ORIGINAL ARTICLE

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

Dipansu Ghosh,[1] Samantha L Jamson,[2] Paul D Baxter,[3] Mark W Elliott[1]

► Additional materials are published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2011-200699).

[1]Department of Respiratory Medicine, St James' University Hospital, Leeds, UK
[2]Safety and Technology Group, Institute for Transport Studies, University of Leeds, Leeds, UK
[3]Division of Biostatistics, LIGHT, Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, UK

Correspondence to
Dr Mark W Elliott, Consultant Respiratory Physician, Department of Respiratory Medicine, Sleep and Non-invasive Ventilation Services, St James' University Hospital, Beckett Street, Leeds LS9 7TF, UK;
mark.elliott@leedsth.nhs.uk

Received 28 June 2011
Accepted 4 April 2012
Published Online First
5 May 2012

ABSTRACT

Introduction Some patients with obstructive sleep apnoea syndrome are at higher risk of being involved in road traffic accidents. It has not been possible to identify this group from clinical and polysomnographic information or using simple simulators. We explore the possibility of identifying this group from variables generated in an advanced PC-based driving simulator.

Methods All patients performed a 90 km motorway driving simulation. Two events were programmed to trigger evasive actions, one subtle and an alert driver should not crash, while for the other, even a fully alert driver might crash. Simulator parameters including standard deviation of lane position (SDLP) and reaction times at the veer event (VeerRT) were recorded. There were three possible outcomes: 'fail', 'indeterminate' and 'pass'. An exploratory study identified the simulator parameters predicting a 'fail' by regression analysis and this was then validated prospectively.

Results 72 patients were included in the exploratory phase and 133 patients in the validation phase. 65 (32%) patients completed the run without any incidents, 45 (22%) failed, 95 (46%) were indeterminate. Prediction models using SDLP and VeerRT could predict 'fails' with a sensitivity of 82% and specificity of 96%. The models were subsequently confirmed in the validation phase.

Conclusions Using continuously measured variables it has been possible to identify, with a high degree of accuracy, a subset of patients with obstructive sleep apnoea syndrome who fail a simulated driving test. This has the potential to identify at-risk drivers and improve the reliability of a clinician's decision-making.

Key messages

What is the key question?

- Is it possible to identify a group of patients with obstructive sleep apnoea syndrome (OSAS) who are potentially at risk of being involved in road traffic accidents using an advanced office-based driving simulator?

What is the bottom line?

- It is possible to identify with a high degree of accuracy a subset of patients with OSAS who are likely to fail a simulated driving scenario, in a way that has 'street credibility' using continuously recorded data from the simulator.

Why read on?

- Using subtle parameters of which the subjects remain unaware, this simulator provides hope for a much needed objective test to help clinicians advise patients with OSAS about driving.

INTRODUCTION

On average, patients with obstructive sleep apnoea syndrome (OSAS) are at increased risk of being involved in a road traffic accident, but not all patients with OSAS are unsafe drivers. Currently advice about an individual's fitness to drive is based on the severity of the sleep-disordered breathing and daytime sleepiness, and their account of their driving.[1–4] Although there is a trend towards increased likelihood of accidents with more severe sleep-disordered breathing, these are not sufficiently robust data on which to base decisions for an individual.[5] There are conflicting data about the relationship between perceived sleepiness and the likelihood of being involved in an accident[6 7] and between subjective and objective tests for increased daytime sleepiness.[8–10] Driving requires alertness and also complex integrated higher cortical function; patients with OSAS may have neurological damage, which may impact on driving.[11] Driving may therefore be impaired for reasons other than those just related to maintenance of alertness. The advice that a patient will receive about driving will also depend on their doctor's attitude to risk and this is likely to be inconsistent in the absence of robust objective criteria. There is therefore a need for an objective test, which can help to inform the advice that clinicians give to patients with OSAS.

Any such test should evaluate as many aspects as possible of all the functions needed for safe driving, and not just alertness. Performing studies during real driving is not feasible. PC-based driving simulators provide objective data and previous studies have shown that patients with OSAS tend to perform worse than normal subjects on driving simulators, but there is considerable overlap.[12–14] Performance on the simulator improves with continuous positive airways pressure (CPAP).[13 15–21] Most of these

studies used simple simulators with graphics which were not very realistic and the simulators have not been validated against real driving. Furthermore, subjects perform in such a way that raises questions of credibility, for example, multiple crashes and off-road events during a short run, about the relationship to real driving.[14 19] This relationship is key if simulators are to be used in advising whether an individual is fit to drive. Fully immersive simulators are close to real driving but are expensive, only available in a few research centres.

The Institute for Transport Studies at the University of Leeds, UK has developed a sophisticated fully immersive driving simulator (UoLDS). This is a full size car with complex audio-visuals providing a realistic driving experience. The 'car' moves and feels real as if it is slowing down, accelerating etc. Driving simulators will never fully replicate the real driving experience, although studies have shown that there is a good correlation at the performance level.[22] Driving simulators offer an alternative environment in which to study driving behaviour and hence inform road safety policy.

Alongside the full-scale simulator, a PC-based simulator (MiniSim) has been developed using the same software. The MiniSim, provides realistic graphics, incorporates steering and foot pedals and, like the UoLDS, allows continuous measurement of variables, which have been shown to relate to driver performance.[22 23]

Because patients may need to perform the simulator test on more than one occasion and/or may be able to 'raise their game' if they know that their licence is at risk, it would be an advantage to assess driving performance using measures of which the patient is unaware. Therefore we have evaluated whether variables that are recorded continuously and unobtrusively on the MiniSim are associated with a one-off event that is credible as being indicative of poor driving.

METHODS

The study was conducted at St James's University Hospital, Leeds, UK. Ethical approval was obtained from the local NHS Research Ethics Committee.

Patients

Patients attending the Sleep Clinic with a confirmed diagnosis of OSAS (apnoea hypopnoea index and/or oxygen desaturation index (ODI-4% dips in saturation) >10/h) on respiratory variable overnight sleep study (Embletta, Medcare Flaga, Reykjavik, Iceland) or overnight oximetry were approached. Recruitment was biased towards patients considered for a trial of CPAP therapy. This was to generate a patient population at risk of road traffic accident and likely to have 'events' on the simulator.

Table 1 Characteristics of the patient populations

Characteristics	Exploratory study (n=72)	Validation study (n=133)
Age (years), median (IQR)	54 (46–61)	52 (44–60)
Men, n (%)	65 (90)	118 (88)
BMI (kg/m^2), median (IQR)	34 (30–37)	33 (30–37)
ESS, mean (IQR)	13 (7–16)	12 (7–16)
AHI (events/h), median (IQR)	32 (20–54)	25 (13–42)
ODI (events/h), median (IQR)	35 (22–55)	24 (13–44)
Years since first driving licence, median (IQR)	34 (25–40)	30 (30–38)

AHI, apnoea hypopnoea index; BMI, body mass index; ESS, Epworth Sleepiness Score; ODI, oxygen desaturation index.

Driving simulator (MiniSim)

Road layout and scenario

A 90 km three-lane motorway was developed with UK standard lane markings and signage. The road is composed of eight sections (each 9 km in length). One section of motorway takes approximately 7 min to drive (at 70 mph) and will be referred to as one epoch. All subjects had the procedures explained and had a four-epoch practice session before commencing the test proper. A 'minor' or 'veer' event was choreographed within epoch 4 of the test run; a vehicle swerves briefly into the driver's lane, requiring an avoidance manoeuvre such as braking or swerving (or both). The vehicle is sufficiently far ahead that an alert driver should easily be able to avoid a collision. Throughout the drive, vehicles manoeuvre in and out of the subjects' lane and they 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; even with full attention some subjects might not avoid a crash.

Measures and endpoints

Task failure was defined as hitting another vehicle, veering completely out of lane (except in response to an event) or

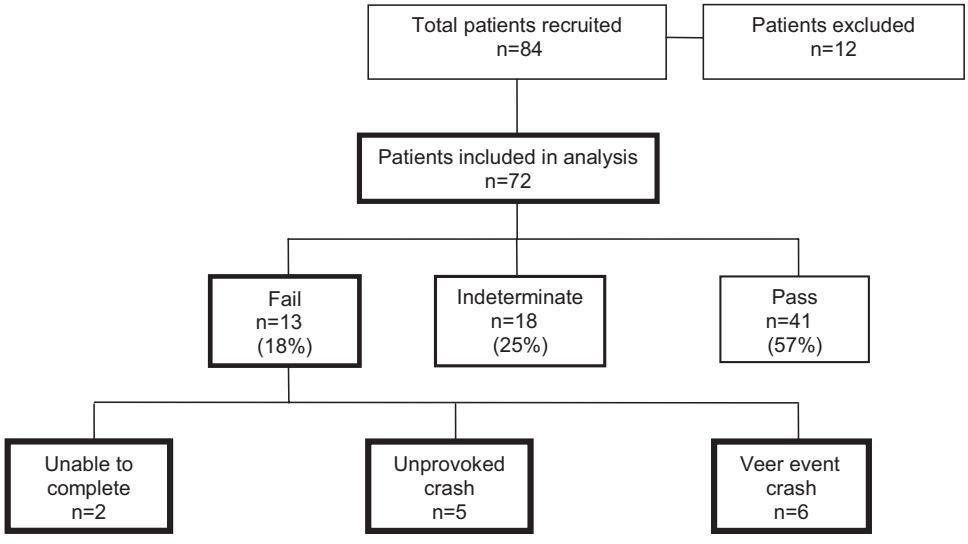

Figure 1 Driving simulator outcomes of all patients in the exploratory phase of the study.

Table 2 Comparisons between the three groups in exploratory phase: fail, indeterminate and pass

Parameters	Fail (n=13), mean (SD)	Indeterminate (n=18), mean (SD)	Pass (n=41), mean (SD)	One-way ANOVA p Value	Bonferroni's multiple testing correction: is p<0.05?		
					Fail vs indeterminate	Fail vs pass	Indeterminate vs pass
SDLP 3	0.66 (0.2)	0.44 (0.13)	0.37 (0.08)	<0.001	Yes	Yes	No
HFS 3	0.37 (0.09)	0.26 (0.06)	0.27 (0.07)	<0.001	Yes	Yes	No
TTC 3	1.97 (1.69)	5.84 (3.34)	6.67 (4.9)	0.002	Yes	Yes	No
Hw1s 3	29.6 (28.9)	17.5 (16.2)	13.5 (13)	0.296	No	No	No
Hw 3	0.21 (0.23)	0.393 (0.21)	0.57 (0.24)	<0.001	No	Yes	Yes
Speed 3	65 (8)	67 (3)	67 (3)	0.535	No	No	No
VeerRT (s)	2.3 (0.78)	1.68 (0.33)	1.38 (0.25)	<0.001	Yes	Yes	Yes
BrakeRT (s)	3.6 (0.6)	3.5 (0.9)	2.2 (0.6)	<0.001	No	Yes	Yes

BrakeRT, reaction time at the brake event; HFS 3, mean of HFS in epoch 3; Hw 3, minimum headway in epoch 3; Hw1s 3, percentage of time spent with headway under 1 s in epoch 3; SDLP 3, mean SDLP in epoch 3; Speed 3, mean speed in epoch 3; TTC 3, mean of TTC in epoch 3; VeerRT, reaction time at the veer event.

spending more than 5% of the total study time (2.5 min) with two wheels out of the middle lane. There were four possible outcomes of the simulator runs; task failure unprovoked during the study; crashing into the vehicle in front at the 'veer' event; crashing into the vehicle in front only at the major event; no task failure at any time during the study run. Unprovoked task failure and crashes at the minor event should not happen during simulated driving and these subjects were considered to have 'failed' the test. Subjects who completed the test without meeting any of the task failure criteria defined above were deemed to have 'passed'. The subjects who only crashed at the 'brake' event were deemed to be 'indeterminate'.

The MiniSim recorded continuous measures of driving behavior, such as time it would take to collide into the lead vehicle were it to stop dead (minimum time headway, Hw), percentage time spent with minimum headway of <1 s (Hw1s), minimum time to collision (TTC) to the preceding vehicle, high-frequency steering (HFS), mean speed, 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 (RTs).

Study design and analysis

The study was divided into two phases. In the first phase we explored whether any of the continuous (eg, SDLP, HFS, TTC, Hw1s) and event specific (eg, RT) simulator variables recorded could predict the outcomes on the MiniSim. We compared these measures of driving performance between different categories of patients using one-way analysis of variance and t tests with Bonferroni's multiple testing correction. Binary logistic regression analysis was used to test the hypothesis that a 'fail' could be predicted from continuous measures of driving behaviour and thereby explore the possibility of developing a predictive model. Receiver operating characteristic (ROC) curve analyses were performed to calculate the discriminative power of the models and identify optimal cutoffs for probability score. The sensitivity, specificity and predictive powers of the models were calculated using the cutoff values. The curves generated for each model were compared using methods described by DeLong *et al*.[24]

In the second phase we validated the findings from the exploratory study in a different population. We compared ROC curves and used a two-sample z-test for comparing proportions.

Detailed methodology and definitions of simulator parameters are provided in the online supplement.

RESULTS

Subject population

Two hundred and twenty-nine patients participated in the study. Eighty-four patients were recruited for the first phase and 145 for the second. Twelve patients were excluded from each of the two phases due to inability to complete the two runs (practice and test), time constraints, simulator sickness (n=4) and inability to sit continuously for 50 min. The characteristics of the 205 patients who completed the studies are described in table 1. There were no differences between the two cohorts except that the apnoea hypopnoea index in patients in the first phase was significantly higher (p=0.009).

Exploratory study

Forty-one (57%) subjects completed the simulator runs successfully. Thirty-one (43%) had some form of task failure (figure 1). Of the 13 subjects who failed, 9 also crashed at the brake event. Two subjects could not complete the full test run as they veered out of lane into the central reservation. One of them fell asleep.

Simulator variables

Comparisons between the three groups 'fail', 'indeterminate' and 'pass' are shown in table 2. There were significant differences in the ability to maintain lane position (figure 2), minimum time to collision with the vehicle in front (TTC), minimum time headway (Hw), HFS (figure 3) and RT at the veer and brake events (figure 4). These significant differences were

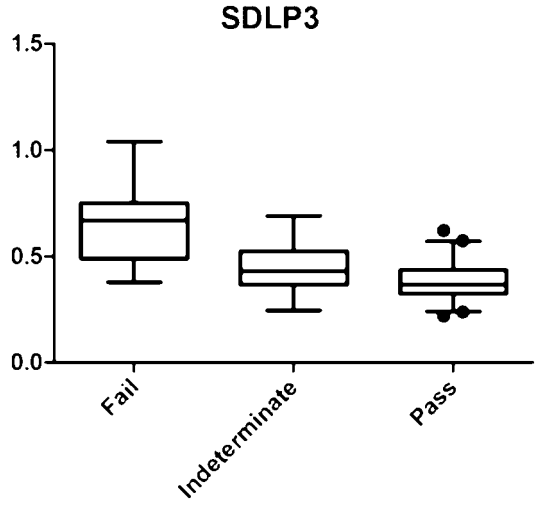

Figure 2 Distribution of mean of SD of lane position (SDLP) in epoch 3 between the three groups.

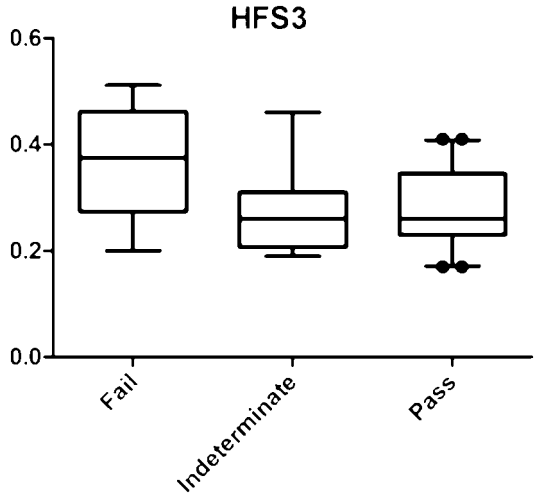

Figure 3 Distribution of mean of proportion of high-frequency steering activity (HFS) in epoch 3 between the three groups.

maintained irrespective of the epochs (3, 5, 6 or 7) used and therefore, for simplicity, ease of presentation and because this may allow the test to be shortened in future, only data from epoch 3 are presented and used in the regression analyses. There was a clear distinction between 'fails' and the rest, but only the VeerRT and Hw 3 were significantly different between the 'pass' and 'indeterminate' groups.

Regression analysis and ROC curves

Two predictive models emerged in differentiating 'fails' from the others. Model 1—including only the SDLP 3, and Model 2—includes a combination of SDLP 3 and VeerRT.

Model 2 had higher predictive power. Figure 5 compares the two models using ROC curve analysis and reinforces the finding from the regression analysis. The first model has an area under the curve (AuC) of 0.89 and the second 0.93. The difference between the two ROC curves was not significant (p=0.132). Table 3 compares the sensitivities and specificities of the two models at the chosen cutoffs. The chosen cutoffs (0.15 for model 1 and 0.3 for model 2) quoted here are a compromise between extremes of high sensitivity and high specificity, giving equal weight to both.

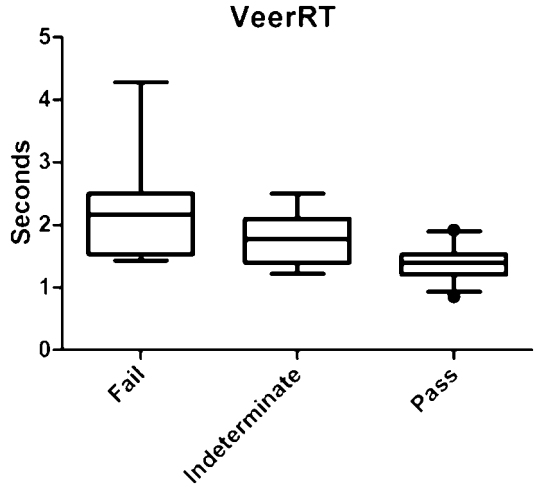

Figure 4 Distribution of mean veer reaction time (VeerRT) between the three groups.

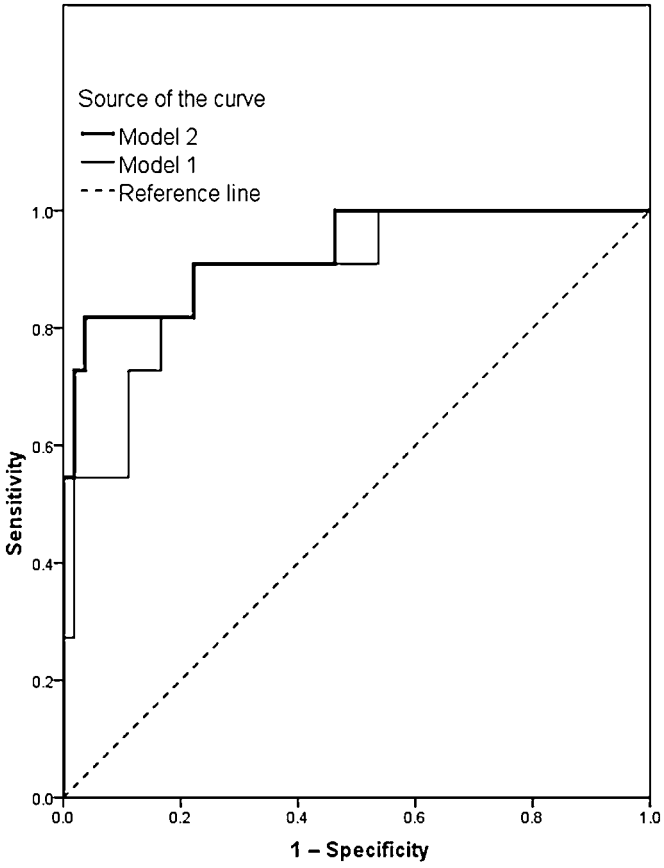

Figure 5 Receiver operating characteristic curves for the two models.

The details of the extremes of cutoffs, regression coefficients and equations are provided in the online supplement.

A similar regression analysis was performed excluding those who had 'failed' to distinguish between 'indeterminates' and 'passes'. AuC in the ROC curve analysis was 0.84.The sensitivity of that model to identify 'pass' was 97% (95% CI 86% to 99%) with 53% (95% CI 26% to 79%) specificity; VeerRT and Hw were the parameters in the equation. The positive predictive value was 84% (95% CI 70% to 93%) and the negative predictive value only 88% (95% CI 51% to 99%) using a cutoff of 0.5. Hence using simulator data the 'failed' group could be predicted with much greater certainty than the 'indeterminate' group.

Validation study

One hundred and thirty-three patients were analysed for the validation study and the outcomes are described in figure 6. Most subjects who failed fulfilled more than one task failure criterion. Of the 13 subjects who could not complete the full run, 4 also had an unprovoked crash and 4 crashed at the veer event. Nine of

Table 3 Comparing optimum sensitivities and specificities of the two models

Models	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Model 1 (cutoff 0.15)	76 (46 to 95)	84 (72 to 92)	53 (25 to 75)	94 (84 to 99)
Model 2 (cutoff 0.3)	82 (48 to 98)	96 (87 to 99)	82 (48 to 98)	96 (87 to 99)

Note: 95% CI.

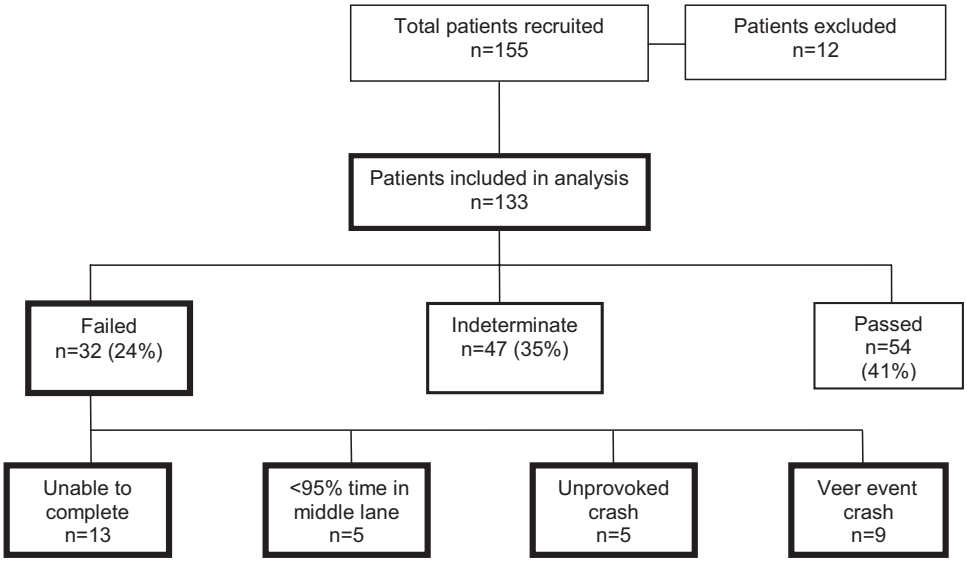

Figure 6 Simulator outcomes of the validation phase of the study.

them also failed on lane criteria. Of the 19 failed subjects who completed the full run, 12 also crashed at the brake event.

Table 4 compares the three categories. The results are similar to the first phase of the study, broadly showing the same relationship between the simulator parameters.

We could apply the equation from model 1 to 133 subjects and model 2 to 113 subjects due to unavailability of either of the two parameters; some subjects did not brake at the veer event and avoided a crash just by swerving; others crashed before epoch 3 and hence SDLP 3 was not available. A ROC curve was constructed with the predicted probability score for model 2 generated by the equation derived in the exploratory study for each subject. The AuC was 0.9 compared with 0.93 in the exploratory study. The difference between the two curves was not significant (p=0.570). Furthermore the calculated sensitivities and specificities at the chosen cutoffs were not significantly different between the two cohorts when compared using z-tests, confirming that the findings of the exploratory study are valid in a different population (table 5).

DISCUSSION

We have shown in two different cohorts that with the MiniSim it is possible to distinguish, with a high degree of confidence, patients with OSAS who unequivocally crash during simulated driving from those who are able to complete a 50 min drive without incident. We can also identify a group, but with less confidence, whose performance is intermediate. Not only was their response at the veer event different from the others but there was a clear hierarchical pattern for other parameters (SDLP, HFS, Hw and BrakeRT); subjects who 'failed' were worst, those who 'passed' the best, with the 'indeterminates' in the middle.

The sensitivities, specificities and the predictive values can be calculated for different cutoffs; the one chosen will depend upon the attitude to risk. At one extreme, all accidents should be prevented and anybody with the slightest possibility of having an accident should be identified. At the other extreme, the emphasis will be on being as sure as possible that an individual is an unsafe driver. We have quoted compromise values for cutoffs, giving equal weight to sensitivity and specificity; others might choose a different value. Examples are given in the online supplement.

We deliberately tried to recruit patients at risk of having problems driving; many completed a 50 min run on a realistic motorway without incident. This can potentially overestimate the sensitivities and specificities compared with a general OSAS cohort. However, in real life this test would probably be used on patients with OSAS who are considered to be high risk. One problem with previous simulator studies was the number of crashes or events during the study. In a study by Juniper *et al*[19] patients with OSAS had a median of 5.2 off-road events per hour. Similarly Turkington *et al*[14] reported an average of 24 off-road events per hour in their study using a divided attention driving simulator. Even patients with severe OSAS do not have multiple events during 20 min of on-road driving. The criteria that we used for 'fail' are realistic and understandable to patients. This is very important if the test is to have credibility; an individual who fails on the simulator because they go off road

Table 4 Comparisons between the three groups in validation study: fail, indeterminate and pass

Parameters	Fail (n=32), mean (SD)	Indeterminate (n=47), mean (SD)	Pass (n=54), mean (SD)	One-way ANOVA p Value	Bonferroni's multiple testing correction: is p<0.05?		
					Fail vs indeterminate	Fail vs pass	Indeterminate vs pass
SDLP 3	0.55 (0.2)	0.42 (0.1)	0.38 (0.09)	<0.001	Yes	Yes	No
HFS 3	0.35 (0.09)	0.3 (0.08)	0.26 (0.08)	0.001	Yes	Yes	No
Hw 3	0.30 (0.22)	0.48 (0.27)	0.59 (0.19)	<0.001	No	Yes	Yes
TTC 3	3.19 (3.2)	8 (18)	7.1 (7.2)	0.195	No	No	No
VeerRT (s)	2.2 (0.55)	1.57 (0.45)	1.44 (0.34)	<0.001	Yes	Yes	No
BrakeRT (s)	3.27 (1)	3.5 (0.8)	2.29 (0.9)	<0.001	No	Yes	Yes

BrakeRT, reaction time at the brake event; HFS 3, mean of HFS in epoch 3; Hw 3, minimum headway in epoch 3; SDLP 3, mean SDLP in epoch 3; TTC 3, mean of TTC in epoch 3; VeerRT, reaction time at the veer event.

Table 5 Sensitivities, specificities and predictive values of the two models at chosen cutoffs applied to the exploratory and validation cohorts and p values for a two-sample z-test comparing the sensitivities and specificities of the chosen cutoffs in the two populations

Models	Sensitivity	Specificity	Positive predictive value	Negative predictive value
Model 1 (cutoff 0.15)				
Exploratory cohort	76 (46 to 95)	84 (72 to 92)	53 (25 to 75)	94 (84 to 99)
Validation cohort	60 (41 to 77)	85 (77 to 91)	55 (36 to 72)	88 (80 to 93)
p Values for z-test	0.507	0.866		
Model 2 (cutoff 0.3)				
Exploratory cohort	82 (48 to 98)	96 (87 to 99)	82 (48 to 98)	96 (87 to 99)
Validation cohort	60 (36 to 81)	93 (85 to 97)	63 (38 to 84)	92 (84 to 96)
p Values for z-test	0.384	0.342		

Note: 95% CI.

multiple times might argue, quite reasonably, that this is not what happens when they drive a real car and therefore that the simulation is not valid.

While it might be reasonable to include an event such as our final 'brake' event, as failure to avoid this is realistic evidence of sub-optimal performance, it has the disadvantage that it may limit the usefulness of the test for repeated use. Longitudinal studies have found that behaviour adapts and changes over time, in driving simulator experiments.[25] Likewise, a patient expecting something may perform differently on subsequent occasions.[26] Furthermore, a patient may drive poorly at other times during the test but perform adequately at the event. Variables that are recorded continuously throughout, and of which the patient is unaware, are preferable. In common with previous studies we found that poor lane control (SDLP) was predictive of a crash.[18 19] Predictive power was increased by the inclusion of reaction time at the veer event. Again previous studies have shown that patients with OSAS who are untreated have worse reaction times than controls and patients with OSAS after CPAP therapy.[18 19] This is likely to be an underestimate as we had to exclude some patients from the analysis; some subjects (n=5) did not brake at all at the veer event and avoided a crash by veering out of lane, a legitimate manoeuvre; others (n=4) did not brake at all and crashed. Although this assessment requires an 'event' it was a subtle extension of routine driving behaviour and is unlikely to be memorable.

Though we have explored hard endpoints there must be scope for the clinician to make decisions on an individual basis. Subjects who visibly struggle to stay awake, utilise various coping strategies to stay awake (eg, one subject sang and thumped the desk throughout the test), should not necessarily be deemed to have passed. Alternatively, subjects who show good lane control 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.

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. 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. 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.

Acknowledgements We are grateful to Michael Daly, Tony Horrobin and Hamish Jamson from the Institute for Transport Studies for their contribution to the development and running of the simulator. We are thankful to the contributions of research nurse Craig Armstrong, sleep physiologist Sue Watts, specialist nurses Martin Latham, Jampa Choedon, Mitchell Nix, Lisa Emmett, healthcare assistant Anne Kellett and Audrey Rowe and Susan Leigh from the Sleep Services administrative team.

Contributors DG: design of study, data collection, analysis, writing the paper. SLJ: design of study, analysis, writing the paper. PDB: design, statistical analysis, writing the paper. MWE: original concept, design of study, writing the paper.

Funding The study was partly supported by an unrestricted grant from ResMed for healthcare assistant time to supervise simulator runs.

Competing interests None.

Patients consent Obtained.

Ethics approval NHS Research Ethics Committee, York.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Arand D,** Bonnet M, Hurwitz T, *et al*. The clinical use of the MSLT and MWT. *Sleep* 2005;**28**:123—44.
2. **Carskadon MA,** Dement WC. Multiple sleep latency tests during the constant routine. *Sleep* 1992;**15**:396—9.
3. **Engleman HM,** Hirst WS, Douglas NJ. Under reporting of sleepiness and driving impairment in patients with sleep apnoea/hypopnoea syndrome. *J Sleep Res* 1997;**6**:272—5.
4. **Reimer B,** D'Ambrosio LA, Coughlin JE, *et al*. Using self-reported data to assess the validity of driving simulation data. *Behav Res Methods* 2006;**38**:314—24.
5. **Ingre M,** Akerstedt T, Peters B, *et al*. Subjective sleepiness and accident risk avoiding the ecological fallacy. *J Sleep Res* 2006;**15**:142—8.
6. **Liu GF,** Han S, Liang DH, *et al*. Driver sleepiness and risk of car crashes in Shenyang, a Chinese northeastern city: population-based case-control study. *Biomed Environ Sci* 2003;**16**:219—26.
7. **Teran-Santos J,** Jimenez-Gomez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. Cooperative Group Burgos-Santander. *N Engl J Med* 1999;**340**:847—51.
8. **Bennett LS,** Stradling JR, Davies RJ. A behavioural test to assess daytime sleepiness in obstructive sleep apnoea. *J Sleep Res* 1997;**6**:142—5.
9. **Punjabi NM,** Bandeen-Roche K, Young T. Predictors of objective sleep tendency in the general population. *Sleep* 2003;**26**:678—83.
10. **Sangal RB,** Sangal JM, Belisle C. Subjective and objective indices of sleepiness (ESS and MWT) are not equally useful in patients with sleep apnea. *Clin Electroencephalogr* 1999;**30**:73—5.
11. **Morrell MJ,** Jackson ML, Twigg GL, *et al*. Changes in brain morphology in patients with obstructive sleep apnoea. *Thorax* 2010;**65**:908—14.
12. **Findley LJ,** Suratt PM, Dinges DF. Time-on-task decrements in 'steer clear' performance of patients with sleep apnea and narcolepsy. *Sleep* 1999;**22**: 804—9.
13. **George CF,** Boudreau AC, Smiley A. Simulated driving performance in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 1996;**154**:175—81.
14. **Turkington PM,** Sircar M, Allgar V, *et al*. Relationship between obstructive sleep apnoea, driving simulator performance, and risk of road traffic accidents. *Thorax* 2001;**56**:800—5.
15. **Findley LJ,** Fabrizio MJ, Knight H, *et al*. Driving simulator performance in patients with sleep apnea. *Am Rev Respir Dis* 1989;**140**:529—30.
16. **George CF.** Reduction in motor vehicle collisions following treatment of sleep apnoea with nasal CPAP. *Thorax* 2001;**56**:508—12.
17. **Hack M,** Davies RJ, Mullins R, *et al*. Randomised prospective parallel trial of therapeutic versus subtherapeutic nasal continuous positive airway pressure on simulated steering performance in patients with obstructive sleep apnoea. *Thorax* 2000;**55**:224—31.
18. **Hack MA,** Choi SJ, Vijayapalan P, *et al*. Comparison of the effects of sleep deprivation, alcohol and obstructive sleep apnoea (OSA) on simulated steering performance. *Respir Med* 2001;**95**:594—601.
19. **Juniper M,** Hack MA, George CF, *et al*. Steering simulation performance in patients with obstructive sleep apnoea and matched control subjects. *Eur Respir J* 2000;**15**:590—5.

20. **Orth M,** Leidag M, Kotterba S, *et al*. [Estimation of accident risk in obstructive sleep apnea syndrome (OSAS) by driving simulation] (In German). *Pneumologie* 2002;**56**:13—18.
21. **Turkington PM,** Sircar M, Saralaya D, *et al*. Time course of changes in driving simulator performance with and without treatment in patients with sleep apnoea hypopnoea syndrome. *Thorax* 2004;**59**:56—9.
22. **Carsten OMJ,** Groeger JA, Blana E, *et al*. *Driver performance in the engineering and physical sciences research council driving simulator: a validation study. Driver Performance in the Engineering and Physical Sciences Research Council Driving Simulator*. Engineering and Physical Sciences Research Council, UK, 1997. Report No. GR/K56162.
23. **Blana E,** Golias J. Differences between vehicle lateral displacement on the road and in a fixed-base simulator. *Hum Factors* 2002;**44**:303—13.
24. **DeLong ER,** DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837—45.
25. **Jamson S,** Lai F, Jamson H. Driving simulators for robust comparisons: a case study evaluating road safety engineering treatments. *Accid Anal Prev* 2010;**42**:961—71.
26. **Engstrom J,** Aust ML, Vistrom M. Effects of working memory load and repeated scenario exposure on emergency braking performance. *Hum Factors* 2010;**52**:551—9.

Journal club

Chronic disease management for tobacco dependence

This randomised control trial compared the efficacy of utilising chronic disease management principles for tobacco dependence using a tailored intervention with standard care. As tobacco dependence is a chronic relapsing condition, the tailored intervention was chosen to account for possible interim setbacks.

Four hundred and forty-three eligible participants received five telephone-counselling calls and 4 weeks of nicotine replacement therapy. They were randomised to receive continuing counselling and nicotine replacement therapy for 1 year (longitudinal care, LC) or to receive one additional call at 8 weeks (evidence-based usual care, UC). The primary outcome was 6 months of prolonged abstinence, measured at 18 months following initial quit date. Secondary outcomes included abstinence rates before 6 months and smoking reduction.

At 18 months, 30.2% of LC participants reported 6 months of abstinence from smoking, compared with 23.5% in UC. Prior to 6 months, abstinence rates were slightly higher with UC than LC. At all time points, those who did not quit had greater smoking reduction with LC than UC (statistically significant only at 12 months). With LC, quit rates rose throughout the year without reaching a plateau, suggesting that extending treatment further may be beneficial.

One limitation of this study was the difficulty in differentiating between the effects of behavioural and medical treatment. Results were not biochemically confirmed, but the study population was believed to be low-risk for incorrectly reporting smoking status. The LC model allowed counsellors to adjust treatment in response to smokers' experiences of quitting and to positively reinforce the option of interim smoking reduction. Chronic management appears to be a feasible approach to increase long-term abstinence.

► **Joseph AM,** Fu SS, Lindgren B, *et al*. Chronic disease management for tobacco dependence. *Arch Intern Med* 2011;**171**:1894—900.

Harshani M Chandrasekera

Correspondence to Dr Harshani M Chandrasekera, FY2, Department of Thoracic Medicine, The Royal Free Hospital, Pond Street, London NW3 2QG, UK; harshani.chandrasekera@nhs.net

Published Online First 27 January 2012

Thorax 2012;**67**:821. doi:10.1136/thoraxjnl-2012-201586

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.