), and Lower My (flexion).

These data scores were all computed from Euro NCAP test data as sufficient factors are collected to create the three test protocols listed above. The ROC analysis for the 3-protocols with accompanying statistics are shown in Figure 1 and Table 2 below.
As noted above, the IIWPG subjective scale was converted into a numerical one for this comparison, which does lose a bit of information compared to the other test scores that use actual values. Nevertheless, this was necessary for the ROC analysis. The analysis showed that any apparent differences between the areas under the curves (the classic way of comparing tests) were indistinguishable from random variation. What these results do show is that any one of these tests has reasonable and statistically significant predictive value for real-world outcomes, but there was no apparent statistical difference observed between the 3-curves. It will take a considerable increase in data before apparent differences between tests will be detectable in the ROC analysis.
Correlation Analysis

Euro NCAP was significantly correlated with the other two protocols. The match between the Euro NCAP and J-NCAP appeared to be slightly stronger (less variation) than between the others as shown in Fig. 2 to 4.

![Fig 2: Euro NCAP and JNCAP scores.](image)

![Fig 3: Euro NCAP and IIWPG scores.](image)

![Fig 4: JNCAP and IIWPG scores.](image)

Furthermore, all the test protocols were significantly correlated well with the risk for pmi as shown in Fig. 5 to 7.

![Fig 5: Risk for pmi vs Euro NCAP score.](image)

![Fig 6: Risk for pmi vs JNCAP score.](image)

![Fig 7: Risk for pmi vs IIWPG score.](image)

Risk of Permanent Medical Impairment by Euro NCAP Crash Pulse

Fig’s 8 to 10 show the association between the risk of impairment (pmi) and the test pulse for the 3-Euro NCAP tests. All three pulses in Euro NCAP correlate with risk for pmi, but the correlation for the low and mid severity pulse was stronger than for the high severity pulse.

![Fig 8: PMI risk vs Euro NCAP low pulse.](image)

![Fig 9: PMI risk vs Euro NCAP mid pulse.](image)

![Fig 10: PMI risk vs Euro NCAP high pulse.](image)

Further examination of the pmi risk between combinations of the three test pulses are shown in Fig’s 11 to 13. The correlation between pmi risk for low and mid pulses and that for low and high pulses were significant, although the correlation between the mid and high pulse was not at the 5% probability level (around 6%).

Kullgren et al. 5
Injury Criteria

Examples of results of the correlation analyses undertaken between the risk of pmi and the test criteria are shown in Fig’s 14 to 19 (plots for all criteria in all three pulses can be seen in the Appendix). The criteria that show correlation (p<0.05) to risk of pmi were NIC in all three pulses (p just above 0.05 in the low pulse), Fz in the high pulse and rebound velocity in the low pulse. Also the full geometry assessment explained risk for pmi while geometry did not. Fx had some outliers in the low and mid pulses, an example can be seen in Fig 17, where high Fx values were measured although the risk for pmi was low. If excluding these outliers a correlation could be verified for these pulses but not for the high pulse.

Correlations with Modified Test Pulses and Criteria

The final set of analyses examined possibilities for fewer test pulses. And based on the correlation analysis of injury criteria an attempt to using a reduced number of criteria was made using; NIC, Fx, Fz and rebound velocity. Fig’s 20 to 22 show the correlations for reduced combinations of Euro NCAP test pulses and test criteria. These should be read in conjunction with Fig 4 shown earlier for the 3 test pulses and the full set of criteria. All these combinations were significant (Euro NCAP 2 pulses and all criteria had a p value just above 0.05). However, it could not be verified which of the combinations that best explained risk for pmi.
DISCUSSION

Efforts to reduce the number of Whiplash Associated Disorders (WAD) are essential due to the societal burden in terms of costs and suffering from impairments (Krafft 1998, Malm et al. 2008, Kullgren et al. 2013). Consumer test programmes are important tools to reduce WAD in the way they promote cars with the most effective whiplash systems. Since the first consumer test programmes were introduced by IIWPG and Folksam/SRA in 2003 and 2004 the European New Car Assessment Program (Euro NCAP) in 2009 introduced a test protocol for assessing the whiplash potential of vehicle seats as part of its vehicle safety rating system in Europe. Also Japan NCAP has introduced whiplash tests as part of the Japanese consumer test programme (JNCAP 2014).

In conducting this study, two methods were adopted in an attempt to investigate associations between the variables and the outcome criteria. The first was a rigorous parametric statistical evaluation using logistic regression and Receiver Operating Curve (ROC) analysis, while the second was a correlation analysis between aspects of the tests and whiplash outcome. While the results suggested that all protocols examined (Euro NCAP, J-NCAP and IIWPG) appeared to have a strong meaningful association with the advent of Permanent Medical Impairment (pmi), there were no statistical differences between any of the three protocols for the data available. Moreover, as noted earlier, it will require a sizable increase in the available database before this could be expected.

In conducting the correlation analysis, there were two options available for measuring real-world outcome due to the injury data available. These included (i) the likelihood of sustaining a whiplash injury leading to symptoms lasting longer than 1 month after the crash event, and/or (ii) the likelihood of sustaining on-going impairment (most often classified as a chronic outcome). Previous research by for example Davidson and Kullgren (2011) showed that while early predictions of whiplash provided greater numbers of outcomes that could be assessed, the measurement of Permanent Medical Impairment (pmi) was a more stable measure.

It was positive to find that the 3 main consumer test programmes; Euro NCAP, JNCAP and IIWPG, were equally able to predict real-world whiplash injury outcome in terms of permanent medical impairment. Previous studies have shown that both IIWPG and Folksam/SRA test programmes correlate with real-world whiplash injury outcomes (Kullgren et al. 2007, Kullgren and Krafft 2008). But this is the first time it has been shown for the Euro NCAP consumer tests. The reader needs to be careful not to infer too much from these findings though. Given the sparseness of data available so far, this can only be considered as a preliminary analysis that needs further confirmation with a much larger data set.

The current Euro NCAP whiplash test protocol includes three crash pulses of various severity and all injury criteria included in the consumer tests developed by Folksam/SRA and IIWPG in 2002 to 2003. It is a mix of 7 injury criteria and a geometrical assessment of the head rest. One of the main reasons for this is to get a robustness of the rating procedure, in the sense that the tests cover a wide range in crash severity and that many possible injury causes are covered as long as we don’t know the exact cause of the whiplash symptoms. Previous studies have shown correlations between various of the injury criteria used and real-world injury outcome (see for example Davidsson and Kullgren 2011). The results reported here showed that some of the criteria used were less able to predict with real-world whiplash injury outcome. In particular, time to head rest contact (THRC) and the acceleration of the T1 vertebrae (T1acc) were not correlated with Permanent Medical Impairment as were Nkm and Fx (although Fx was significant however for the Mid pulse, although not for the Low and High pulses). In addition, for the car models with the highest measured Fx values the risk for pmi was very low. Note that the numbers of observations for these points in the plots involved only one or two crashes.
Even though the accident material covered 16 years (1998 to 2013) of crashes reported to Folksam Insurance, the number of available cases was too low to provide statistically robust results. Nevertheless, the findings reported here for reductions in the number of tests and outcome criteria were promising, suggesting possible options for explaining the risk for pmi and as a basis for potentially developing the Euro NCAP protocol in the years ahead.

Limitations
By far the biggest limitation experienced with this analysis was in the number of suitable cases of Permanent Medical Impairment in the current Folksam database. While this could be overcome to some degree by using the outcome of whiplash injury with symptoms lasting longer than 1 month after the crash event, from previous studies, this may be offset by the degree of variation in the findings.

To increase the data available for analysis it was necessary to use non published Euro NCAP data nased on tests performed before 2009. It was not fully verified if these tests fully met all Euro NCAP protocol requeirements.

The ROC analysis is well-established as a means to compare the performance of test criteria against a gold-standard outcome measure (here, real-world injury). Substantially more data would be required before the tests studied can be distinguished using these criteria.

SUMMARY AND CONCLUSIONS
The analyses conducted here were aimed at identifying real-world whiplash associations with existing test protocols used by Euro NCAP, J-NCAP, and IIWPG. Given the paucity in the data available, the finding here should be regarded as preliminary findings at this stage. Of interest, Euro NCAP, JNCAP and IIWPG were all found to be significantly correlated with each other and correlated to some degree with the risk for WAD leading to permanent medical impairment. There were signs that that there could be refinements in the number of test criteria. There was a suggestion that reductions in both the number of tests and criteria could still provide significant associations with Permanent Medical Impairment, but that further research is warranted to further test its robustness.

REFERENCES


Thatcham (2013) www.thatcham.org


APPENDIX

Fig A1. Risk for pmi vs low+mid pulses.

Fig A2. Risk for pmi vs low+high pulses.

Fig A3. Risk for pmi vs mid+high pulses.

Fig A4. Risk for pmi vs NIC low pulse.

Fig A5. Risk for pmi vs Nkm low pulse.

Fig A6. Risk for pmi vs reb vel low pulse.

Fig A7. Risk for pmi vs Fx low pulse.

Fig A8. Risk for pmi vs Fz low pulse.

Fig A9. Risk for pmi vs T1acc low pulse.

Fig A10. Risk for pmi vs THRC low pulse.

Fig A11. Risk for pmi vs NIC mid pulse.

Fig A12. Risk for pmi vs Nkm mid pulse.

Fig A13. Risk for pmi vs reb vel mid pulse.

Fig A14. Risk for pmi vs Fz mid pulse.
Fig A15. Risk for pmi vs T1 acc mid pulse.  
Fig A16. Risk for pmi vs THRC mid pulse.  
Fig A17. Risk for pmi vs NIC high pulse.  
Fig A18. Risk for pmi vs Nkm high pulse.  
Fig A19. Risk for pmi vs reb vel high pulse.  
Fig A20. Risk for pmi vs Fx high pulse.  
Fig A21. Risk for pmi vs Fz high pulse.  
Fig A22. Risk for pmi vs T1acc high pulse.  
Fig A23. Risk for pmi vs THRC (high pulse).  
Fig A24. Ris for pmi vs geometry.  
Fig A25. Risk for pmi vs full geometry.  
Fig A26. Risk for pmi vs Euro NCAP 2 pulses (mid and high) and all criteria.  
Fig A27. Risk for pmi vs all Euro NCAP pulse and 4 criteria; NIC, Fx, Fz and reb vel.  
Fig A28. Risk for pmi vs Euro NCAP 2 pulses mid and high and 4 criteria; NIC, Fx, Fz and reb vel.