Continuous measures of driving performance on an advanced office based driving simulator can be used to predict simulator task failure in patients with obstructive sleep apnoea syndrome.

ONLINE DATA SUPPLEMENT

METHODS

The study was conducted in the Department of Respiratory Medicine at St. James' University Hospital, Leeds, UK. The software and hardware support for the miniSim was provided by the Institute for Transport Studies, University of Leeds, UK. Ethical approval was obtained from the local NHS Research Ethics Committee.

Patients

Patients attending the Sleep Clinic at St. James's University Hospital with a confirmed diagnosis of OSA [Apnoea Hypopnoea Index (AHI) and/or Oxygen Desaturation Index (ODI) and/or 4% dips in saturation >10/ hour] on respiratory variable overnight sleep study (Embletta®) or overnight oximetry were approached. Recruitment was biased towards patients who were considered for a trial of CPAP therapy. This was to generate a patient population who might be at a higher risk of RTA and have a number of "events" on the simulator. Included patients had their demographic (age, BMI etc), clinical (ESS) and polysomnographic characteristics (AHI and ODI) recorded. Patients with other causes of sleepiness; eg shift workers, patients on sedative medications, etc were excluded.

Driving Simulator (miniSim)

All patients were asked not to drink any caffeinated drinks within two hours of the start of the study.

Road layout and scenario:

A 90km three-lane motorway was developed with UK standard lane markings and signage. The road is comprised of 8 junctions, including entry and exit slip roads, separated by 8 sections of road (each 9km in length). One junction and one section of motorway together take approximately 7 minutes to drive (at 70mph) and will be referred to as one epoch. Thus the entire road comprises eight epochs. All subjects had the procedures explained and had a 4 epoch practice session (20 - 25 minutes) before commencing the test proper. In the main experimental driving session a "minor" or "veeer" event was choreographed within epoch 4. This entailed a scenario whereby a vehicle swerves briefly into the driver's lane just ahead of them. This requires an avoidance manoeuvre such as braking or swerving (or both), but the vehicle is sufficiently far ahead that it was anticipated that an alert, competent driver should easily be able to avoid a collision. It should be noted that all through the drive vehicles manoeuvre in and out of the subjects' lane and it is expected that drivers would react to them as they would in real life. The "minor" event is an extension of these manoeuvres. A "major" or "brake" event was inserted into epoch 8 and this also signalled the end of the run. Here, a vehicle ahead brakes heavily, requiring the driver to be fully attentive and reactive in order to avoid a collision. However, even with full attention some subjects might not be able to avoid a crash. The scenario was coordinated such that all drivers would be at the same time-to-collision when the car ahead starts to brake; thus all drivers are faced with a comparable task. A review of the literature and experience from other studies using simulators suggests that "minor" events may precede “major” events [1]. All subjects were instructed to drive in the middle lane and were asked not to change lanes to overtake the vehicle in front but to try to keep up with it. This generates comparable and consistent data.

Measures and end points:

Task failure was defined as: hitting another vehicle, veering completely out of lane (except in response to an event) or spending more than 5% of the total study time (2 ½ minutes) with two wheels out of the middle lane. There were four possible outcomes of the simulator runs; (i) task failure unprovoked during the study; (ii) crashing into the vehicle in front at the veer event (iii) crashing into the vehicle in front only at the brake event; (iv) no task failure at any time during the study run. Unprovoked task failure and crashes at the minor event should not happen during normal simulated driving and any subject falling into this category was considered to have “failed“ the simulator test. Subjects who completed the test without meeting any of the task failure criteria defined above were deemed to have “passed”. The major event was choreographed such that it was harder to avoid a crash and those who only crashed at this event were deemed to be “indeterminate”.

At 60Hz, various parameters of driving behaviour were recorded. These included continuous measures of driving behaviour such as: minimum time headway (Hw), percentage time spent with minimum headway of less than 1 second (Hw1s), minimum time-to-collision (TTC) to the preceding vehicle, high frequency steering (HFS), mean speed and speed variation, standard deviation of lane position (SDLP), lane changes. For the purpose of analysis we used the mean values for each parameter in epochs 3,5,6 and 7, which were free of events and just require steady driving at approximately 70 mph. In addition, specific measures at the programmed events were also recorded, including: speed on approach to collision and reaction times (RT).

Definitions of various driving simulator parameters

Time to Collision (TTC) [seconds] is defined as the instantaneous time it would take to collide into the lead vehicle if vehicle speeds are kept constant. TTC reflects risk margin; the lower the TTC, the less margin for error. Mean and median values of the TTC-minima and the number of TTC-minima less than one second have been used as indicators of risk of collision; the lower the value of TTC-minima, the higher the risk.[2]

Time Headway (Hw) [seconds] to lead vehicle is defined as the time it would take to collide into the lead vehicle were it to stop dead. Time Headway is a measure of longitudinal risk margin. The closer and faster a subject travels behind a lead vehicle, the less the chance of managing to avoid a collision if the lead vehicle reduces their speed. For a small headway, the time a subject can be distracted by another task without a highly increased risk of accident, is much less than if the time headway is large. The proportion of the time headway less than one second (Hw1s) has been used as a risk indicator for car following situations. A higher proportion of time spent with headway less than 1 second (Hw1s) is an indicator of worse performance and dangerous driving**.** [3,4] Minimum time headway is the minimum value of headway reached in a particular epoch. Again a lower value indicates poorer driving performance.

Standard Deviation of Lateral Position (SDLP) : Less lateral control may be observed as an increase in standard deviation of lateral position (SDLP). In several studies, driver sleepiness (drugs, sleep deprivation) has been shown to cause an increase in SDLP; the steering control has become less stable. However, SDLP is influenced by overtaking and voluntary changes in lateral position due to road curvature; effects that may not be related to driving performance. Hence in this study subjects were asked to stay in the middle lane all through the runs and we took into account the SDLP only from the straight sections of the road. Higher SDLP relates to worse vehicle control. [5,6] It is measured in meters.

High frequency steering activity (HFS) : The high frequency component of steering activity is measured as a ratio between steering movements of 3 - 6 hz to all other steering activity. Higher HFS indicates poorer control. [7,8]

Reaction time (RT) [Seconds] : Time between the lead vehicle commencing veering or braking manoeuvre and participant commencing braking. If the patients failed to brake the reaction time was infinity and if they veered out of lane to avoid crash no RT was recorded. In the case of the former a RT of > 2.5 sec was assigned to prevent these patients from being excluded from the analysis. All but one subject with a recorded RT had RT >2.2 sec. Recent studies of perception-reaction times have shown 85th percentile of 1.9 seconds and 2.5 seconds for the 95th percentile time; hence we chose 2.5 secs as the maximum cut off. [9]

Study Design and analysis

The study was divided into two phases. In the first phase we explored whether any of the continuous and event specific variables recorded during simulator runs could predict the outcomes on the MiniSim. We compared various continuously recorded (eg. SDLP, HFS, TTC, Hw1s) and event specific (eg. RT) measures of driving performance between different categories of patients using one way ANOVA and t tests with Bonferroni's multiple testing correction.

Binary logistic regression analysis was performed to test the hypothesis that a "fail" could be predicted from continuous measures of driving behaviour and thereby explore the possibility of developing predictive model/s. The regression analysis for the exploratory study was done in two stages. In the first stage the outcome or dependent variables were "Fail" and "Other than fail" ie. "indeterminate" and "pass". The independent variables included in analysis were the simulator parameters which showed significant differences in univariate analysis (ANOVA), the rest were excluded. All the simulator parameters were continuous variables. The binary logistic regression analysis method used was backward stepwise (conditional). A similar analysis was carried out in the second stage where the dependent variables were "Pass" and "indeterminate". During model building the criterion for selecting a predictor was $p<0.05$ and for rejection was $p>0.1$.

Receiver Operative Characteristic (ROC) curve analyses was performed to identify optimal cut offs for sensitivity, specificity and predictive powers of the models. The analysis was carried out with outcomes being "Fail" and "Other than fail" against the predicted probability scores generated from the regression analysis. In order to compare the discriminatory powers of the models, the area under the curves were compared using methods described by DeLong et al.[10] It was hypothesized that the choice of cut offs and thereby sensitivity and specificity of the models could be influenced by various factors, eg. by one's attitude towards risk, drivers' individual situation or economic impact of identifying too many false positives versus too many false negatives. On one end it might be the view taken by the driving regulatory authority that anybody with the slightest possibility of failing this test should be considered to have failed ie. high sensitivity to identify “fails” but with relatively higher number of false positives. On the other hand the drivers’ perspective would be to identify only the ones who are definitely failing ie. very high specificity with loss of sensitivity. There also could be compromise between the sensitivity and specificity which we chose to report in the main paper. So we calculated the predictive values of the models at three cut offs, two extremes and the compromise.

In the second phase we validated the findings from the Exploratory Study in a different population. The simulator parameters obtained from the validation cohort was applied to the regression equations derived in the exploratory phase. The probability scores thus generated were used to create new ROC curves. The areas under the curves of the two studies were compared using MedCalc ® statistical software. The sensitivities and specificities of the models on the validation cohort at cut off values chosen from the exploratory models were calculated. These values were compared with the sensitivity and specificity of the exploratory cohort using two sample Z test for comparing proportions.

The rest of the statistical analyses was carried out using SPSS 17 ® and Prism Graphpad 5 ® statistical software packages.

RESULTS

Exploratory Study

The regression equation for Model 1 is as follows :

Probability of “fail” = A / (1+A), where A = exp (SDLP3 * 12.087 – 7.566)

The regression equation for Model 2 is as follows :

Probability of “fail” = A / (1+A), where A = exp (SDLP3 * 10.705 + VeerRT * 2.406 – 11.42)

Table 2 show the variables included in the equation for Model 2.

The results of the regression analysis did not change significantly regardless of which epoch was used.

Effect of different cut offs

The tables (Tables 3, 4, 5, 6) below show the effect of three different cut offs (two extremes and one compromise) of Models 1 and 2 chosen from ROC curve analysis when applied to the exploratory and validation cohort in differentiating “fails” from “others” (mentioned in the table as “pass” for simplicity.

	True Fail	True Pass	Sensitivity	Specificity	PPV	NPV	LR
Predicted Fail	12	15	92.3 (64-100)	74.14 (61-85)	44.44 (25-65)	97.73 (88-100)	3.56
Predicted Pass	1	43	**Cut off 0.1**				
Predicted Fail	10	9	76.9 (46-94)	84.5 (72-93)	52.6 (29-75)	94.2 (84-99)	4.95
Predicted Pass	3	49	**Cut off 0.15**				
Predicted Fail	8	2	61.5 (32-86)	96.6 (88-99)	80 (44-97)	91.8 (82-97)	17.85
Predicted Pass	5	56	**Cut off 0.527**				

Table 3: Effect of 3 different cut offs of Model 1 on the exploratory cohort
(Figures in brackets are 95% CI, PPV = positive predictive value, NPV = Negative predictive value)

	True Fail	True Pass	Sensitivity	Specificity	PPV	NPV	LR
Predicted Fail	20	25	66.67 (47-83)	75.2 (66-83)	44.44 (30-60)	88.37 (80-94)	2.69
Predicted Pass	10	76	**Cut off 0.1**				
Predicted Fail	18	15	60 (40-77)	85.15 (77-91)	54.55 (36-72)	87.76 (80-93)	4.04
Predicted Pass	12	86	**Cut off 0.15**				
Predicted Fail	13	5	43.33 (25-62)	79.19 (58-93)	72.22 (47-90)	52.78 (35-70)	2.08
Predicted Pass	17	96	**Cut off 0.527**				

Table 4: Effect of the 3 different cut offs as Table 3 on the validation cohort
(Figures in brackets are 95% CI, PPV = positive predictive value, NPV = Negative predictive value)

	True Fail	True Pass	Sensitivity	Specificity	PPV	NPV	LR
Predicted Fail	10	12	90.9 (59-100)	77.8 (64-88)	45.45 (24-68)	97.67 (88-100)	4.09
Predicted Pass	1	42	**Cut off 0.0675**				
Predicted Fail	9	2	81.8 (48-98)	96.3 (87-99)	81.8 (48-98)	96.3 (87-99)	22.09
Predicted Pass	2	52	**Cut off 0.3**				
Predicted Fail	8	1	72.7 (39-94)	98.15 (90-100)	88.89 (52-100)	94.64 (85-99)	39.27
Predicted Pass	3	53	**Cut off 0.5**				

Table 5: Effect of 3 different cut offs of Model 2 on the exploratory cohort
(Figures in brackets are 95% CI, PPV = positive predictive value, NPV = Negative predictive value)

	True Fail	True Pass	Sensitivity	Specificity	PPV	NPV	LR
Predicted Fail	17	23	85.00 (62-97)	75.79 (66-84)	42.5 (27-59)	96.00 (89-99)	3.51
Predicted Pass	3	72	**Cut off 0.0675**				
Predicted Fail	12	7	60 (36-81)	92.6 (85-97)	63.16 (38-84)	91.67 (84-96)	8.14
Predicted Pass	8	88	**Cut off 0.3**				
Predicted Fail	6	3	30 (12-54)	96.8 (91-99)	66.6 (30-92)	86.79 (80-93)	9.5
Predicted Pass	14	92	**Cut off 0.5**				

Table 6: Effect of the 3 different cut offs as Table 5 on the validation cohort
(Figures in brackets are 95% CI, PPV = positive predictive value, NPV = Negative predictive value)

DISCUSSION

The sensitivities, specificities and the predictive values of the models can be calculated for different cut offs; the one chosen will depend upon the attitude to risk. At one extreme road safety organisations might purport that all accidents should be prevented and anybody with the slightest possibility of having an accident should be identified (i.e. high sensitivity to identify "fails" but with a high number of false positives). At the other extreme, where the primary consideration is the individual whose livelihood depends on driving, the emphasis will be on being as sure as possible that an individual is an unsafe driver (i.e. very high specificity with loss of sensitivity). Furthermore the economic impact of identifying too many false positives versus too many false negatives will need to be taken into account. We have quoted compromise values for cut offs for which we have given equal weight to sensitivity and specificity; others might choose a different value.

Even though we have explored hard endpoints for our study there must be scope for the clinician to make decisions on a case by case basis. Subjects who visibly struggle to stay awake during the test, utilize various coping strategies to stay awake (eg. one subject sang, whistled and thumped the desk throughout the test), should not necessarily be deemed to have passed. Alternatively there are subjects who show good lane control throughout the study but happen to crash due to a momentary lapse might still be considered to have passed. This second group are the subjects who could not be identified correctly by the regression analysis. The "indeterminate" group warrant further study; while it can be argued that failure at the brake event alone does not necessarily indicate unsafe driving, their performance across a wide range of measures was clearly worse than those who passed. The test must be used as part of an overall clinical assessment and cannot stand alone.

With any test, an individual may perform differently under test conditions than under normal everyday conditions. This is likely to be particularly true of a test where the result has a bearing on their licence. This is difficult to overcome but is less likely when the parameters of interest are those of which an individual is unaware and can be measured unobtrusively and continuously.

Many questions still need to be answered before the MiniSim, or similar advanced simulators, can be used to help advise patients with OSAS about driving but it does hold promise. Further studies are in progress. It has significant advantages over previously described simulators in terms of realism and results that are credible in terms of their relationship to normal driving. The fact that events can be reliably predicted from parameters of which the patient is unaware is an advantage. An objective test is an advance over the current situation in which inconsistent advice is given, based upon unreliable data, coloured by the clinician's individual stance on driving and accident risk. The work described here is the first step towards the development of an objective test that could have a major impact on the reliability of the advice provided.

REFERENCE

1 Turkington PM, Sircar M, Allgar V, Elliott MW. Relationship between obstructive sleep apnoea, driving simulator performance, and risk of road traffic accidents. *Thorax* 2001 Oct;56(10):800-5.

2 Horst R, Hogema J. Time-to-collision and collision avoidance systems. 1993; Salzburg 1993.

3 Evans L, Wasielewski P. Do accident-involved drivers exhibit riskier everyday driving behaviour? *Accident Analysis and Prevention* 1982;14(1):57-64.

4 Evans L, Wasielewski P. Risky driving related to driver and vehicle characteristics. *Accident Analysis and Prevention* 1983;15(2):121-36.

5 Hack MA, Choi SJ, Vijayapalan P, Davies RJ, Stradling JR. Comparison of the effects of sleep deprivation, alcohol and obstructive sleep apnoea (OSA) on simulated steering performance. *Respir Med* 2001 Jul;95(7):594-601.

6 Juniper M, Hack MA, George CF, Davies RJ, Stradling JR. Steering simulation performance in patients with obstructive sleep apnoea and matched control subjects. *Eur Respir J* 2000 Mar;15(3):590-5.

7 Jamson H, Merat N. Can low cost road engineering measures combat driver fatigue A driving simulator investigation. Montana,USA 2009.

8 MacDonald WA, Hoffman ER. Review of Relationships Between Steering Wheel Reversal Rate and Driving Task Demand. *Human Factors* 1980 Dec 1;6:733-9.

9 The Kiewit Center for Infrastructure and Transportation Oregon State University. Stopping sight distance and decision sight distance. Oregon State University,Corvallis, Oregon; 2004.

10 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988 Sep;44(3):837-45.