BMJ Open

SOLVING THE PROBLEM OF RELIABILITY IN UNSUPERVISED APNOEA DETECTION WITH A NEW WEARABLE MEDICAL DEVICE

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005299
Article Type:	Research
Date Submitted by the Author:	19-Mar-2014
Complete List of Authors:	Rodriguez-Villegas, Esther; Imperial College London, Chen, Guangwei; Imperial College London, Dept of Electrical and Electronic Engineering Radcliffe, Jeremy; Institute of Neurology, Duncan, John; University College London,
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Diagnostics
Keywords:	RESPIRATORY MEDICINE (see Thoracic Medicine), Adult intensive & critical care < ANAESTHETICS, SLEEP MEDICINE

SOLVING THE PROBLEM OF RELIABILITY IN UNSUPERVISED APNEA DETECTION WITH A NEW WEARABLE MEDICAL DEVICE

Rodriguez-Villegas Esther[1], PhD, SMIEEE; Gwangwei Chen[1]; Radcliffe Jeremy[2], FRCA, Duncan John[3], FRCP

[1]Dept of Electrical and Electronic Engineering,
Imperial College London, Exhibition Road
London SW7 2AZ, UK

[2]Dept of Neuroanaesthesia
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

[3]Dept of Clinical and Experimental Epilepsy
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

Corresponding author:
Esther Rodriguez-Villegas
Circuits and Systems
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road
London, SW7 2AZ, UK
e.rodriguez@imperial.ac.uk, +4475946261, Fax: +44 207 581 4419

Keywords: Apnea, Breathing, Low Power Electronics, Medical Devices, Physiological Monitoring

Word Count. 3500

ABSTRACT

Objectives. Current techniques for monitoring patients for apnea suffer from significant limitations. These include insufficient availability to meet diagnostic needs, cost, accuracy of results in the presence of artifacts, or difficulty of use in unsupervised conditions. We created and clinically tested a novel miniature medical device that overcomes these limitations.

Participants. We studied 20 healthy control subjects and 10 sleep apnea patients.

Primary Outcomes. The performance of the new system and also of the FDA approved SOMNO clinical system, conventionally used for sleep apnea diagnosis was evaluated in the same conditions. Both systems were tested during a normal night of sleep in both controls and patients. Their performance was quantified in terms of detection of apnea and hypopnea in individual 10 second epochs, which were compared with scoring of signals by a blinded clinician.

Results. For spontaneous apneas during natural sleep and considering the clinician scorer as the gold standard the new wearable apnea detection device had 88.6% (CI: 85.4-91.8) sensitivity and 99.6% (CI: 99.6-99.7) specificity. In comparison the SOMNO system had 14·3% (CI: 10.8-17.8) sensitivity and 99·3% (CI: 99.2-99.4) specificity. The novel device had been specifically designed to detect apnea, but if both apnea and hypopneas during sleep were considered in the assessment, the sensitivity and specificity were 77·1% (CI: 73.8-80.5) and 99·7% (CI: 99.7-99.8) respectively; versus 54% (CI: 50.0-57.9) and 98.5% (CI: 98.4-98.6) for the SOMNO.

Conclusions. The performance of the novel device compares very well to the scoring by an experienced clinician even in the presence of breathing artifacts. This can potentially make it a real solution for apnea home monitoring.

SUMMARY

Strenghts:

- We present the smallest, least intrusive technology to automatically detect apneas/hypopneas
- Performance characterization in normal signal conditions and with signal artefacts, showing excellent agreement with expert- 60,000 epochs assessed in controls and patients
- Sensitivity six times better than a state of the art commercial system, and excellent scoring in terms of user acceptance.

Limitations:

- The size of the study is limited. This is however justified by the fact that it was an initial pilot study to prove the strength of this novel technology to detect individual events even in the presence of artefacts (study goals of 95% confidence intervals for sensitivity and specificity values).
- The technology is still not optimized for hypopnea detection.

INTRODUCTION

Apnea may occur acutely in the context of infectious, respiratory, cardiac and neurological disease;[1-5] be caused by medication;[6-8] and on occasion death may be averted with urgent intervention.[1, 6, 9-11] Apnea may also occur recurrently either as a co-morbidity in chronic conditions including asthma, gastro-oesophageal reflux, neuromuscular disorders and diabetes;[12-16] or on its own in sleep apnea syndrome.[17-27]

The importance of monitoring and quantifying apneas is widely acknowledged. Apneas are one of the two leading causes of Sudden Death in Epilepsy (SUDEP), which only in the UK affects more people than cot death and AIDS together.[28-29] Apneas are also known to be a major problem due to its potentially disastrous consequences in anaesthesia recovery rooms.[30-31] And just sleep apnea may affect between 2% and 10% of the adult population [24] and 1% to 3% of the pediatric population,[18] and is heavily under-diagnosed.[19] The indirect medical costs of under diagnosed adult patients, in the years preceding the diagnosis, is estimated to increase by up to a two-fold, even after correcting for chronic disease status.[22-23] This, added to the potential social consequences, in the form of accidents, increased morbidity and impact on work efficiency makes the condition a major public health issue.[24]

Currently existing techniques for monitoring and quantification of apneas are not satisfactory. In sleep apnea diagnosis, polysomnography is the gold standard but the lack of sleep labs, sleep specialists and the associated cost, either make it difficult for the family physician to confirm the suspicion, or delays diagnosis.[25] The importance of the problem has led Medicare and Medicaid in the USA to recently authorize payment of treatment for adults diagnosed with unattended home sleep monitoring devices.[24] Unfortunately

existing home monitoring devices suffer from one or several of the following limitations: the sensors can be difficult to place resulting in invalid recordings, they still require considerable specialist time in order to interpret the results, automatic interpretation is very inaccurate mainly due to the inability to deal with artifacts, sensors can be cumbersome or intrusive so affecting the quality of sleep.[26-27] Furthermore, there is no portable apnea monitoring system that can detect apneas with high enough sensitivity and specificity in real time to potentially be used to alert carers of life-threatening situations due to acute apnea that can occur in the context of other clinical scenarios such as epilepsy or in anaesthesia. In these scenarios also, the alternative of relying on devices that might be able to detect the sequelae of apnea (for example pulse oximeters to detect hypoxeamia or heart rate monitors to detect bradycardia) might result in fatal consequences due to a delayed response to the apnea.

We present the results of the first clinical study of a new wearable apnea detection device (WADD) specifically designed to overcome the limitations of all other existing technologies.

METHODS

Device

We determined that the strongest externally detectable signal related to breathing corresponded to turbulence in the trachea. This signal was detected with a customized acoustic chamber that optimized the signal transmission. The signal detected by the sensor has components corresponding to both the wanted “signal” (breathing) and undesired “noise” caused by artifacts (cardiac signal, external noise (eg speech, music, wind), movement causing rubbing against the sensor and electromagnetic interference). A novel signal processing algorithm was developed to differentiate “signal” from “noise”. The

algorithm evaluates both the temporal and frequency characteristics of the signal obtained from the sensor. Over 15 different features are analyzed with parametric functions that dynamically adapt over time, to compensate for changes in both the subject and the environment. No pre-calibration or subject specific knowledge or modification is needed for the sensor or the signal processing algorithm.[32-34] Part of the algorithm was implemented on hardware and incorporated into the sensor. This reduces the amount of data that is needed for wireless transmission and consequently the amount of power required from the battery; hence the small size of the device. The wireless receiver and the remaining part of the algorithm were run on a laptop computer.

The WADD was wireless, measured 3·74 by 2·4 by 2·1 cm, weighed 17 grams, and was fixed to the skin on the neck with hydrocolloid colostomy adhesive patches of approximately 4cm diameter (Boots). The preferred location was over the trachea, halfway between the lower margin of the thyroid cartilage and the supra-sternal notch (Figure 1(a)). If the skin in that location was loose, as was common in subjects over 40 years of age, the device was placed antero-laterally, anterior to the sternomastoid muscle. The device was left in place overnight, for approximately 14 hours.

Participants

The study was conducted in a sleep study room of the National Hospital for Neurology and Neurosurgery (UK). We studied 20 healthy controls and 10 patients who were admitted for diagnostic monitoring of sleep-related disorders of breathing because these were likely to have spontaneous apnea events. They also had a variety of neurological conditions, including epilepsy, dementia, neuropathy and motor neuron disease. The reasons to recruit patients who had been referred for diagnosis of possible sleep-related disorders of

breathing, as opposed to those who had been already diagnosed, were twofold. First, the purpose of this study was not to evaluate the WADD for sleep apnea diagnosis, but rather to evaluate its ability to detect individual events, both during controlled conditions to assess the robustness to artifact rejection, and during spontaneous sleep. Good performance on individual event identification would however be expected to translate in a good performance in the context of the different clinical applications. Secondly, non-diagnosed patients were recruited because studying diagnosed individuals would have involved either delay or interruption of their treatment. The decision on the number of patients was based on obtaining a large enough number of events that would lead to the study goals of 95% confidence intervals for sensitivity and specificity values. A larger number of controls were included to be able to assess specificity amongst those who were most likely to be disease free, and also in the presence of artifacts. The patient group comprised 2 females and 8 males with: a median age of 44·5 years of age (range 25-82); a median weight of 74 Kg (range 41-187); a median height of 177 cm (range 160-188); a median body mass index (BMI) of 23 Kg/cm^2 (range 17-61); and a median neck circumference of 40 cm (range 30-43). The control group comprised 3 females and 17 males with: a median age of 33·5 years of age (range 23-63); a median weight of 81.5 Kg (range 60-120); a median height of 176 cm (range 145-185); a median body mass index (BMI) of 26·5 Kg/cm2 (range 20-36); and a median neck circumference of 38 cm (range 34-48). Overall 40% of the subjects were overweight and 24% were obese. The study was approved by the Medicine and Healthcare products Regulatory Agency (MHRA) and the Research Ethics Committee of the UK National Hospital for Neurology and Neurosurgery.

Procedure

All subjects also had simultaneous, clinically standard respiratory monitoring comprising: finger oximetry; oro-nasal airflow sensors; thoracic and abdominal expansion bands; and ECG; using the SOMNO polysomnography system (SOMNOscreen ™ RC kombi . SOMNO Medics, Germany)- Figure 1 (b). Additionally, to further facilitate expert interpretation of polysomnography data, a second pulse oximeter (Pulsox-300i, Konica Minolta, Japan) was attached to the free hand. After attachment of the WADD and the SOMNOmedics polysomnography system, controls subjects participated in a series of exercises, comprising:

T1. Normal breathing for 5 minutes.

T2. Shallow breathing for 5 minutes.

T3. Normal breathing for 45 secs alternating with 15 secs instructed breath holds for 5 min.

T4. Normal breathing for 30 secs alternating with 30 secs instructed breath holds for 10min.

T5. As in 4 but with loud music in the background.

T6. Normal breathing while walking for 5 minutes.

T7. Normal breathing for 30 secs alternating with 30 secs instructed breath holds while lying prone for 5 min.

These exercises were designed to be representative of the worse case of artifact situations affecting the WADD following previous, very exhaustive, lab based research and testing. Following the exercises subjects were allowed to prepare for sleep and were left undisturbed overnight.

Data analysis

The breathing exercises data were analyzed by the automated WADD software and the automated SOMNO software. Instructed apneas were considered to be the "true events". The last six hours of sleep were blindly analyzed by: the automated WADD software, the automated SOMNO software, and by the experienced clinician who reviewed the raw signals from all SOMNO sensors, and had no knowledge about how WADD had been designed or worked. The reason to evaluate the last six hours of sleep was to try to keep the same amount of sleep data in as many subjects as possible in order to prevent biasing of the results. The pulse oximeter was also used by the clinician to support the diagnostic decisions and also individual event classification mostly in those cases in which the signals from the other SOMNO pulse oximeter was corrupted by artifacts. After the separate classification of WADD and SOMNO data, a further investigator compared the results.

The breathing exercises data were analyzed in 15 seconds epochs because this was the shortest duration of an instructed apnea. The sleep data was analyzed in 10 second epochs.

Two assessments were carried out of the sleep data. In the first assessment there was no pre-assumption of a gold standard, and the three systems (WADD, SOMNO and expert marker) were put under test and treated indistinctively. An epoch would be classified as true positive apnea or true positive hypopnea if at least two out of the three systems concurred on the classification. In the second assessment the final classification of these epochs would be that of the expert market, or in other words the expert marker was considered to be the gold standard deciding, and the performance of both SOMNO and WADD system was evaluated. The SOMNO was evaluated as well as the WADD, as there is

little or no quantitative information about the accuracy of automated polysomnography systems.

In both assessments epochs could be classified as:

a) True Positive Apnea (cessation of breathing signal, with correspondent absence of respiratory airflow)
b) True Positive Hypopnea (over 50% reduction in oronasal signal and in thoracoabdominal movement together with over 2% decrease in oxygen saturation).
c) False Positive Hypopneas (if a system had classified a breathing epoch as a hypopnea).
d) False Positive Apnea (if a system had classified a breathing epoch as an apnea)
e) False Classification Apnea as Hypopnea (if a system had classified an apnea epoch as hypopnea).
f) False Classification Hypopnea as Apnea (if a system had classified hypopnea as apnea).
g) False Negative Apnea (if a system classifies an apnea as breathing).
h) False Negative Hypopnea (if a system classifies a hypopnea as breathing).

The breathing exercises data were analyzed in the same way, but the instructed apneas and breathing sections were considered the absolute truth and hence there was no independent expert review.

The performance of the three systems was evaluated using the following metrics:

Sensitivity=(TP)/(TP+FN)

Specificity= (FP)/(TN+FP)

(TP=True Positive, TN=True Negative, FP=False Positive,FN=False Negative).

For each one of the two assessments (i.e. not presuming a gold standard, and considering the expert to be the gold standard), two different analyses were carried out. Firstly only apneas were considered to be true positives. Hence any hypopnea would be regarded as breathing (true negative); False Classification of Hypopneas as Apneas were re-classified as false positives; and False classification of Apneas as Hypopneas were re-classified as false negatives. Secondly, apneas and hypopneas were considered indistinctively, and hence true events of both variety would be also considered together.

These two analyses were carried out as they would be relevant to different clinical scenarios. For example, high sensitivity for detecting apnea would be crucial for identification of sudden apnea if monitoring those with epilepsy; whereas for diagnosis of sleep-related breathing dysfunction, which generally relies on the Apnea Hypopnea Index, the differentiation between apnea and hypopnea might be clinically less important.

RESULTS

Breathing exercises data

Data were available in 3956 15 second epochs for the controls performing the breathing exercises (132 in total). Table 1 summarizes the performance of WADD and SOMNO in the seven breathing exercises. Figure 2 illustrates examples of the signals obtained from the different sensors. Table 1 is divided in three parts. Part (a) and (b) quantify performance considering different scenarios for wrongly classified hypopneas. Although the real events were apneas, both systems had the ability to indicate hypopneas too. This resulted in some real apnea and breathing epochs being wrongly marked as hypopneas. In order to account for these, Table 1 (a) shows the sensitivity and specificity when only apneas are considered

as events (i.e. hypopneas would be regarded as breathing). In part (b) of the table hypopneas and apneas are indistinctively considered. Part (c) illustrates the total number of epochs that fall into a specific classification for both systems. The combined sensitivity and specificity for all the exercises across all the subjects for the WADD was 97·7% and 99·6% (considering hypopneas as breathing); or 99·2% and 99·5% (considering hypopneas as events). With the same criteria the sensitivity and specificity for the SOMNO was only 37·8% sensitivity, 96·5% specificity; or 62·8% sensitivity, 90·5% specificity.

WADD versus Somno performance in instructed exercises with hypopneas NOT considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99.2
T2	NA	NA	100	90.6
T3	94·6	38	99	96·9
T4	98·9	38·8	99·7	94·5
T5	99·2	31·4	99·2	99·7
T6	NA	NA	100	96·5
T7	94·2	48·2	98·5	99
Total	**97·7**	**37·8**	**99·6**	**96·5**

(a)

WADD versus Somno performance in instructed exercises with hypopneas and apneas indistinctively considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99
T2	NA	NA	100	81.4
T3	96·7	66.3	99	87·5
T4	100	64·6	99·7	87·9
T5	99·2	59.2	99·2	95·1
T6	NA	NA	100	89·8
T7	99	64·4	97	93.4
Total	**99·2**	**62·8**	**99·5**	**90·5**

(b)

Summary of classification of the different epochs										
Exercise	**TP**		**TN**		**FC**		**FP**		**FN**	
	WADD	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**
T1	0	0	380	376	0	0	0	4	0	0
T2	0	0	360	293	0	0	0	67	0	0
T3	89	61	285	252	2	26	3	36	3	31
T4	356	230	363	320	4	92	1	44	0	126
T5	357	213	365	350	0	100	3	18	3	147
T6	0	0	400	359	0	0	0	41	0	0
T7	189	123	191	184	9	31	6	13	2	68
Total	**991**	**627**	**2344**	**2134**	**15**	**249**	**13**	**223**	**8**	**372**

(c)

Table 1: Summary of performance for the WADD and SOMNO across the seven breathing exercises (as detailed in Procedure). TP= true positive (apnea), TN=true negative (breathing), FC= false classification, FP= False Positive, FN=False Negative. Part (a) of the table shows the sensitivity and specificity not considering hypopneas as events (i.e. all hypopneas are considered breathing). Based on this all False Classifications, FC (apneas wrongly classified as hypopneas) are considered False Negatives (FN); and all False Positives hypopneas are considered True Negatives (TN). Part (b) shows the sensitivity and specificity considering apnea and hypopnea as indistinctive events. Based on this all False Classifications are re-classified as True Positives (TP); and all False Positives hypopneas are False Positives (FP). Part (c) details the number of epochs corresponding to a particular classification.

Sleep data

For the sleep data 62,727 10 second epochs were analyzed in total. 34 true apnea epochs and 40 true hypopnea epochs were identified for the controls (36 and 37 if the clinician scorer was considered to be the gold standard); and 312 apnea epochs and 181 hypopnea epochs for the patients (342 and 200 if the clinician scorer was considered to be the gold standard). The average number of apnea epochs for the patient group throughout the night was 32. All patients had episodes of apnea or hypopnea. There was only two patients who did not have any episode of apnea. For one control, only 3·2 hours of data were recorded, because of an ICT error. For one patient, only 3 hours were analyzed as more than one

SOMNO sensor including the nasal cannula and the pulse oximeters detached prematurely. The results in terms of sensitivity and specificity, for the control group, patient group and overall are presented in Table 2. Table 2 is divided in four parts: the first and second evaluate the performance for apnea and apnea/hypopnea combined detection respectively without assuming a gold standard (i.e. the consensus of the majority determines a true event); and the third and fourth parts present the same evaluation but considering the expert as the gold standard.

Apnea detection (% sensitivity and specificity)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**47·1**	**97·1**	**94·1**	**99·3**	**99·8**	**100**
(95 CI)	(30.3-63.8)	(91.4-100)	(86.2-100)	(99.2-99.3)	(99.7/99.8)	(100-100)
Patients	**14·7**	**99·4**	**98·1**	**99·5**	**99·5**	**99·9**
(95 CI)	(10.8-18.7)	(98.5-100)	(96.6-99.6)	(99.5-99.6)	(99.4-99.6)	(99.8-99.9)
All	**17·9**	**99·1**	**97·7**	**99·4**	**99·7**	**99·9**
(95 CI)	(13.9-22.0)	(98.2/100)	(96.1-99.3)	(99.3-99.4)	(99.6-99.7)	(99.9-100)

Apnea and Hypopnea combined detection (%)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
All	**57·8**	**84·1**	**98·2**	**98·4**	**99·5**	**99·9**
(95 CI)	(53.8-61.9)	(81.1-87.1)	(97.2-99.3)	(98.3-98.5)	(99.5-99.6)	(99.9-100)

Apnea Detection with the clinician scorer as Gold Standard reference (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**38·9**	**86·1**	**99·2**	**99·7**
(95 CI)	(23.0-54.8)	(74.8-97.4)	(99.1-99.3)	(99.7-99.8)
Patients	**11·7**	**88·9**	**99·5**	**99·4**
(95 CI)	(8.3-15.1)	(85.6-92.2)	(99.4-99.6)	(99.3-99.5)
All	**14·3**	**88·6**	**99·3**	**99·6**
(95 CI)	(10.8-17.8)	(85.4-91.8)	(99.2-99.4)	(99.6-99.7)

Apnea and Hypopnea combined detection with clinician scorer as Gold Standard (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
All	**54·0**	**77·1**	**98·5**	**99·7**
(95 CI)	(50.0-57.9)	(73.8-80.5)	(98.4-98.6)	(99.7-99.8)

Table 2: Summary of performance for the WADD, SOMNO and clinician scorer systems for detection of apnea and hypopnea in 15 second epochs of overnight recordings.

The WADD also had the added feature of being able to differentiate between central and obstructive apnea. 90% of the central apneas were rightly marked as central. 96% were rightly marked as obstructive. Approximately 60% of the total apneas were obstructive in origin.

Device comfort

After the overnight study, the devices were detached and the subjects scored the comfort of the devices and quality of sleep (rating 1 to 5, with 5 representing maximum comfort and quality). Skin irritation caused by the WADD's adhesive was also rated from 1 to 5 (5 representing no irritation, 4 mild transient, redness, and 1 severe irritation). The median rating for WADD comfort was 5 (range 4-5). The median rating for SOMNO comfort was 3 (range 1-5 for controls and 2-5 for patients). The median rating for irritation caused by the WADD plaster on the neck was 5 (range 5-5 for controls and 4-5 for patients).

DISCUSSION

Main findings

WADD had very high sensitivity and specificity for detecting apnea in 15 second epochs in a series of breathing and breath-holding exercises in a variety of conditions, including the presence of external background noise, movement and posture. The tolerability of WADD was superior to the portable polysomnography system (SOMNO) during overnight recordings.

WADD had 97·7-99·2% sensitivity to detect instructed apneas and 88.6-99·1% for 10 seconds spontaneous apneas during natural sleep, with similar performance in controls and patients. The WADD also detected all apneas over 30 seconds and there were only 3 over

30 seconds false positives. For short apneas, in most cases, disagreement between the clinician scorer and the WADD were caused by the WADD identifying as apnea epochs that the expert classified as hypopneas.

As expected, the WADD performance was less good when apneas and hypopneas were considered together (minimum sensitivity 77·1%).This is not surprising since the WADD was designed to identify apnea, not hypopneas, and the latter were detected from the transmitted signal which had already been pre-processed for apnea detection. From the table, it can be observed that the degradation of performance was more evident in the controls because the controls had a large number of shorter hypopneas (under 22.5 seconds) which the WADD did not detect properly. In the patients, who often demonstrated apneas, the hypopnea events were longer and these were detected by WADD. Although the lower sensitivity in hypopnea detection might in principle seem problematic if the WADD was to be used in the context of sleep apnea diagnosis (hypopneas are very common events in sleep labs), it is worth noting that: 1) there is no other reported automatic system that gets anywhere close to this with similar specificity and apnea detection performance; 2) the variations between different sleep labs due to the non-uniform definition of hypopneas already leads to much larger diagnostic variations than the limitation in sensitivity of the WADD;[35-37] 3) assuming the worse case scenario for the WADD, this is that a patient only had hypopneas throughout the night, this reduced sensitivity would be a problem that would translate to non-diagnosis of sleep apnea for patients that with 100% sensitivity would have had a sleep apnea hypopnea index (AHI) between 5 and 6.5 (i.e very mild cases of sleep apnea). Patients with no sleep apnea, moderate sleep apnea, severe sleep apnea

and those with mild sleep apnea with AHI between 6.5 and 15 would have been rightly diagnosed.

SOMNO Performance

The automatic analysis of the SOMNO apparatus, an FDA approved and clinically accepted system, based on assessing apnea from a variety of different sensors, significantly differed from that of the expert marker, with an average sensitivity value of around 14%. The results obtained from the instructed apneas tests also showed that even in the absence of artifacts, apneas were not well detected by the SOMNO system, with an average sensitivity of 37·8%. This demonstrates the need for caution if relying on current automated assessment methods for diagnosing apnea. Whilst performance might be improved by optimizing parameters for individual patients, this is not practical for single overnight recordings or use as an alerting monitor. The WADD does not require any parameter optimization or subject specific calibration.

The SOMNO system performance improved in the event of indistinct classification of apneas and hypopneas, but was still poor compared with the clinician scorer (54% sensitivity). This sensitivity was at the cost of reduced specificity: for every true hypopnea detected there were approximately four false detections. Overall, the performance of the WADD in hypopnea/apnea combined detection was significantly better than the SOMNO, in sensitivity (77·1% vs 54% if considering the clinician scorer as a gold standard, and 84·1% vs 57·8% otherwise), but also in specificity, as the WADD only detected one false hypopnea epoch for every four true events.

Limitations of current design. Future improvements

The WADD is obviously no substitute to a full night study in a sleep clinic, since it does not provide all the information that a full polysomnography system would. There are advantages and disadvantages to this device with respect to full polysomnography. The WADD can be used to determine the Apnea Hypopnea Index (AHI), which is used in sleep apnea diagnosis to ascertain whether a patient has sleep apnea and to score the severity of the condition. The main advantage is that it can be used for at home assessment or monitoring, and from that point of view it is clearly superior to any of the other existing devices (highly resilient to artifacts, very easy to attach and durable in position, low cost, much more comfortable, and accurate). Considering the restricted resources for sleep clinic referral this device could be a very useful tool to determine at very low cost who should be referred to a specialist centre for full polysomnography. The disadvantage is that there are other parameters that could be used for extra assessment that the device does not measure, such as microarousal or full cardiac activity.

The WADD device used in the current study relied on wireless transmission to a PC. However changing the PC to a dedicated mobile phone sized receiver poses no technological challenge. A subsequent version that is being developed is smaller (2.4 by 2.4 by 1.2 cm, weighing 7.5 grams) and can operate continuously on hearing aid batteries for over 48 hours. It has a separate dedicated receiver of comparable size to a mobile phone which can be located up to 10 metres from the subject.

Funding. UCL Business and Imperial Innovations.

Authors Contributions

John Duncan and Esther Rodriguez- Villegas had the initial idea of developing a wearable breathing monitoring device. Esther Rodriguez-Villegas and Guangwei Chen were the main creators of the novel device carrying out the electronic design which includes both hardware and software. John Duncan contributed through the process helping to define the engineering specifications to meet the clinical need. John Duncan designed the protocol for the clinical study. Esther Rodriguez-Villegas and Guangwei Chen created the documentation for MHRA approval. Jeremy Radcliffe led the clinical studies and blindly marked all the data using the sensors of the SOMNO system as well as the Konica Minolta pulse oximeter. Esther Rodriguez-Villegas and Guangwei Chen compared the results of the data scored by the WADD, SOMNO and blinded expert, and carried out the data analysis. All the authors contributed to the interpretation of data. John Duncan, Jeremy Radcliffe and Esther Rodriguez- Villegas wrote the paper.

Data Sharing. Extra information may be available by emailing e.rodriguez@imperial.ac.uk.

Acknowledgments: Eduardo Aguilar-Pelaez and for assisting with the data collection at the initial phases of the clinical study; UCL Business, Epilepsy Society and Staff on MITU at NHNN.

Conflict of interest: There are no competing interests.

References

1 Shorvon S, Tomnson T. Sudden unexpected death in epilepsy. *Lancet* 2011;378(9808):2028–38.

2 Mace SE. Acute bacterial meningitis. *Emerg Med Clin North Am* 2008;**26**(2):281–317.

3 Cordero L, Rau R, Taylor D, et al. Enteric gram-negative bacilli bloodstream infections: 17 years' experience in a neonatal intensive care unit. *Am J Infect Control* 2004;**32**(4):189–95.

4 Powers, KS. Acute pulmonary infections. In: Lucking SE, Maffei, FA, Tamburro, RF and Thomas, NJ, editors. Pediatric critical care study guide. London: Springer 2012; 514–534.

5 Moruzzi P, Sarzi-Braga S, Rossi M, et al. Sleep apnea in ischaemic heart disease: differences between acute and chronic coronary syndromes. *Heart* 1999;**82**(3):343–7.

6 Van Vliet G, Deal CL, Crock PA, et al. Sudden death in growth hormone-treated children with Prader-Willi syndrome. *J Pediatr* 2004;**144**(1):129–31.

7 Taylor MB, Grounds RM, Mulrooney PD, et al. Ventilatory effects of propofol during induction of anaesthesia. *Anaesthesia* 1986;**41**(8):816–20.

8 Murphy JJ, Swanson T, Ansermino M, et al. The frequency of apneas in premature infants after inguinal hernia repair: do they need overnight monitoring in the intensive care unit? *J Pediatr Surg* 2008;**43**(5):865-8.

9 Naeem N, Montenegro H. Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation* 2005;**67**(1): 13-23.

10 Moon RY, Horne RS, Hauck FR. Sudden infant death syndrome. *Lancet* 2007;**370**(9598):1578–87.

11 Gami AS, Howard DE, Olson EJ, et al. Day-night pattern of sudden death in obstructive sleep apnea, *N Engl J Med* 2005;**352**(12):1206–14.

12 Julien JY, Martin JG, Ernst P, et al. Prevalence of obstructive sleep apnea–hypopnea in severe versus moderate asthma. *J Allergy Clin Immunol* 2009;**124**(2):371–6.

13 Emilsson O, Janson C, Benediktsdóttir B,et al. Nocturnal gastroesophageal reflux, lung function and symptoms of obstructive sleep apnea: Results from an epidemiological survey. *Respir Med* 2012;**106**(3):459–66.

14 Reichmuth K, Austin D, Skatrud JB, et al. Association of sleep apnea and type II diabetes. A population-based study. *Am J Respir Crit Care Med* 2005;**172**(12):1590–5.

15 Hanly PJ. Mechanisms and management of central sleep apnea. *Lung* 1992;**170**(1):1–17.

16 Culebras A. Sleep and neuromuscular disorders. *Neurol Clin* 2005;**23**(4):1209–23.

17 Young T, Palta M, Dempsey J, et al. The occurrence of sleep disordered breathing among middle aged adults. *N Engl J Med* 1993;**328**(17):1230–5.

18 Faroux B. What's new in pediatric sleep? *Pediatr Respir Rev* 2007;**8**(1):85–9

19 Pang KP, Terris DJ. Screening for obstructive sleep apnea; an evidence-based analysis. *Am J Otolaryngol* 2006;**27**(2):112–8.

20 Redline S, Quan SF. Sleep apnea: a common mechanism for the deadly triad--cardiovascular disease, diabetes, and cancer? *Am J Respir Crit Care Med* 2012;**186**(2):123–124.

21 Hirshkowitz M. The clinical consequences of obstructive sleep apnea and associated excessive sleepiness. *J Fam Pract* 2008;**57**(8 Suppl):S9–16.

22 Wittmann V, Rodenstein DO. Health care costs and the sleep apnea syndrome. *Sleep Medicine Rev* 2004;**8**(4):269–279

23 Smith R, Ronald J, Delaive K, et al. What are obstructive sleep apnea patients being treated for prior to this diagnosis? *Chest* 2002;**12**(1):164-172.

24 Leger D, Bayon V, Laaban JP, et al. Impact of sleep apnea on economics. *Sleep Med Rev* 2012;**16**(5):455–62.

25 Flemons WW, Douglas NJ, Kuna ST, et al. Access to diagnosis and treatment of patients with suspected sleep apnea. *Am J Respir Crit Care Med* 2004;**169**(6):668–72.

26 Oliveira MG, Nery LE, Santos-Silva R, et al. Is portable monitoring accurate in the diagnosis of obstructive sleep apnea syndrome in chronic pulmonary obstructive disease? *Sleep Med*. 2012;**13**(8):1033-8.

27 Flemons WW, Littner MR; Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest* 2003;**124**(4):1543–79

28 McDonald F. Avoidable Epilepsy-Related Deaths: The case for prevention. *Epilepsy Bereaved Report*, 2011.

29 Surges R, Sander JW. Sudden unexpected death in epilepsy: mechanisms, prevalence and prevention. *Curr Opin Neurol* 2012;**2**:201–207.

30 Lofsky A. Sleep apnea and narcotic postoperative pain medication: a morbidity and mortality risk. *Anaesthesia Patient Safety Foundation Newsletter*. Summer 2002;**17**:24–25.

31 Bhananker SM, Posner KL, Cheney FW, et al. Injury and liability associated with monitored anesthesia care: A closed claims analysis. *Anesthesiology* 2006;**104**:228–34.

32 Rodriguez-Villegas E, Corbishley P, Duncan JS, et al. A method and apparatus for monitoring respiratory activity. Patent *WIPO;* WO/2007/099314.

33 Corbishley PG, Rodriguez-Villegas E. Breathing detection: towards a miniaturized wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering* 2008; **55**(1):196–204.

34 Rodriguez-Villegas E, Chen G, Aguilar-Pelaez E. Feature characterization for breathing monitor. Patent *WIPO*; WO/2012/007719.

35 Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. *Medicina* 2003;**63**(2):119–24.

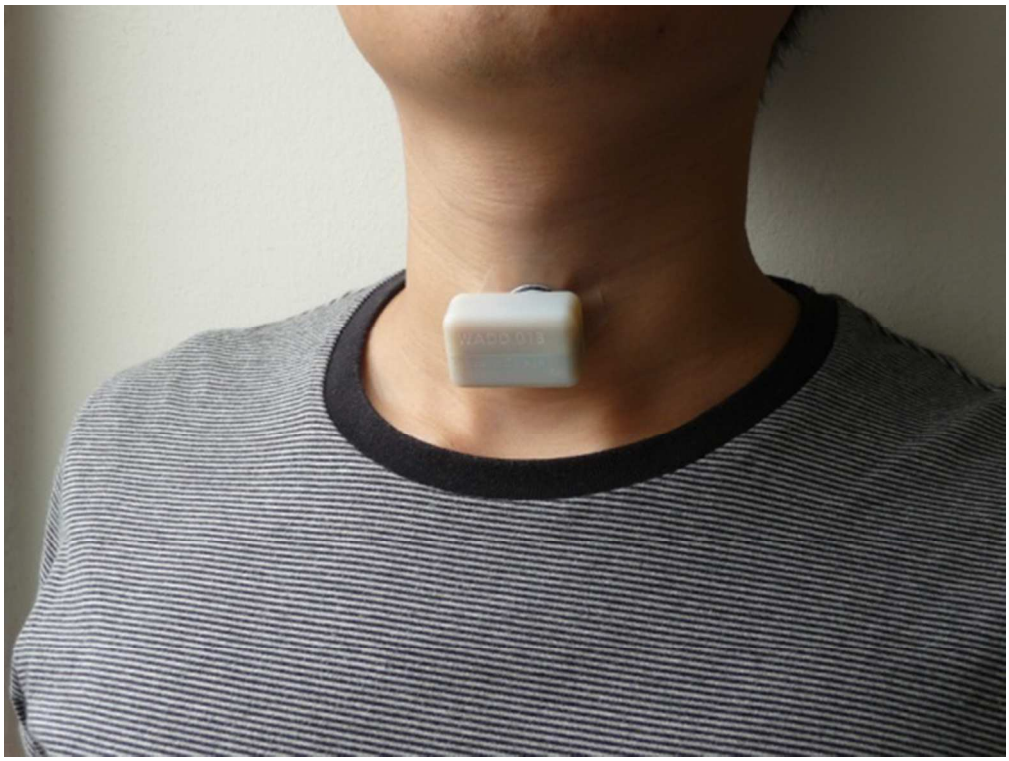
36 Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med* 2000;**161** (2):369–74.

37 Ruehland WR, Rochford PD, O'Donoghue FJ, et al. The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index, *Sleep* 2009;**32**(2):150–157.

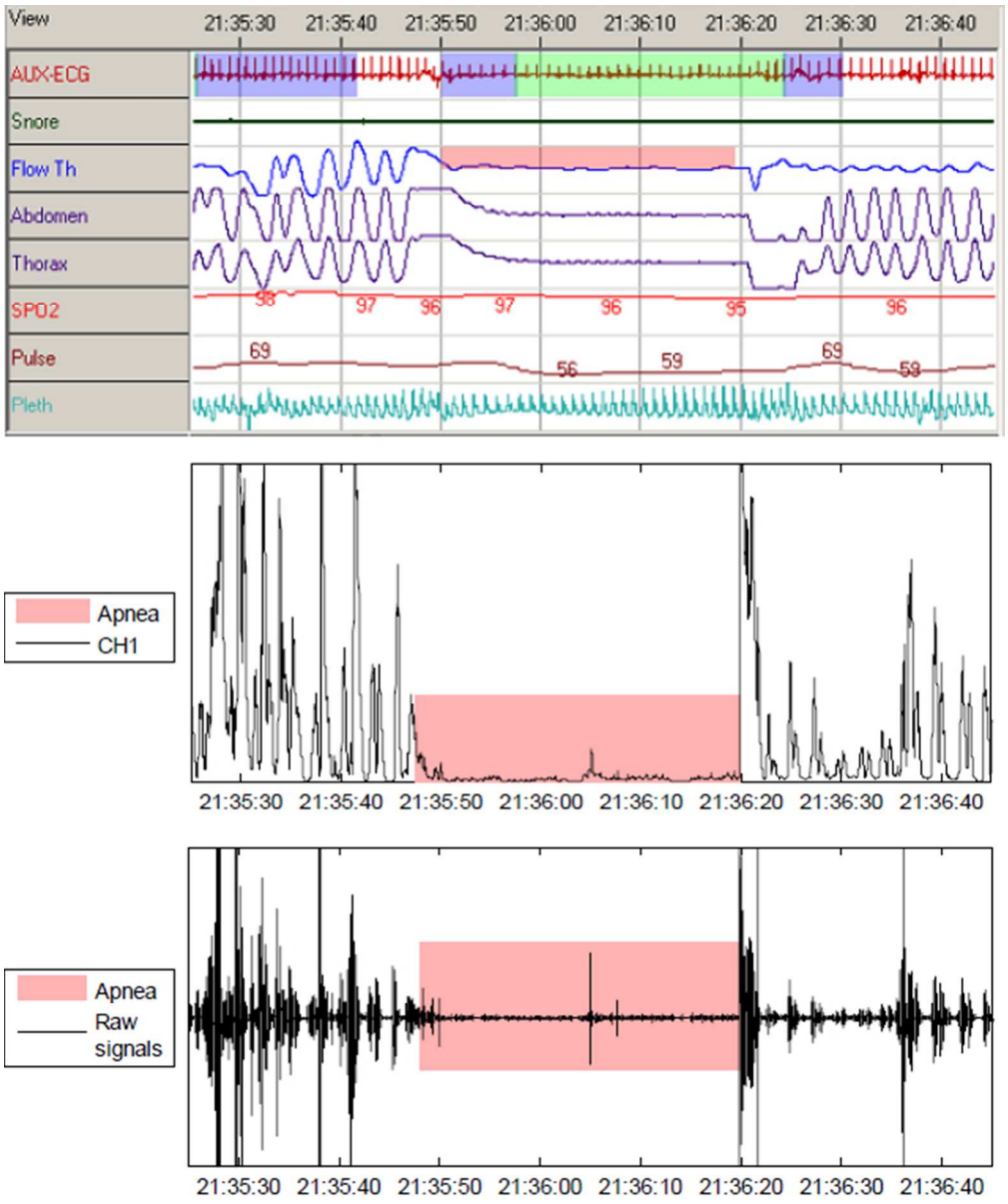
Figure Legends:

Figure 1: (a) WADD worn by one of the investigators. (b) Subject wearing an existing state of the art ambulatory apnea monitoring system (SOMNO), comprising finger oximetry; oro-nasal flow sensors; thoracic and abdominal expansion bands; and ECG.

Figure 2: Illustration of the SOMNO and WADD output signals showing an apnea event: (top) raw signals from the different SOMNO sensors, (middle) processed WADD signal, (bottom) WADD output signal.



67x50mm (300 x 300 DPI)



53x64mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(included, page 1)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(included, page 1)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(included, pages 3-4)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(included, 3-4)**
Methods		
Study design	4	Present key elements of study design early in the paper **(included, pages 5-7)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(included, pages 5-6)**
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants **(included, pages 8-9)**
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case **(NA)**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(included, pages 8-9)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(included, pages 8-10)**
Bias	9	Describe any efforts to address potential sources of bias **(included, pages 8, 10)**
Study size	10	Explain how the study size was arrived at **(included, page 6)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(included, pages 8-10)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(included, pages 6, 14)**
		(*b*) Describe any methods used to examine subgroups and interactions **(NA)**
		(*c*) Explain how missing data were addressed **(page 12)**
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy **(NA)**
		(*e*) Describe any sensitivity analyses **(included)**

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(included, page 6)**
		(b) Give reasons for non-participation at each stage **(NA)**
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(included, page 6)**
		(b) Indicate number of participants with missing data for each variable of interest **(NA)**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) **(NA)**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time (NA)
		Case-control study—Report numbers in each exposure category, or summary measures of exposure **(NA)**
		Cross-sectional study—Report numbers of outcome events or summary measures**(NA)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **(included, page 14)**
		(*b*) Report category boundaries when continuous variables were categorized **(included, pages 8-10)**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **(NA)**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **(NA)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(included, pages 11-15)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(included, page 17)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(included, page 17)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(included, page 17)**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **(included, 19)**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

A Pilot Study of A Wearable Apnea Detection Device

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005299.R1
Article Type:	Research
Date Submitted by the Author:	06-Aug-2014
Complete List of Authors:	Rodriguez-Villegas, Esther; Imperial College London, Electrical and Electronic Engineering Chen, Guangwei; Imperial College London, Dept of Electrical and Electronic Engineering Radcliffe, Jeremy; UCL Institute of Neurology,UK National Hospital for Neurology and Neurosurgery, Dept of Neuroanaesthesia Duncan, John; UCL Institute of Neurology, UK National Hospital for Neurology and Neurosurgery, Dept of Clinical and Experimental Epilepsy
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Diagnostics
Keywords:	RESPIRATORY MEDICINE (see Thoracic Medicine), Adult intensive & critical care < ANAESTHETICS, SLEEP MEDICINE

A PILOT STUDY OF A WEARABLE APNEA DETECTION DEVICE

Rodriguez-Villegas Esther[1], PhD, SMIEEE; Gwangwei Chen[1]; Radcliffe Jeremy[2], FRCA, Duncan John[3], FRCP

[1]Dept of Electrical and Electronic Engineering,
Imperial College London, Exhibition Road
London SW7 2AZ, UK

[2]Dept of Neuroanaesthesia
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

[3]Dept of Clinical and Experimental Epilepsy
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

Corresponding author:
Esther Rodriguez-Villegas
Circuits and Systems
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road
London, SW7 2AZ, UK
e.rodriguez@imperial.ac.uk, +4475946261, Fax: +44 207 581 4419

Keywords: Apnea, Breathing, Low Power Electronics, Medical Devices, Physiological Monitoring

Word Count. 3500

ABSTRACT

Rationale. Current techniques for monitoring patients for apnea suffer from significant limitations. These include insufficient availability to meet diagnostic needs, cost, accuracy of results in the presence of artifacts, or difficulty of use in unsupervised conditions.

Objectives. We created and clinically tested a novel miniature medical device that targets to overcome these limitations.

Methods. We studied 20 healthy control subjects and 10 sleep apnea patients. The performance of the new system and also of the FDA approved SOMNO clinical system, conventionally used for sleep apnea diagnosis was evaluated in the same conditions. Both systems were tested during a normal night of sleep in both controls and patients. Their performance was quantified in terms of detection of apnea and hypopnea in individual 10 second epochs, which were compared with scoring of signals by a blinded clinician.

Main Results. For spontaneous apneas during natural sleep and considering the clinician scorer as the gold standard the new wearable apnea detection device had 88.6% sensitivity and 99.6% specificity. In comparison the SOMNO system had 14·3% sensitivity and 99·3% specificity. The novel device had been specifically designed to detect apnea, but if both apnea and hypopneas during sleep were considered in the assessment, the sensitivity and specificity were 77·1% and 99·7% respectively; versus 54% and 98.5% for the SOMNO.

Conclusions. The performance of the novel device compares very well to the scoring by an experienced clinician even in the presence of breathing artifacts, in this small pilot study. This can potentially make it a real solution for apnea home monitoring.

Word Count: 259

SUMMARY

Strenghts:

- We present the smallest, least intrusive technology to automatically detect apneas/hypopneas
- Performance characterization in normal signal conditions and with signal artefacts, showing excellent agreement with expert- 60,000 epochs assessed in controls and patients
- Sensitivity six times better than a state of the art commercial system, and excellent scoring in terms of user acceptance.

Limitations:

- The size of the study is limited. This is however justified by the fact that it was an initial pilot study to prove the strength of this novel technology to detect individual events even in the presence of artefacts (study goals of 95% confidence intervals for sensitivity and specificity values).

- The technology is still not optimized for hypopnea detection.

INTRODUCTION

Apnea may occur acutely in the context of infectious, respiratory, cardiac and neurological disease;[1-5] be caused by medication;[6-8] and on occasion death may be averted with urgent intervention.[1, 6, 9-11] Apnea may also occur recurrently either as a co-morbidity in chronic conditions including asthma, gastro-oesophageal reflux, neuromuscular disorders and diabetes;[12-16] or on its own in sleep apnea syndrome.[17-27]

The importance of monitoring and quantifying apneas is widely acknowledged. Apneas are one of the two leading causes of Sudden Death in Epilepsy (SUDEP), which only in the UK affects more people than cot death and AIDS together.[28-29] Apneas are also known to be a major problem due to its potentially disastrous consequences in anaesthesia recovery rooms.[30-31] And just sleep apnea may affect between 2% and 10% of the adult population [24] and 1% to 3% of the pediatric population,[18] and is heavily under-diagnosed.[19] The indirect medical costs of under diagnosed adult patients, in the years preceding the diagnosis, is estimated to increase by up to a two-fold, even after correcting for chronic disease status.[22-23] This, added to the potential social consequences, in the form of accidents, increased morbidity and impact on work efficiency makes the condition a major public health issue.[24]

Currently existing techniques for monitoring and quantification of apneas are not satisfactory. In sleep apnea diagnosis, polysomnography is the gold standard but the lack of sleep labs, sleep specialists and the associated cost, either make it difficult for the family physician to confirm the suspicion, or delays diagnosis.[25] The importance of the problem has led Medicare and Medicaid in the USA to recently authorize payment of treatment for adults diagnosed with unattended home sleep monitoring devices.[24] Unfortunately

existing home monitoring devices suffer from one or several of the following limitations: the sensors can be difficult to place resulting in invalid recordings, they still require considerable specialist time in order to interpret the results, automatic interpretation is very inaccurate mainly due to the inability to deal with artifacts, sensors can be cumbersome or intrusive so affecting the quality of sleep.[26-27] Furthermore, there is no portable apnea monitoring system that can detect apneas with high enough sensitivity and specificity in real time to potentially be used to alert carers of life-threatening situations due to acute apnea that can occur in the context of other clinical scenarios such as epilepsy or in anaesthesia. In these scenarios also, the alternative of relying on devices that might be able to detect the sequelae of apnea (for example pulse oximeters to detect hypoxeamia or heart rate monitors to detect bradycardia) might result in fatal consequences due to a delayed response to the apnea.

We present the results of the first clinical study of a new wearable apnea detection device (WADD) specifically designed to overcome the limitations of all other existing technologies.

METHODS

Device

We determined that the strongest externally detectable signal related to breathing corresponded to turbulence in the trachea. This signal was detected with a customized acoustic chamber that optimized the signal transmission. The signal detected by the sensor has components corresponding to both the wanted “signal” (breathing) and undesired “noise” caused by artifacts (cardiac signal, external noise (eg speech, music, wind), movement causing rubbing against the sensor and electromagnetic interference). A novel signal processing algorithm was developed to differentiate “signal” from “noise”. The

algorithm evaluates both the temporal and frequency characteristics of the signal obtained from the sensor. Over 15 different features are analyzed with parametric functions that dynamically adapt over time, to compensate for changes in both the subject and the environment. No pre-calibration or subject specific knowledge or modification is needed for the sensor or the signal processing algorithm.[32-34] Part of the algorithm was implemented on hardware and incorporated into the sensor. This reduces the amount of data that is needed for wireless transmission and consequently the amount of power required from the battery; hence the small size of the device. The wireless receiver and the remaining part of the algorithm were run on a laptop computer.

The WADD was wireless, measured 3·74 by 2·4 by 2·1 cm, weighed 17 grams, and was fixed to the skin on the neck with hydrocolloid colostomy adhesive patches of approximately 4cm diameter (Boots). The preferred location was over the trachea, halfway between the lower margin of the thyroid cartilage and the supra-sternal notch (Figure 1(a)). If the skin in that location was loose, as was common in subjects over 40 years of age, the device was placed antero-laterally, anterior to the sternomastoid muscle. The device was left in place overnight, for approximately 14 hours.

Participants

The study was conducted in a sleep study room of the National Hospital for Neurology and Neurosurgery (UK). We studied 20 healthy controls and 10 patients, as they were sequentially admitted for diagnostic monitoring of sleep-related disorders of breathing, because these were likely to have spontaneous apnea events. The patients and controls were not matched. Patients also had a variety of neurological conditions, including epilepsy, dementia, neuropathy and motor neuron disease. The reasons to recruit patients who had

been referred for diagnosis of possible sleep-related disorders of breathing, as opposed to those who had been already diagnosed, were twofold. First, the purpose of this study was not to evaluate the WADD for sleep apnea diagnosis, but rather to evaluate its ability to detect individual events, both during controlled conditions to assess the robustness to artifact rejection, and during spontaneous sleep. Good performance on individual event identification would however be expected to translate in a good performance in the context of the different clinical applications. Secondly, non-diagnosed patients were recruited because studying diagnosed individuals would have involved either delay or interruption of their treatment. The decision on the number of patients was based on obtaining a large enough number of events that would lead to the study goals of 95% confidence intervals for sensitivity and specificity values (based on clinical experience on the minimum number of apnea events which would be expected per subject referred for sleep apnea diagnosis, per night). A larger number of controls were included to be able to assess specificity amongst those who were most likely to be disease free, and also in the presence of artifacts. The patient group comprised 2 females and 8 males with: a median age of 44·5 years of age (range 25-82); a median weight of 74 Kg (range 41-187); a median height of 177 cm (range 160-188); a median body mass index (BMI) of 23 Kg/cm^2 (range 17-61); and a median neck circumference of 40 cm (range 30-43). The control group comprised 3 females and 17 males with: a median age of 33·5 years of age (range 23-63); a median weight of 81.5 Kg (range 60-120); a median height of 176 cm (range 145-185); a median body mass index (BMI) of 26·5 Kg/cm2 (range 20-36); and a median neck circumference of 38 cm (range 34-48). Overall 40% of the subjects were overweight and 24% were obese. The study was approved by the Medicine and Healthcare products Regulatory Agency (MHRA) and the Research Ethics Committee of the UK National Hospital for Neurology and Neurosurgery.

Procedure

All subjects also had simultaneous, clinically standard respiratory monitoring comprising: finger oximetry; oro-nasal airflow sensors; thoracic and abdominal expansion bands; and ECG; using the SOMNO polysomnography system (SOMNOscreen ™ RC kombi . SOMNO Medics, Germany)- Figure 1 (b). Additionally, to further facilitate expert interpretation of polysomnography data, a second pulse oximeter (Pulsox-300i, Konica Minolta, Japan) was attached to the free hand. After attachment of the WADD and the SOMNOmedics polysomnography system, controls subjects participated in a series of exercises, comprising:

T1. Normal breathing for 5 minutes.

T2. Shallow breathing for 5 minutes.

T3. Normal breathing for 45 secs alternating with 15 secs instructed breath holds for 5 min.

T4. Normal breathing for 30 secs alternating with 30 secs instructed breath holds for 10min.

T5. As in 4 but with loud music in the background.

T6. Normal breathing while walking for 5 minutes.

T7. Normal breathing for 30 secs alternating with 30 secs instructed breath holds while lying prone for 5 min.

These exercises were designed to be representative of the worse case of artifact situations affecting the WADD following previous, very exhaustive, lab based research and testing. Following the exercises subjects were allowed to prepare for sleep and were left undisturbed overnight.

Data analysis

The breathing exercises data were analyzed by the automated WADD software and the automated SOMNO software. Instructed apneas were considered to be the "true events". The last six hours of sleep were blindly analyzed by: the automated WADD software, the automated SOMNO software, and by the experienced clinician who reviewed the raw signals from all SOMNO sensors, and had no knowledge about how WADD had been designed or worked. The reason to evaluate the last six hours of sleep was to try to keep the same amount of sleep data in as many subjects as possible in order to prevent biasing of the results. The pulse oximeter was also used by the clinician to support the diagnostic decisions and also individual event classification mostly in those cases in which the signals from the other SOMNO pulse oximeter was corrupted by artifacts. After the separate classification of WADD and SOMNO data, a further investigator compared the results.

The breathing exercises data were analyzed in 15 seconds epochs because this was the shortest duration of an instructed apnea. The sleep data was analyzed in 10 second epochs.

Two assessments were carried out of the sleep data. In the first assessment there was no pre-assumption of a gold standard, and the three systems (WADD, SOMNO and expert marker) were put under test and treated indistinctively. An epoch would be classified as true positive apnea or true positive hypopnea if at least two out of the three systems concurred on the classification. In the second assessment the final classification of these epochs would be that of the expert market, or in other words the expert marker was considered to be the gold standard deciding, and the performance of both SOMNO and

WADD system was evaluated. The SOMNO was evaluated as well as the WADD, as there is little or no quantitative information about the accuracy of automated polysomnography systems.

In both assessments epochs could be classified as:

a) True Positive Apnea (cessation of breathing signal, with correspondent absence of respiratory airflow)
b) True Positive Hypopnea (over 50% reduction in oronasal signal and in thoracoabdominal movement together with over 2% decrease in oxygen saturation).
c) False Positive Hypopneas (if a system had classified a breathing epoch as a hypopnea).
d) False Positive Apnea (if a system had classified a breathing epoch as an apnea)
e) False Classification Apnea as Hypopnea (if a system had classified an apnea epoch as hypopnea).
f) False Classification Hypopnea as Apnea (if a system had classified hypopnea as apnea).
g) False Negative Apnea (if a system classifies an apnea as breathing).
h) False Negative Hypopnea (if a system classifies a hypopnea as breathing).

The breathing exercises data were analyzed in the same way, but the instructed apneas and breathing sections were considered the absolute truth and hence there was no independent expert review.

The performance of the three systems was evaluated using the following metrics:

Sensitivity=(TP)/(TP+FN)

Specificity= (FP)/(TN+FP)

(TP=True Positive, TN=True Negative, FP=False Positive,FN=False Negative).

For each one of the two assessments (i.e. not presuming a gold standard, and considering the expert to be the gold standard), two different analyses were carried out. Firstly only apneas were considered to be true positives. Hence any hypopnea would be regarded as breathing (true negative); False Classification of Hypopneas as Apneas were re-classified as false positives; and False classification of Apneas as Hypopneas were re-classified as false negatives. Secondly, apneas and hypopneas were considered indistinctively, and hence true events of both variety would be also considered together.

These two analyses were carried out as they would be relevant to different clinical scenarios. For example, high sensitivity for detecting apnea would be crucial for identification of sudden apnea if monitoring those with epilepsy; whereas for diagnosis of sleep-related breathing dysfunction, which generally relies on the Apnea Hypopnea Index, the differentiation between apnea and hypopnea might be clinically less important.

RESULTS

Breathing exercises data

Data were available in 3956 15 second epochs for the controls performing the breathing exercises (132 in total). Table 1 summarizes the performance of WADD and SOMNO in the seven breathing exercises. Figure 2 illustrates examples of the signals obtained from the different sensors. Table 1 is divided in three parts. Part (a) and (b) quantify performance considering different scenarios for wrongly classified hypopneas. Although the real events were apneas, both systems had the ability to indicate hypopneas too. This resulted in some real apnea and breathing epochs being wrongly marked as hypopneas. In order to account

for these, Table 1 (a) shows the sensitivity and specificity when only apneas are considered as events (i.e. hypopneas would be regarded as breathing). In part (b) of the table hypopneas and apneas are indistinctively considered. Part (c) illustrates the total number of epochs that fall into a specific classification for both systems. The combined sensitivity and specificity for all the exercises across all the subjects for the WADD was 97·7% and 99·6% (considering hypopneas as breathing); or 99·2% and 99·5% (considering hypopneas as events). With the same criteria the sensitivity and specificity for the SOMNO was only 37·8% sensitivity, 96·5% specificity; or 62·8% sensitivity, 90·5% specificity.

WADD versus Somno performance in instructed exercises with hypopneas NOT considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99.2
T2	NA	NA	100	90.6
T3	94·6	38	99	96·9
T4	98·9	38·8	99·7	94·5
T5	99·2	31·4	99·2	99·7
T6	NA	NA	100	96·5
T7	94·2	48·2	98·5	99
Total	**97·7**	**37·8**	**99·6**	**96·5**

(a)

WADD versus Somno performance in instructed exercises with hypopneas and apneas indistinctively considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99
T2	NA	NA	100	81.4
T3	96·7	66.3	99	87·5
T4	100	64·6	99·7	87·9
T5	99·2	59.2	99·2	95·1
T6	NA	NA	100	89·8
T7	99	64·4	97	93.4
Total	**99·2**	**62·8**	**99·5**	**90·5**

(b)

Summary of classification of the different epochs										
Exercise	**TP**		**TN**		**FC**		**FP**		**FN**	
	WADD	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**
T1	0	0	380	376	0	0	0	4	0	0
T2	0	0	360	293	0	0	0	67	0	0
T3	89	61	285	252	2	26	3	36	3	31
T4	356	230	363	320	4	92	1	44	0	126
T5	357	213	365	350	0	100	3	18	3	147
T6	0	0	400	359	0	0	0	41	0	0
T7	189	123	191	184	9	31	6	13	2	68
Total	**991**	**627**	**2344**	**2134**	**15**	**249**	**13**	**223**	**8**	**372**

(c)

Table 1: Summary of performance for the WADD and SOMNO across the seven breathing exercises (as detailed in Procedure). TP= true positive (apnea), TN=true negative (breathing), FC= false classification, FP= False Positive, FN=False Negative. Part (a) of the table shows the sensitivity and specificity not considering hypopneas as events (i.e. all hypopneas are considered breathing). Based on this all False Classifications, FC (apneas wrongly classified as hypopneas) are considered False Negatives (FN); and all False Positives hypopneas are considered True Negatives (TN). Part (b) shows the sensitivity and specificity considering apnea and hypopnea as indistinctive events. Based on this all False Classifications are re-classified as True Positives (TP); and all False Positives hypopneas are False Positives (FP). Part (c) details the number of epochs corresponding to a particular classification.

Sleep data

For the sleep data 62,727 10 second epochs were analyzed in total. 34 true apnea epochs and 40 true hypopnea epochs were identified for the controls (36 and 37 if the clinician scorer was considered to be the gold standard); and 312 apnea epochs and 181 hypopnea epochs for the patients (342 and 200 if the clinician scorer was considered to be the gold standard). The average number of apnea epochs for the patient group throughout the night was 32. All patients had episodes of apnea or hypopnea. There was only two patients who did not have any episode of apnea. For one control, only 3·2 hours of data were recorded,

because of an ICT error. For one patient, only 3 hours were analyzed as more than one SOMNO sensor including the nasal cannula and the pulse oximeters detached prematurely. The results in terms of sensitivity and specificity, for the control group, patient group and overall are presented in Table 2. Table 2 is divided in four parts: the first and second evaluate the performance for apnea and apnea/hypopnea combined detection respectively without assuming a gold standard (i.e. the consensus of the majority determines a true event); and the third and fourth parts present the same evaluation but considering the expert as the gold standard.

Apnea detection (% sensitivity and specificity)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**47·1**	**97·1**	**94·1**	**99·3**	**99·8**	**100**
(95 CI)	(30.3-63.8)	(91.4-100)	(86.2-100)	(99.2-99.3)	(99.7/99.8)	(100-100)
Patients	**14·7**	**99·4**	**98·1**	**99·5**	**99·5**	**99·9**
(95 CI)	(10.8-18.7)	(98.5-100)	(96.6-99.6)	(99.5-99.6)	(99.4-99.6)	(99.8-99.9)
All	**17·9**	**99·1**	**97·7**	**99·4**	**99·7**	**99·9**
(95 CI)	(13.9-22.0)	(98.2/100)	(96.1-99.3)	(99.3-99.4)	(99.6-99.7)	(99.9-100)

Apnea and Hypopnea combined detection (%)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**87·8**	**58·1**	**94·6**	**98·6**	**99·7**	**100**
(95 CI)	(80.4-95.3)	(46.9-69.4)	(89.4-99.8)	(98.5-98.7)	(99.6-99.7)	(100-100)
Patients	**53·3**	**88·2**	**98·8**	**97·9**	**99·5**	**99·8**
(95 CI)	(48.9-57.8)	(85.4-91.1)	(97.8-99.8)	(97.7-98.1)	(99.4-99.6)	(99.8-99.9)
All	**57·8**	**84·1**	**98·2**	**98·4**	**99·5**	**99·9**
(95 CI)	(53.8-61.9)	(81.1-87.1)	(97.2-99.3)	(98.3-98.5)	(99.5-99.6)	(99.9-100)

Apnea Detection with the clinician scorer as Gold Standard reference (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**38·9**	**86·1**	**99·2**	**99·7**
(95 CI)	(23.0-54.8)	(74.8-97.4)	(99.1-99.3)	(99.7-99.8)
Patients	**11·7**	**88·9**	**99·5**	**99·4**
(95 CI)	(8.3-15.1)	(85.6-92.2)	(99.4-99.6)	(99.3-99.5)
All	**14·3**	**88·6**	**99·3**	**99·6**
(95 CI)	(10.8-17.8)	(85.4-91.8)	(99.2-99.4)	(99.6-99.7)

Apnea and Hypopnea combined detection with clinician scorer as Gold Standard (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**86·3**	**54·8**	**98·6**	**99·6**
(95 CI)	(63/73)	(40/73)	(41539/42139)	(41987/42139)
Patients	**49·6**	**80·1**	**98·4**	**100**
(95 CI)	(45.4-53.8)	(75.2-82.1)	(98.2-98.6)	(99.9-100)
All	**54·0**	**77·1**	**98·5**	**99·7**
(95 CI)	(50.0-57.9)	(73.8-80.5)	(98.4-98.6)	(99.7-99.8)

Table 2: Summary of performance for the WADD, SOMNO and clinician scorer systems for detection of apnea and hypopnea in 15 second epochs of overnight recordings.

The WADD also had the added feature of being able to differentiate between central and obstructive apnea. 90% of the central apneas were rightly marked as central. 96% were rightly marked as obstructive. Approximately 60% of the total apneas were obstructive in origin.

Device comfort

After the overnight study, the devices were detached and the subjects scored the comfort of the devices and quality of sleep (rating 1 to 5, with 5 representing maximum comfort and quality). Skin irritation caused by the WADD's adhesive was also rated from 1 to 5 (5 representing no irritation, 4 mild transient, redness, and 1 severe irritation). The median rating for WADD comfort was 5 (range 4-5). The median rating for SOMNO comfort was 3 (range 1-5 for controls and 2-5 for patients). The median rating for irritation caused by the WADD plaster on the neck was 5 (range 5-5 for controls and 4-5 for patients).

DISCUSSION

Main findings

WADD had very high sensitivity and specificity for detecting apnea in 15 second epochs in a series of breathing and breath-holding exercises in a variety of conditions, including the presence of external background noise, movement and posture. The tolerability of WADD was superior to the portable polysomnography system (SOMNO) during overnight recordings.

WADD had 97·7-99·2% sensitivity to detect instructed apneas and 88.6-99·1% for 10 seconds spontaneous apneas during natural sleep, with similar performance in controls and patients. The WADD also detected all apneas over 30 seconds and there were only 3 over 30 seconds false positives. For short apneas, in most cases, disagreement between the clinician scorer and the WADD were caused by the WADD identifying as apnea epochs that the expert classified as hypopneas.

As expected, the WADD performance was less good when apneas and hypopneas were considered together (minimum sensitivity 77·1%).This is not surprising since the WADD was designed to identify apnea, not hypopneas, and the latter were detected from the transmitted signal which had already been pre-processed for apnea detection. From the table, it can be observed that the degradation of performance was more evident in the controls because the controls had a large number of shorter hypopneas (under 22.5 seconds) which the WADD did not detect properly. In the patients, who often demonstrated apneas, the hypopnea events were longer and these were detected by WADD. Although the

lower sensitivity in hypopnea detection might in principle seem problematic if the WADD was to be used in the context of sleep apnea diagnosis (hypopneas are very common events in sleep labs), it is worth noting that: 1) there is no other reported automatic system that gets anywhere close to this with similar specificity and apnea detection performance; 2) the variations between different sleep labs due to the non-uniform definition of hypopneas already leads to much larger diagnostic variations than the limitation in sensitivity of the WADD;[35-37] 3) assuming the worse case scenario for the WADD, this is that a patient only had hypopneas throughout the night, this reduced sensitivity would be a problem that would translate to non-diagnosis of sleep apnea for patients that with 100% sensitivity would have had a sleep apnea hypopnea index (AHI) between 5 and 6 (i.e very mild cases of sleep apnea). Patients with no sleep apnea, moderate sleep apnea, severe sleep apnea and those with mild sleep apnea with AHI between 6 and 15 would have been rightly diagnosed.

The median difference between the WADD calculated Apnea Hypopnea Index (AHI) and the one obtained by the gold standard was 0 (average=0.7).

SOMNO Performance

The automatic analysis of the SOMNO apparatus, an FDA approved and clinically accepted system, based on assessing apnea from a variety of different sensors, significantly differed from that of the expert marker, with an average sensitivity value of around 14%. The results obtained from the instructed apneas tests also showed that even in the absence of artifacts, apneas were not well detected by the SOMNO system, with an average sensitivity of 37·8%. This demonstrates the need for caution if relying on current automated assessment methods for diagnosing apnea. Whilst performance might be improved by optimizing parameters for individual patients, this is not practical for single overnight recordings or use

as an alerting monitor. The WADD does not require any parameter optimization or subject specific calibration.

The SOMNO system performance improved in the event of indistinct classification of apneas and hypopneas, but was still poor compared with the clinician scorer (54% sensitivity). This sensitivity was at the cost of reduced specificity: for every true hypopnea detected there were approximately four false detections. Overall, the performance of the WADD in hypopnea/apnea combined detection was significantly better than the SOMNO, in sensitivity (77·1% vs 54% if considering the clinician scorer as a gold standard, and 84·1% vs 57·8% otherwise), but also in specificity, as the WADD only detected one false hypopnea epoch for every four true events.

Limitations. Future improvements

The study described in this paper is a small pilot study and hence further more comprehensive clinical evaluation of the technology will be necessary before it can be used. The size of the study was however adequate to assess the potential of the technology; to determine whether the initial performance results in controlled conditions were equivalent to those obtained in real scenarios; and to inform a clinical trial. Based on these positive results it is expected that a fully powered clinical trial, focused on diagnosis rather than on individual event identification, will follow in the future.

The calculation of the sensitivity and specificity has assumed that all apnea events were independent, which for some events might not be completely correct. Nonetheless it was observed that the characteristics of the breathing signal changed as much within the same subject (depending on timing, position, external artefacts, etc.) than between different

subjects. A different statistical analysis, possibly comparing pooled with non-pooled data will be the subject of investigation when the technology undergoes a larger clinical trial.

The WADD is obviously no substitute to a full night study in a sleep clinic, since it does not provide all the information that a full polysomnography system would. There are advantages and disadvantages to this device with respect to full polysomnography. The WADD can be used to determine the Apnea Hypopnea Index (AHI), which is used in sleep apnea diagnosis to ascertain whether a patient has sleep apnea and to score the severity of the condition. The main advantage is that it can be used for at home assessment or monitoring, and from that point of view it is clearly superior to any of the other existing devices (highly resilient to artifacts, very easy to attach and durable in position, low cost, much more comfortable, and accurate). Considering the restricted resources for sleep clinic referral this device could be a very useful tool to determine at very low cost who should be referred to a specialist centre for full polysomnography. The disadvantage is that there are other parameters that could be used for extra assessment that the device does not measure, such as microarousal or full cardiac activity. Furthermore, the WADD does not allow to assess the hypoxic load or autonomic activation and therefore impact the cardiovascular or stroke risk associated with OSA syndrome.

The WADD device used in the current study relied on wireless transmission to a PC. However changing the PC to a dedicated mobile phone sized receiver poses no technological challenge. A subsequent version that is being developed is smaller (2.4 by 2.4 by 1.2 cm, weighing 7.5 grams) and can operate continuously on hearing aid batteries for over 48 hours. It has a separate dedicated receiver of comparable size to a mobile phone which can be located up to 10 metres from the subject.

Acknowledgments: Eduardo Aguilar-Pelaez for assisting with the data collection at the initial phases of the clinical study and proof reading one of the versions of the manuscript; Pauline Rogers (Statistician, UCL/H Biomedical Research Unit) for carrying out the statistical analysis leading to the number of tested subjects; UCL Business; Epilepsy Society; and Staff on MITU at NHNN.

Sources of Support. UCL Business and Imperial Innovations.

Authors' contributions: John Duncan and Esther Rodriguez- Villegas had the initial idea of developing a wearable breathing monitoring device. Esther Rodriguez-Villegas and Guangwei Chen were the main creators of the novel device carrying out the electronic design which includes both hardware and software. John Duncan contributed through the process helping to define the engineering specifications to meet the clinical need. John Duncan designed the protocol for the clinical study. Esther Rodriguez-Villegas created the documentation for MHRA approval. Jeremy Radcliffe led the clinical studies and blindly marked all the data using the sensors of the SOMNO system as well as the Konica Minolta pulse oximeter. Esther Rodriguez-Villegas and Guangwei Chen compared the results of the data scored by the WADD, SOMNO and blinded expert, and carried out the data analysis. All the authors contributed to the interpretation of data. John Duncan, Jeremy Radcliffe and Esther Rodriguez- Villegas wrote the paper.

Conflict of interest: There are no competing interests.

Data sharing: No additional data available

References

1 Shorvon S, Tomnson T. Sudden unexpected death in epilepsy. *Lancet* 2011;378(9808):2028–38.

2 Mace SE. Acute bacterial meningitis. *Emerg Med Clin North Am* 2008;**26**(2):281–317.

3 Cordero L, Rau R, Taylor D, et al. Enteric gram-negative bacilli bloodstream infections: 17 years' experience in a neonatal intensive care unit. *Am J Infect Control* 2004;**32**(4):189–95.

4 Powers, KS. Acute pulmonary infections. In: Lucking SE, Maffei, FA, Tamburro, RF and Thomas, NJ, editors. Pediatric critical care study guide. London: Springer 2012; 514–534.

5 Moruzzi P, Sarzi-Braga S, Rossi M, et al. Sleep apnea in ischaemic heart disease: differences between acute and chronic coronary syndromes. *Heart* 1999;**82**(3):343–7.

6 Van Vliet G, Deal CL, Crock PA, et al. Sudden death in growth hormone-treated children with Prader-Willi syndrome. *J Pediatr* 2004;**144**(1):129–31.

7 Taylor MB, Grounds RM, Mulrooney PD, et al. Ventilatory effects of propofol during induction of anaesthesia. *Anaesthesia* 1986;**41**(8):816–20.

8 Murphy JJ, Swanson T, Ansermino M, et al. The frequency of apneas in premature infants after inguinal hernia repair: do they need overnight monitoring in the intensive care unit? *J Pediatr Surg* 2008;**43**(5):865-8.

9 Naeem N, Montenegro H. Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation* 2005;**67**(1): 13-23.

10 Moon RY, Horne RS, Hauck FR. Sudden infant death syndrome. *Lancet* 2007;**370**(9598):1578–87.

11 Gami AS, Howard DE, Olson EJ, et al. Day-night pattern of sudden death in obstructive sleep apnea, *N Engl J Med* 2005;**352**(12):1206–14.

12 Julien JY, Martin JG, Ernst P, et al. Prevalence of obstructive sleep apnea–hypopnea in severe versus moderate asthma. *J Allergy Clin Immunol* 2009;**124**(2):371–6.

13 Emilsson O, Janson C, Benediktsdóttir B,et al. Nocturnal gastroesophageal reflux, lung function and symptoms of obstructive sleep apnea: Results from an epidemiological survey. *Respir Med* 2012;**106**(3):459–66.

14 Reichmuth K, Austin D, Skatrud JB, et al. Association of sleep apnea and type II diabetes. A population-based study. *Am J Respir Crit Care Med* 2005;**172**(12):1590–5.

15 Hanly PJ. Mechanisms and management of central sleep apnea. *Lung* 1992;**170**(1):1–17.

16 Culebras A. Sleep and neuromuscular disorders. *Neurol Clin* 2005;**23**(4):1209–23.

17 Young T, Palta M, Dempsey J, et al. The occurrence of sleep disordered breathing among middle aged adults. *N Engl J Med* 1993;**328**(17):1230–5.

18 Faroux B. What's new in pediatric sleep? *Pediatr Respir Rev* 2007;**8**(1):85–9

19 Pang KP, Terris DJ. Screening for obstructive sleep apnea; an evidence-based analysis. *Am J Otolaryngol* 2006;**27**(2):112–8.

20 Redline S, Quan SF. Sleep apnea: a common mechanism for the deadly triad--cardiovascular disease, diabetes, and cancer? *Am J Respir Crit Care Med* 2012;**186**(2):123–124.

21 Hirshkowitz M. The clinical consequences of obstructive sleep apnea and associated excessive sleepiness. *J Fam Pract* 2008;**57**(8 Suppl):S9–16.

22 Wittmann V, Rodenstein DO. Health care costs and the sleep apnea syndrome. *Sleep Medicine Rev* 2004;**8**(4):269–279

23 Smith R, Ronald J, Delaive K, et al. What are obstructive sleep apnea patients being treated for prior to this diagnosis? *Chest* 2002;**12**(1):164-172.

24 Leger D, Bayon V, Laaban JP, et al. Impact of sleep apnea on economics. *Sleep Med Rev* 2012;**16**(5):455–62.

25 Flemons WW, Douglas NJ, Kuna ST, et al. Access to diagnosis and treatment of patients with suspected sleep apnea. *Am J Respir Crit Care Med* 2004;**169**(6):668–72.

26 Oliveira MG, Nery LE, Santos-Silva R, et al. Is portable monitoring accurate in the diagnosis of obstructive sleep apnea syndrome in chronic pulmonary obstructive disease? *Sleep Med*. 2012;**13**(8):1033-8.

27 Flemons WW, Littner MR; Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest* 2003;**124**(4):1543–79

28 McDonald F. Avoidable Epilepsy-Related Deaths: The case for prevention. *Epilepsy Bereaved Report*, 2011.

29 Surges R, Sander JW. Sudden unexpected death in epilepsy: mechanisms, prevalence and prevention. *Curr Opin Neurol* 2012;**2**:201–207.

30 Lofsky A. Sleep apnea and narcotic postoperative pain medication: a morbidity and mortality risk. *Anaesthesia Patient Safety Foundation Newsletter*. Summer 2002;**17**:24–25.

31 Bhananker SM, Posner KL, Cheney FW, et al. Injury and liability associated with monitored anesthesia care: A closed claims analysis. *Anesthesiology* 2006;**104**:228–34.

32 Rodriguez-Villegas E, Corbishley P, Duncan JS, et al. A method and apparatus for monitoring respiratory activity. Patent *WIPO;* WO/2007/099314.

33 Corbishley PG, Rodriguez-Villegas E. Breathing detection: towards a miniaturized wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering* 2008; **55**(1):196–204.

34 Rodriguez-Villegas E, Chen G, Aguilar-Pelaez E. Feature characterization for breathing monitor. Patent *WIPO*; WO/2012/007719.

35 Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. *Medicina* 2003;**63**(2):119–24.

36 Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med* 2000;**161** (2):369–74.

37 Ruehland WR, Rochford PD, O'Donoghue FJ, et al. The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index, *Sleep* 2009;**32**(2):150–157.

Figure Legends:

Figure 1: (a) WADD worn by one of the investigators. (b) Subject wearing an existing state of the art ambulatory apnea monitoring system (SOMNO), comprising finger oximetry; oro-nasal flow sensors; thoracic and abdominal expansion bands; and ECG.

Figure 2: Illustration of the SOMNO and WADD output signals showing an apnea event: (top) raw signals from the different SOMNO sensors, (middle) processed WADD signal, (bottom) WADD output signal.

A PILOT STUDY OF A WEARABLE APNEA DETECTION DEVICE

Rodriguez-Villegas Esther[1], PhD, SMIEEE; Gwangwei Chen[1]; Radcliffe Jeremy[2], FRCA, Duncan John[3], FRCP

[1]Dept of Electrical and Electronic Engineering,
Imperial College London, Exhibition Road
London SW7 2AZ, UK

[2]Dept of Neuroanaesthesia
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

[3]Dept of Clinical and Experimental Epilepsy
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

Corresponding author:
Esther Rodriguez-Villegas
Circuits and Systems
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road
London, SW7 2AZ, UK
e.rodriguez@imperial.ac.uk, +4475946261, Fax: +44 207 581 4419

Keywords: Apnea, Breathing, Low Power Electronics, Medical Devices, Physiological Monitoring

Word Count. 3500

ABSTRACT

Rationale. Current techniques for monitoring patients for apnea suffer from significant limitations. These include insufficient availability to meet diagnostic needs, cost, accuracy of results in the presence of artifacts, or difficulty of use in unsupervised conditions.

Objectives. We created and clinically tested a novel miniature medical device that targets to overcome these limitations.

Methods. We studied 20 healthy control subjects and 10 sleep apnea patients. The performance of the new system and also of the FDA approved SOMNO clinical system, conventionally used for sleep apnea diagnosis was evaluated in the same conditions. Both systems were tested during a normal night of sleep in both controls and patients. Their performance was quantified in terms of detection of apnea and hypopnea in individual 10 second epochs, which were compared with scoring of signals by a blinded clinician.

Main Results. For spontaneous apneas during natural sleep and considering the clinician scorer as the gold standard the new wearable apnea detection device had 88.6% sensitivity and 99.6% specificity. In comparison the SOMNO system had 14·3% sensitivity and 99·3% specificity. The novel device had been specifically designed to detect apnea, but if both apnea and hypopneas during sleep were considered in the assessment, the sensitivity and specificity were 77·1% and 99·7% respectively; versus 54% and 98.5% for the SOMNO.

Conclusions. The performance of the novel device compares very well to the scoring by an experienced clinician even in the presence of breathing artifacts, in this small pilot study. This can potentially make it a real solution for apnea home monitoring.

Word Count: 259

SUMMARY

Strenghts:

- We present the smallest, least intrusive technology to automatically detect apneas/hypopneas
- Performance characterization in normal signal conditions and with signal artefacts, showing excellent agreement with expert- 60,000 epochs assessed in controls and patients
- Sensitivity six times better than a state of the art commercial system, and excellent scoring in terms of user acceptance.

Limitations:

- The size of the study is limited. This is however justified by the fact that it was an initial pilot study to prove the strength of this novel technology to detect individual events even in the presence of artefacts (study goals of 95% confidence intervals for sensitivity and specificity values).

- The technology is still not optimized for hypopnea detection.

INTRODUCTION

Apnea may occur acutely in the context of infectious, respiratory, cardiac and neurological disease;[1-5] be caused by medication;[6-8] and on occasion death may be averted with urgent intervention.[1, 6, 9-11] Apnea may also occur recurrently either as a co-morbidity in chronic conditions including asthma, gastro-oesophageal reflux, neuromuscular disorders and diabetes;[12-16] or on its own in sleep apnea syndrome.[17-27]

The importance of monitoring and quantifying apneas is widely acknowledged. Apneas are one of the two leading causes of Sudden Death in Epilepsy (SUDEP), which only in the UK affects more people than cot death and AIDS together.[28-29] Apneas are also known to be a major problem due to its potentially disastrous consequences in anaesthesia recovery rooms.[30-31] And just sleep apnea may affect between 2% and 10% of the adult population [24] and 1% to 3% of the pediatric population,[18] and is heavily under-diagnosed.[19] The indirect medical costs of under diagnosed adult patients, in the years preceding the diagnosis, is estimated to increase by up to a two-fold, even after correcting for chronic disease status.[22-23] This, added to the potential social consequences, in the form of accidents, increased morbidity and impact on work efficiency makes the condition a major public health issue.[24]

Currently existing techniques for monitoring and quantification of apneas are not satisfactory. In sleep apnea diagnosis, polysomnography is the gold standard but the lack of sleep labs, sleep specialists and the associated cost, either make it difficult for the family physician to confirm the suspicion, or delays diagnosis.[25] The importance of the problem has led Medicare and Medicaid in the USA to recently authorize payment of treatment for adults diagnosed with unattended home sleep monitoring devices.[24] Unfortunately

existing home monitoring devices suffer from one or several of the following limitations: the sensors can be difficult to place resulting in invalid recordings, they still require considerable specialist time in order to interpret the results, automatic interpretation is very inaccurate mainly due to the inability to deal with artifacts, sensors can be cumbersome or intrusive so affecting the quality of sleep.[26-27] Furthermore, there is no portable apnea monitoring system that can detect apneas with high enough sensitivity and specificity in real time to potentially be used to alert carers of life-threatening situations due to acute apnea that can occur in the context of other clinical scenarios such as epilepsy or in anaesthesia. In these scenarios also, the alternative of relying on devices that might be able to detect the sequelae of apnea (for example pulse oximeters to detect hypoxeamia or heart rate monitors to detect bradycardia) might result in fatal consequences due to a delayed response to the apnea.

We present the results of the first clinical study of a new wearable apnea detection device (WADD) specifically designed to overcome the limitations of all other existing technologies.

METHODS

Device

We determined that the strongest externally detectable signal related to breathing corresponded to turbulence in the trachea. This signal was detected with a customized acoustic chamber that optimized the signal transmission. The signal detected by the sensor has components corresponding to both the wanted “signal” (breathing) and undesired “noise” caused by artifacts (cardiac signal, external noise (eg speech, music, wind), movement causing rubbing against the sensor and electromagnetic interference). A novel signal processing algorithm was developed to differentiate “signal” from “noise”. The

algorithm evaluates both the temporal and frequency characteristics of the signal obtained from the sensor. Over 15 different features are analyzed with parametric functions that dynamically adapt over time, to compensate for changes in both the subject and the environment. No pre-calibration or subject specific knowledge or modification is needed for the sensor or the signal processing algorithm.[32-34] Part of the algorithm was implemented on hardware and incorporated into the sensor. This reduces the amount of data that is needed for wireless transmission and consequently the amount of power required from the battery; hence the small size of the device. The wireless receiver and the remaining part of the algorithm were run on a laptop computer.

The WADD was wireless, measured 3·74 by 2·4 by 2·1 cm, weighed 17 grams, and was fixed to the skin on the neck with hydrocolloid colostomy adhesive patches of approximately 4cm diameter (Boots). The preferred location was over the trachea, halfway between the lower margin of the thyroid cartilage and the supra-sternal notch (Figure 1(a)). If the skin in that location was loose, as was common in subjects over 40 years of age, the device was placed antero-laterally, anterior to the sternomastoid muscle. The device was left in place overnight, for approximately 14 hours.

Participants

The study was conducted in a sleep study room of the National Hospital for Neurology and Neurosurgery (UK). We studied 20 healthy controls and 10 patients, as they were sequentially admitted for diagnostic monitoring of sleep-related disorders of breathing, because these were likely to have spontaneous apnea events. The patients and controls were not matched. Patients also had a variety of neurological conditions, including epilepsy, dementia, neuropathy and motor neuron disease. The reasons to recruit patients who had

been referred for diagnosis of possible sleep-related disorders of breathing, as opposed to those who had been already diagnosed, were twofold. First, the purpose of this study was not to evaluate the WADD for sleep apnea diagnosis, but rather to evaluate its ability to detect individual events, both during controlled conditions to assess the robustness to artifact rejection, and during spontaneous sleep. Good performance on individual event identification would however be expected to translate in a good performance in the context of the different clinical applications. Secondly, non-diagnosed patients were recruited because studying diagnosed individuals would have involved either delay or interruption of their treatment. The decision on the number of patients was based on obtaining a large enough number of events that would lead to the study goals of 95% confidence intervals for sensitivity and specificity values (based on clinical experience on the minimum number of apnea events which would be expected per subject referred for sleep apnea diagnosis, per night). A larger number of controls were included to be able to assess specificity amongst those who were most likely to be disease free, and also in the presence of artifacts. The patient group comprised 2 females and 8 males with: a median age of 44·5 years of age (range 25-82); a median weight of 74 Kg (range 41-187); a median height of 177 cm (range 160-188); a median body mass index (BMI) of 23 Kg/cm^2 (range 17-61); and a median neck circumference of 40 cm (range 30-43). The control group comprised 3 females and 17 males with: a median age of 33·5 years of age (range 23-63); a median weight of 81.5 Kg (range 60-120); a median height of 176 cm (range 145-185); a median body mass index (BMI) of 26·5 Kg/cm2 (range 20-36); and a median neck circumference of 38 cm (range 34-48). Overall 40% of the subjects were overweight and 24% were obese. The study was approved by the Medicine and Healthcare products Regulatory Agency (MHRA) and the Research Ethics Committee of the UK National Hospital for Neurology and Neurosurgery.

Procedure

All subjects also had simultaneous, clinically standard respiratory monitoring comprising: finger oximetry; oro-nasal airflow sensors; thoracic and abdominal expansion bands; and ECG; using the SOMNO polysomnography system (SOMNOscreen ™ RC kombi . SOMNO Medics, Germany)- Figure 1 (b). Additionally, to further facilitate expert interpretation of polysomnography data, a second pulse oximeter (Pulsox-300i, Konica Minolta, Japan) was attached to the free hand. After attachment of the WADD and the SOMNOmedics polysomnography system, controls subjects participated in a series of exercises, comprising:

T1. Normal breathing for 5 minutes.

T2. Shallow breathing for 5 minutes.

T3. Normal breathing for 45 secs alternating with 15 secs instructed breath holds for 5 min.

T4. Normal breathing for 30 secs alternating with 30 secs instructed breath holds for 10min.

T5. As in 4 but with loud music in the background.

T6. Normal breathing while walking for 5 minutes.

T7. Normal breathing for 30 secs alternating with 30 secs instructed breath holds while lying prone for 5 min.

These exercises were designed to be representative of the worse case of artifact situations affecting the WADD following previous, very exhaustive, lab based research and testing. Following the exercises subjects were allowed to prepare for sleep and were left undisturbed overnight.

Data analysis

The breathing exercises data were analyzed by the automated WADD software and the automated SOMNO software. Instructed apneas were considered to be the "true events". The last six hours of sleep were blindly analyzed by: the automated WADD software, the automated SOMNO software, and by the experienced clinician who reviewed the raw signals from all SOMNO sensors, and had no knowledge about how WADD had been designed or worked. The reason to evaluate the last six hours of sleep was to try to keep the same amount of sleep data in as many subjects as possible in order to prevent biasing of the results. The pulse oximeter was also used by the clinician to support the diagnostic decisions and also individual event classification mostly in those cases in which the signals from the other SOMNO pulse oximeter was corrupted by artifacts. After the separate classification of WADD and SOMNO data, a further investigator compared the results.

The breathing exercises data were analyzed in 15 seconds epochs because this was the shortest duration of an instructed apnea. The sleep data was analyzed in 10 second epochs.

Two assessments were carried out of the sleep data. In the first assessment there was no pre-assumption of a gold standard, and the three systems (WADD, SOMNO and expert marker) were put under test and treated indistinctively. An epoch would be classified as true positive apnea or true positive hypopnea if at least two out of the three systems concurred on the classification. In the second assessment the final classification of these epochs would be that of the expert market, or in other words the expert marker was considered to be the gold standard deciding, and the performance of both SOMNO and

WADD system was evaluated. The SOMNO was evaluated as well as the WADD, as there is little or no quantitative information about the accuracy of automated polysomnography systems.

In both assessments epochs could be classified as:

a) True Positive Apnea (cessation of breathing signal, with correspondent absence of respiratory airflow)
b) True Positive Hypopnea (over 50% reduction in oronasal signal and in thoracoabdominal movement together with over 2% decrease in oxygen saturation).
c) False Positive Hypopneas (if a system had classified a breathing epoch as a hypopnea).
d) False Positive Apnea (if a system had classified a breathing epoch as an apnea)
e) False Classification Apnea as Hypopnea (if a system had classified an apnea epoch as hypopnea).
f) False Classification Hypopnea as Apnea (if a system had classified hypopnea as apnea).
g) False Negative Apnea (if a system classifies an apnea as breathing).
h) False Negative Hypopnea (if a system classifies a hypopnea as breathing).

The breathing exercises data were analyzed in the same way, but the instructed apneas and breathing sections were considered the absolute truth and hence there was no independent expert review.

The performance of the three systems was evaluated using the following metrics:

Sensitivity=(TP)/(TP+FN)

Specificity= (FP)/(TN+FP)

(TP=True Positive, TN=True Negative, FP=False Positive,FN=False Negative).

For each one of the two assessments (i.e. not presuming a gold standard, and considering the expert to be the gold standard), two different analyses were carried out. Firstly only apneas were considered to be true positives. Hence any hypopnea would be regarded as breathing (true negative); False Classification of Hypopneas as Apneas were re-classified as false positives; and False classification of Apneas as Hypopneas were re-classified as false negatives. Secondly, apneas and hypopneas were considered indistinctively, and hence true events of both variety would be also considered together.

These two analyses were carried out as they would be relevant to different clinical scenarios. For example, high sensitivity for detecting apnea would be crucial for identification of sudden apnea if monitoring those with epilepsy; whereas for diagnosis of sleep-related breathing dysfunction, which generally relies on the Apnea Hypopnea Index, the differentiation between apnea and hypopnea might be clinically less important.

RESULTS

Breathing exercises data

Data were available in 3956 15 second epochs for the controls performing the breathing exercises (132 in total). Table 1 summarizes the performance of WADD and SOMNO in the seven breathing exercises. Figure 2 illustrates examples of the signals obtained from the different sensors. Table 1 is divided in three parts. Part (a) and (b) quantify performance considering different scenarios for wrongly classified hypopneas. Although the real events were apneas, both systems had the ability to indicate hypopneas too. This resulted in some real apnea and breathing epochs being wrongly marked as hypopneas. In order to account

for these, Table 1 (a) shows the sensitivity and specificity when only apneas are considered as events (i.e. hypopneas would be regarded as breathing). In part (b) of the table hypopneas and apneas are indistinctively considered. Part (c) illustrates the total number of epochs that fall into a specific classification for both systems. The combined sensitivity and specificity for all the exercises across all the subjects for the WADD was 97·7% and 99·6% (considering hypopneas as breathing); or 99·2% and 99·5% (considering hypopneas as events). With the same criteria the sensitivity and specificity for the SOMNO was only 37·8% sensitivity, 96·5% specificity; or 62·8% sensitivity, 90·5% specificity.

WADD versus Somno performance in instructed exercises with hypopneas NOT considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99.2
T2	NA	NA	100	90.6
T3	94·6	38	99	96·9
T4	98·9	38·8	99·7	94·5
T5	99·2	31·4	99·2	99·7
T6	NA	NA	100	96·5
T7	94·2	48·2	98·5	99
Total	**97·7**	**37·8**	**99·6**	**96·5**

(a)

WADD versus Somno performance in instructed exercises with hypopneas and apneas indistinctively considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99
T2	NA	NA	100	81.4
T3	96·7	66.3	99	87·5
T4	100	64·6	99·7	87·9
T5	99·2	59.2	99·2	95·1
T6	NA	NA	100	89·8
T7	99	64·4	97	93.4
Total	**99·2**	**62·8**	**99·5**	**90·5**

(b)

Summary of classification of the different epochs										
Exercise	**TP**		**TN**		**FC**		**FP**		**FN**	
	WADD	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**
T1	0	0	380	376	0	0	0	4	0	0
T2	0	0	360	293	0	0	0	67	0	0
T3	89	61	285	252	2	26	3	36	3	31
T4	356	230	363	320	4	92	1	44	0	126
T5	357	213	365	350	0	100	3	18	3	147
T6	0	0	400	359	0	0	0	41	0	0
T7	189	123	191	184	9	31	6	13	2	68
Total	**991**	**627**	**2344**	**2134**	**15**	**249**	**13**	**223**	**8**	**372**

(c)

Table 1: Summary of performance for the WADD and SOMNO across the seven breathing exercises (as detailed in Procedure). TP= true positive (apnea), TN=true negative (breathing), FC= false classification, FP= False Positive, FN=False Negative. Part (a) of the table shows the sensitivity and specificity not considering hypopneas as events (i.e. all hypopneas are considered breathing). Based on this all False Classifications, FC (apneas wrongly classified as hypopneas) are considered False Negatives (FN); and all False Positives hypopneas are considered True Negatives (TN). Part (b) shows the sensitivity and specificity considering apnea and hypopnea as indistinctive events. Based on this all False Classifications are re-classified as True Positives (TP); and all False Positives hypopneas are False Positives (FP). Part (c) details the number of epochs corresponding to a particular classification.

Sleep data

For the sleep data 62,727 10 second epochs were analyzed in total. 34 true apnea epochs and 40 true hypopnea epochs were identified for the controls (36 and 37 if the clinician scorer was considered to be the gold standard); and 312 apnea epochs and 181 hypopnea epochs for the patients (342 and 200 if the clinician scorer was considered to be the gold standard). The average number of apnea epochs for the patient group throughout the night was 32. All patients had episodes of apnea or hypopnea. There was only two patients who did not have any episode of apnea. For one control, only 3·2 hours of data were recorded,

because of an ICT error. For one patient, only 3 hours were analyzed as more than one SOMNO sensor including the nasal cannula and the pulse oximeters detached prematurely. The results in terms of sensitivity and specificity, for the control group, patient group and overall are presented in Table 2. Table 2 is divided in four parts: the first and second evaluate the performance for apnea and apnea/hypopnea combined detection respectively without assuming a gold standard (i.e. the consensus of the majority determines a true event); and the third and fourth parts present the same evaluation but considering the expert as the gold standard.

Apnea detection (% sensitivity and specificity)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**47·1**	**97·1**	**94·1**	**99·3**	**99·8**	**100**
(95 CI)	(30.3-63.8)	(91.4-100)	(86.2-100)	(99.2-99.3)	(99.7/99.8)	(100-100)
Patients	**14·7**	**99·4**	**98·1**	**99·5**	**99·5**	**99·9**
(95 CI)	(10.8-18.7)	(98.5-100)	(96.6-99.6)	(99.5-99.6)	(99.4-99.6)	(99.8-99.9)
All	**17·9**	**99·1**	**97·7**	**99·4**	**99·7**	**99·9**
(95 CI)	(13.9-22.0)	(98.2/100)	(96.1-99.3)	(99.3-99.4)	(99.6-99.7)	(99.9-100)

Apnea and Hypopnea combined detection (%)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**87·8**	**58·1**	**94·6**	**98·6**	**99·7**	**100**
(95 CI)	(80.4-95.3)	(46.9-69.4)	(89.4-99.8)	(98.5-98.7)	(99.6-99.7)	(100-100)
Patients	**53·3**	**88·2**	**98·8**	**97·9**	**99·5**	**99·8**
(95 CI)	(48.9-57.8)	(85.4-91.1)	(97.8-99.8)	(97.7-98.1)	(99.4-99.6)	(99.8-99.9)
All	**57·8**	**84·1**	**98·2**	**98·4**	**99·5**	**99·9**
(95 CI)	(53.8-61.9)	(81.1-87.1)	(97.2-99.3)	(98.3-98.5)	(99.5-99.6)	(99.9-100)

Apnea Detection with the clinician scorer as Gold Standard reference (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**38·9**	**86·1**	**99·2**	**99·7**
(95 CI)	(23.0-54.8)	(74.8-97.4)	(99.1-99.3)	(99.7-99.8)
Patients	**11·7**	**88·9**	**99·5**	**99·4**
(95 CI)	(8.3-15.1)	(85.6-92.2)	(99.4-99.6)	(99.3-99.5)
All	**14·3**	**88·6**	**99·3**	**99·6**
(95 CI)	(10.8-17.8)	(85.4-91.8)	(99.2-99.4)	(99.6-99.7)

Apnea and Hypopnea combined detection with clinician scorer as Gold Standard (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**86·3**	**54·8**	**98·6**	**99·6**
(95 CI)	(63/73)	(40/73)	(41539/42139)	(41987/42139)
Patients	**49·6**	**80·1**	**98·4**	**100**
(95 CI)	(45.4-53.8)	(75.2-82.1)	(98.2-98.6)	(99.9-100)
All	**54·0**	**77·1**	**98·5**	**99·7**
(95 CI)	(50.0-57.9)	(73.8-80.5)	(98.4-98.6)	(99.7-99.8)

Table 2: Summary of performance for the WADD, SOMNO and clinician scorer systems for detection of apnea and hypopnea in 15 second epochs of overnight recordings.

The WADD also had the added feature of being able to differentiate between central and obstructive apnea. 90% of the central apneas were rightly marked as central. 96% were rightly marked as obstructive. Approximately 60% of the total apneas were obstructive in origin.

Device comfort

After the overnight study, the devices were detached and the subjects scored the comfort of the devices and quality of sleep (rating 1 to 5, with 5 representing maximum comfort and quality). Skin irritation caused by the WADD's adhesive was also rated from 1 to 5 (5 representing no irritation, 4 mild transient, redness, and 1 severe irritation). The median rating for WADD comfort was 5 (range 4-5). The median rating for SOMNO comfort was 3 (range 1-5 for controls and 2-5 for patients). The median rating for irritation caused by the WADD plaster on the neck was 5 (range 5-5 for controls and 4-5 for patients).

DISCUSSION

Main findings

WADD had very high sensitivity and specificity for detecting apnea in 15 second epochs in a series of breathing and breath-holding exercises in a variety of conditions, including the presence of external background noise, movement and posture. The tolerability of WADD was superior to the portable polysomnography system (SOMNO) during overnight recordings.

WADD had 97·7-99·2% sensitivity to detect instructed apneas and 88.6-99·1% for 10 seconds spontaneous apneas during natural sleep, with similar performance in controls and patients. The WADD also detected all apneas over 30 seconds and there were only 3 over 30 seconds false positives. For short apneas, in most cases, disagreement between the clinician scorer and the WADD were caused by the WADD identifying as apnea epochs that the expert classified as hypopneas.

As expected, the WADD performance was less good when apneas and hypopneas were considered together (minimum sensitivity 77·1%).This is not surprising since the WADD was designed to identify apnea, not hypopneas, and the latter were detected from the transmitted signal which had already been pre-processed for apnea detection. From the table, it can be observed that the degradation of performance was more evident in the controls because the controls had a large number of shorter hypopneas (under 22.5 seconds) which the WADD did not detect properly. In the patients, who often demonstrated apneas, the hypopnea events were longer and these were detected by WADD. Although the

lower sensitivity in hypopnea detection might in principle seem problematic if the WADD was to be used in the context of sleep apnea diagnosis (hypopneas are very common events in sleep labs), it is worth noting that: 1) there is no other reported automatic system that gets anywhere close to this with similar specificity and apnea detection performance; 2) the variations between different sleep labs due to the non-uniform definition of hypopneas already leads to much larger diagnostic variations than the limitation in sensitivity of the WADD;[35-37] 3) assuming the worse case scenario for the WADD, this is that a patient only had hypopneas throughout the night, this reduced sensitivity would be a problem that would translate to non-diagnosis of sleep apnea for patients that with 100% sensitivity would have had a sleep apnea hypopnea index (AHI) between 5 and 6 (i.e very mild cases of sleep apnea). Patients with no sleep apnea, moderate sleep apnea, severe sleep apnea and those with mild sleep apnea with AHI between 6 and 15 would have been rightly diagnosed.

The median difference between the WADD calculated Apnea Hypopnea Index (AHI) and the one obtained by the gold standard was 0 (average=0.7).

SOMNO Performance

The automatic analysis of the SOMNO apparatus, an FDA approved and clinically accepted system, based on assessing apnea from a variety of different sensors, significantly differed from that of the expert marker, with an average sensitivity value of around 14%. The results obtained from the instructed apneas tests also showed that even in the absence of artifacts, apneas were not well detected by the SOMNO system, with an average sensitivity of 37·8%. This demonstrates the need for caution if relying on current automated assessment methods for diagnosing apnea. Whilst performance might be improved by optimizing parameters for individual patients, this is not practical for single overnight recordings or use

as an alerting monitor. The WADD does not require any parameter optimization or subject specific calibration.

The SOMNO system performance improved in the event of indistinct classification of apneas and hypopneas, but was still poor compared with the clinician scorer (54% sensitivity). This sensitivity was at the cost of reduced specificity: for every true hypopnea detected there were approximately four false detections. Overall, the performance of the WADD in hypopnea/apnea combined detection was significantly better than the SOMNO, in sensitivity (77·1% vs 54% if considering the clinician scorer as a gold standard, and 84·1% vs 57·8% otherwise), but also in specificity, as the WADD only detected one false hypopnea epoch for every four true events.

Limitations. Future improvements

The study described in this paper is a small pilot study and hence further more comprehensive clinical evaluation of the technology will be necessary before it can be used. The size of the study was however adequate to assess the potential of the technology; to determine whether the initial performance results in controlled conditions were equivalent to those obtained in real scenarios; and to inform a clinical trial. Based on these positive results it is expected that a fully powered clinical trial, focused on diagnosis rather than on individual event identification, will follow in the future.

The calculation of the sensitivity and specificity has assumed that all apnea events were independent, which for some events might not be completely correct. Nonetheless it was observed that the characteristics of the breathing signal changed as much within the same subject (depending on timing, position, external artefacts, etc.) than between different

subjects. A different statistical analysis, possibly comparing pooled with non-pooled data will be the subject of investigation when the technology undergoes a larger clinical trial.

The WADD is obviously no substitute to a full night study in a sleep clinic, since it does not provide all the information that a full polysomnography system would. There are advantages and disadvantages to this device with respect to full polysomnography. The WADD can be used to determine the Apnea Hypopnea Index (AHI), which is used in sleep apnea diagnosis to ascertain whether a patient has sleep apnea and to score the severity of the condition. The main advantage is that it can be used for at home assessment or monitoring, and from that point of view it is clearly superior to any of the other existing devices (highly resilient to artifacts, very easy to attach and durable in position, low cost, much more comfortable, and accurate). Considering the restricted resources for sleep clinic referral this device could be a very useful tool to determine at very low cost who should be referred to a specialist centre for full polysomnography. The disadvantage is that there are other parameters that could be used for extra assessment that the device does not measure, such as microarousal or full cardiac activity. Furthermore, the WADD does not allow to assess the hypoxic load or autonomic activation and therefore impact the cardiovascular or stroke risk associated with OSA syndrome.

The WADD device used in the current study relied on wireless transmission to a PC. However changing the PC to a dedicated mobile phone sized receiver poses no technological challenge. A subsequent version that is being developed is smaller (2.4 by 2.4 by 1.2 cm, weighing 7.5 grams) and can operate continuously on hearing aid batteries for over 48 hours. It has a separate dedicated receiver of comparable size to a mobile phone which can be located up to 10 metres from the subject.

Acknowledgments: Eduardo Aguilar-Pelaez for assisting with the data collection at the initial phases of the clinical study and proof reading one of the versions of the manuscript; Pauline Rogers (Statistician, UCL/H Biomedical Research Unit) for carrying out the statistical analysis leading to the number of tested subjects; UCL Business; Epilepsy Society; and Staff on MITU at NHNN.

Sources of Support. UCL Business and Imperial Innovations.

Authors' contributions: John Duncan and Esther Rodriguez- Villegas had the initial idea of developing a wearable breathing monitoring device. Esther Rodriguez-Villegas and Guangwei Chen were the main creators of the novel device carrying out the electronic design which includes both hardware and software. John Duncan contributed through the process helping to define the engineering specifications to meet the clinical need. John Duncan designed the protocol for the clinical study. Esther Rodriguez-Villegas created the documentation for MHRA approval. Jeremy Radcliffe led the clinical studies and blindly marked all the data using the sensors of the SOMNO system as well as the Konica Minolta pulse oximeter. Esther Rodriguez-Villegas and Guangwei Chen compared the results of the data scored by the WADD, SOMNO and blinded expert, and carried out the data analysis. All the authors contributed to the interpretation of data. John Duncan, Jeremy Radcliffe and Esther Rodriguez- Villegas wrote the paper.

Conflict of interest: There are no competing interests.

References

1 Shorvon S, Tomnson T. Sudden unexpected death in epilepsy. *Lancet* 2011;378(9808):2028–38.

2 Mace SE. Acute bacterial meningitis. *Emerg Med Clin North Am* 2008;**26**(2):281–317.

3 Cordero L, Rau R, Taylor D, et al. Enteric gram-negative bacilli bloodstream infections: 17 years' experience in a neonatal intensive care unit. *Am J Infect Control* 2004;**32**(4):189–95.

4 Powers, KS. Acute pulmonary infections. In: Lucking SE, Maffei, FA, Tamburro, RF and Thomas, NJ, editors. Pediatric critical care study guide. London: Springer 2012; 514–534.

5 Moruzzi P, Sarzi-Braga S, Rossi M, et al. Sleep apnea in ischaemic heart disease: differences between acute and chronic coronary syndromes. *Heart* 1999;**82**(3):343–7.

6 Van Vliet G, Deal CL, Crock PA, et al. Sudden death in growth hormone-treated children with Prader-Willi syndrome. *J Pediatr* 2004;**144**(1):129–31.

7 Taylor MB, Grounds RM, Mulrooney PD, et al. Ventilatory effects of propofol during induction of anaesthesia. *Anaesthesia* 1986;**41**(8):816–20.

8 Murphy JJ, Swanson T, Ansermino M, et al. The frequency of apneas in premature infants after inguinal hernia repair: do they need overnight monitoring in the intensive care unit? *J Pediatr Surg* 2008;**43**(5):865-8.

9 Naeem N, Montenegro H. Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation* 2005;**67**(1): 13-23.

10 Moon RY, Horne RS, Hauck FR. Sudden infant death syndrome. *Lancet* 2007;**370**(9598):1578–87.

11 Gami AS, Howard DE, Olson EJ, et al. Day-night pattern of sudden death in obstructive sleep apnea, *N Engl J Med* 2005;**352**(12):1206–14.

12 Julien JY, Martin JG, Ernst P, et al. Prevalence of obstructive sleep apnea–hypopnea in severe versus moderate asthma. *J Allergy Clin Immunol* 2009;**124**(2):371–6.

13 Emilsson O, Janson C, Benediktsdóttir B,et al. Nocturnal gastroesophageal reflux, lung function and symptoms of obstructive sleep apnea: Results from an epidemiological survey. *Respir Med* 2012;**106**(3):459–66.

14 Reichmuth K, Austin D, Skatrud JB, et al. Association of sleep apnea and type II diabetes. A population-based study. *Am J Respir Crit Care Med* 2005;**172**(12):1590–5.

15 Hanly PJ. Mechanisms and management of central sleep apnea. *Lung* 1992;**170**(1):1–17.

16 Culebras A. Sleep and neuromuscular disorders. *Neurol Clin* 2005;**23**(4):1209–23.

17 Young T, Palta M, Dempsey J, et al. The occurrence of sleep disordered breathing among middle aged adults. *N Engl J Med* 1993;**328**(17):1230–5.

18 Faroux B. What's new in pediatric sleep? *Pediatr Respir Rev* 2007;**8**(1):85–9

19 Pang KP, Terris DJ. Screening for obstructive sleep apnea; an evidence-based analysis. *Am J Otolaryngol* 2006;**27**(2):112–8.

20 Redline S, Quan SF. Sleep apnea: a common mechanism for the deadly triad--cardiovascular disease, diabetes, and cancer? *Am J Respir Crit Care Med* 2012;**186**(2):123–124.

21 Hirshkowitz M. The clinical consequences of obstructive sleep apnea and associated excessive sleepiness. *J Fam Pract* 2008;**57**(8 Suppl):S9–16.

22 Wittmann V, Rodenstein DO. Health care costs and the sleep apnea syndrome. *Sleep Medicine Rev* 2004;**8**(4):269–279

23 Smith R, Ronald J, Delaive K, et al. What are obstructive sleep apnea patients being treated for prior to this diagnosis? *Chest* 2002;**12**(1):164-172.

24 Leger D, Bayon V, Laaban JP, et al. Impact of sleep apnea on economics. *Sleep Med Rev* 2012;**16**(5):455–62.

25 Flemons WW, Douglas NJ, Kuna ST, et al. Access to diagnosis and treatment of patients with suspected sleep apnea. *Am J Respir Crit Care Med* 2004;**169**(6):668–72.

26 Oliveira MG, Nery LE, Santos-Silva R, et al. Is portable monitoring accurate in the diagnosis of obstructive sleep apnea syndrome in chronic pulmonary obstructive disease? *Sleep Med*. 2012;**13**(8):1033-8.

27 Flemons WW, Littner MR; Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest* 2003;**124**(4):1543–79

28 McDonald F. Avoidable Epilepsy-Related Deaths: The case for prevention. *Epilepsy Bereaved Report*, 2011.

29 Surges R, Sander JW. Sudden unexpected death in epilepsy: mechanisms, prevalence and prevention. *Curr Opin Neurol* 2012;**2**:201–207.

30 Lofsky A. Sleep apnea and narcotic postoperative pain medication: a morbidity and mortality risk. *Anaesthesia Patient Safety Foundation Newsletter*. Summer 2002;**17**:24–25.

31 Bhananker SM, Posner KL, Cheney FW, et al. Injury and liability associated with monitored anesthesia care: A closed claims analysis. *Anesthesiology* 2006;**104**:228–34.

32 Rodriguez-Villegas E, Corbishley P, Duncan JS, et al. A method and apparatus for monitoring respiratory activity. Patent *WIPO;* WO/2007/099314.

33 Corbishley PG, Rodriguez-Villegas E. Breathing detection: towards a miniaturized wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering* 2008; **55**(1):196–204.

34 Rodriguez-Villegas E, Chen G, Aguilar-Pelaez E. Feature characterization for breathing monitor. Patent *WIPO*; WO/2012/007719.

35 Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. *Medicina* 2003;**63**(2):119–24.

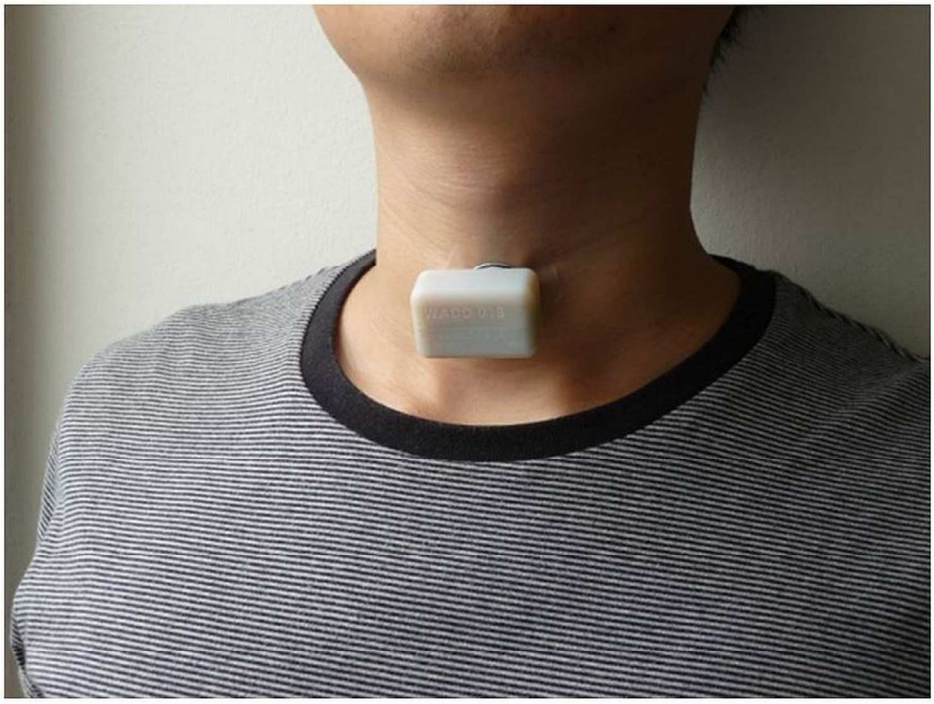
36 Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med* 2000;**161** (2):369–74.

37 Ruehland WR, Rochford PD, O'Donoghue FJ, et al. The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index, *Sleep* 2009;**32**(2):150–157.

Figure Legends:

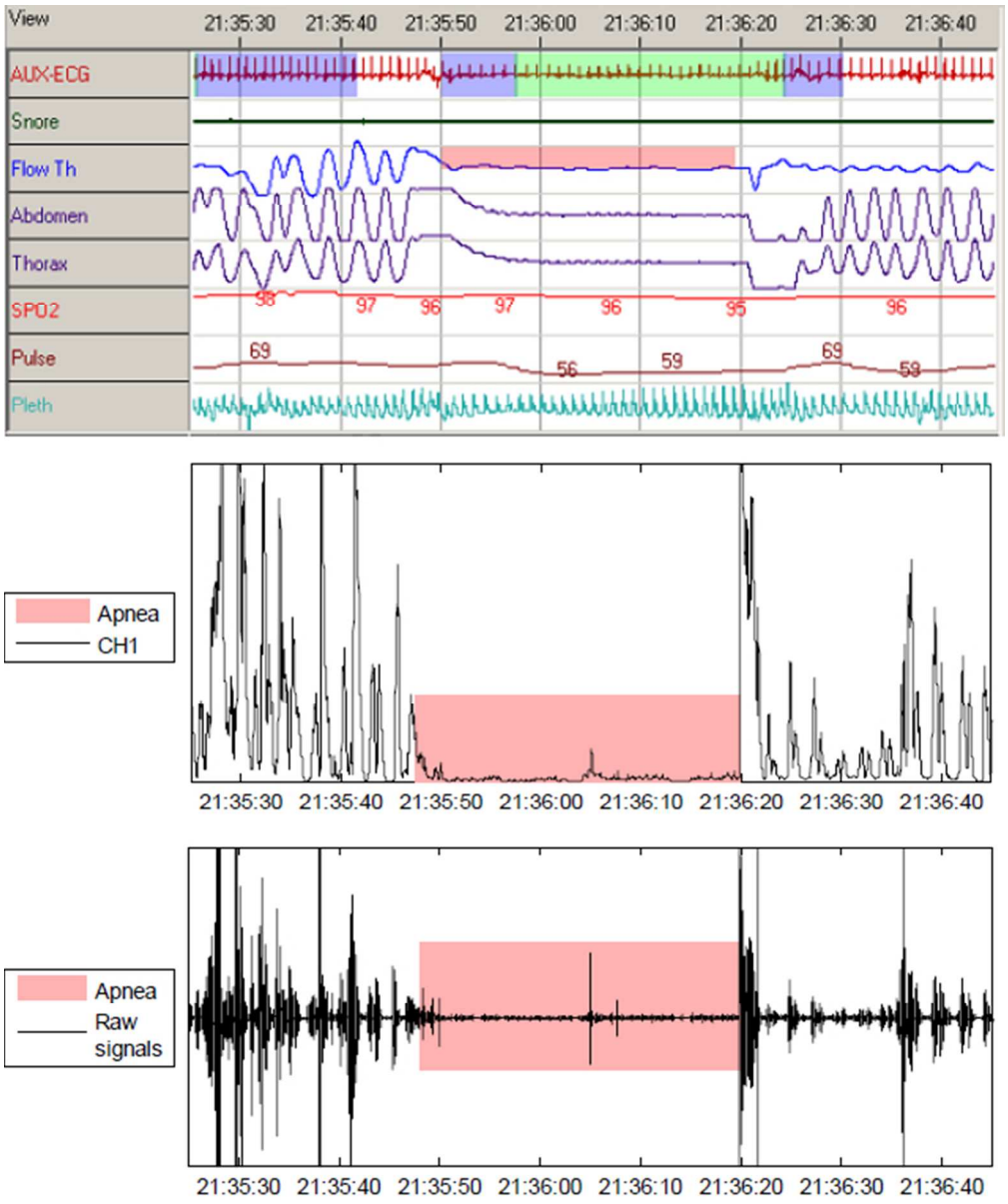
Figure 1: (a) WADD worn by one of the investigators. (b) Subject wearing an existing state of the art ambulatory apnea monitoring system (SOMNO), comprising finger oximetry; oro-nasal flow sensors; thoracic and abdominal expansion bands; and ECG.

Figure 2: Illustration of the SOMNO and WADD output signals showing an apnea event: (top) raw signals from the different SOMNO sensors, (middle) processed WADD signal, (bottom) WADD output signal.



90x69mm (300 x 300 DPI)

90x115mm (300 x 300 DPI)



53x64mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(included, page 1)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(included, page 1)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(included, pages 3-4)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(included, 3-4)**
Methods		
Study design	4	Present key elements of study design early in the paper **(included, pages 5-7)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(included, pages 5-6)**
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants **(included, pages 8-9)**
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case **(NA)**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(included, pages 8-9)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(included, pages 8-10)**
Bias	9	Describe any efforts to address potential sources of bias **(included, pages 8, 10)**
Study size	10	Explain how the study size was arrived at **(included, page 6)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(included, pages 8-10)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(included, pages 6, 14)**
		(*b*) Describe any methods used to examine subgroups and interactions **(NA)**
		(*c*) Explain how missing data were addressed **(page 12)**
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy **(NA)**
		(*e*) Describe any sensitivity analyses **(included)**

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(included, page 6)**
		(b) Give reasons for non-participation at each stage **(NA)**
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(included, page 6)**
		(b) Indicate number of participants with missing data for each variable of interest **(NA)**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) **(NA)**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time (NA)
		Case-control study—Report numbers in each exposure category, or summary measures of exposure **(NA)**
		Cross-sectional study—Report numbers of outcome events or summary measures**(NA)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **(included, page 14)**
		(*b*) Report category boundaries when continuous variables were categorized **(included, pages 8-10)**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **(NA)**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **(NA)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(included, pages 11-15)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(included, page 17)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(included, page 17)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(included, page 17)**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **(included, 19)**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

A Pilot Study of A Wearable Apnea Detection Device

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005299.R2
Article Type:	Research
Date Submitted by the Author:	22-Aug-2014
Complete List of Authors:	Rodriguez-Villegas, Esther; Imperial College London, Electrical and Electronic Engineering Chen, Guangwei; Imperial College London, Dept of Electrical and Electronic Engineering Radcliffe, Jeremy; UCL Institute of Neurology,UK National Hospital for Neurology and Neurosurgery, Dept of Neuroanaesthesia Duncan, John; UCL Institute of Neurology, UK National Hospital for Neurology and Neurosurgery, Dept of Clinical and Experimental Epilepsy
Primary Subject Heading:	Respiratory medicine
Secondary Subject Heading:	Diagnostics
Keywords:	RESPIRATORY MEDICINE (see Thoracic Medicine), Adult intensive & critical care < ANAESTHETICS, SLEEP MEDICINE

A PILOT STUDY OF A WEARABLE APNEA DETECTION DEVICE

Rodriguez-Villegas Esther[1], PhD, SMIEEE; Gwangwei Chen[1]; Radcliffe Jeremy[2], FRCA, Duncan John[3], FRCP

[1]Dept of Electrical and Electronic Engineering,
Imperial College London, Exhibition Road
London SW7 2AZ, UK

[2]Dept of Neuroanaesthesia
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

[3]Dept of Clinical and Experimental Epilepsy
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

Corresponding author:
Esther Rodriguez-Villegas
Circuits and Systems
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road
London, SW7 2AZ, UK
e.rodriguez@imperial.ac.uk, +4475946261, Fax: +44 207 581 4419

Keywords: Apnea, Breathing, Low Power Electronics, Medical Devices, Physiological Monitoring

Word Count. 3500

ABSTRACT

Rationale. Current techniques for monitoring patients for apnea suffer from significant limitations. These include insufficient availability to meet diagnostic needs, cost, accuracy of results in the presence of artifacts, or difficulty of use in unsupervised conditions.

Objectives. We created and clinically tested a novel miniature medical device that targets to overcome these limitations.

Methods. We studied 20 healthy control subjects and 10 sleep apnea patients. The performance of the new system and also of the FDA approved SOMNO clinical system, conventionally used for sleep apnea diagnosis was evaluated in the same conditions. Both systems were tested during a normal night of sleep in both controls and patients. Their performance was quantified in terms of detection of apnea and hypopnea in individual 10 second epochs, which were compared with scoring of signals by a blinded clinician.

Main Results. For spontaneous apneas during natural sleep and considering the clinician scorer as the gold standard the new wearable apnea detection device had 88.6% sensitivity and 99.6% specificity. In comparison the SOMNO system had 14·3% sensitivity and 99·3% specificity. The novel device had been specifically designed to detect apnea, but if both apnea and hypopneas during sleep were considered in the assessment, the sensitivity and specificity were 77·1% and 99·7% respectively; versus 54% and 98.5% for the SOMNO.

Conclusions. The performance of the novel device compares very well to the scoring by an experienced clinician even in the presence of breathing artifacts, in this small pilot study. This can potentially make it a real solution for apnea home monitoring.

Word Count: 259

SUMMARY

Strenghts:

- We present the smallest, least intrusive technology to automatically detect apneas/hypopneas
- Performance characterization in normal signal conditions and with signal artefacts, showing excellent agreement with expert- 60,000 epochs assessed in controls and patients
- Sensitivity six times better than a state of the art commercial system, and excellent scoring in terms of user acceptance.

Limitations:

- The size of the study is limited. This is however justified by the fact that it was an initial pilot study to prove the strength of this novel technology to detect individual events even in the presence of artefacts (study goals of 95% confidence intervals for sensitivity and specificity values).

- The technology is still not optimized for hypopnea detection.

INTRODUCTION

Apnea may occur acutely in the context of infectious, respiratory, cardiac and neurological disease;[1-5] be caused by medication;[6-8] and on occasion death may be averted with urgent intervention.[1, 6, 9-11] Apnea may also occur recurrently either as a co-morbidity in chronic conditions including asthma, gastro-oesophageal reflux, neuromuscular disorders and diabetes;[12-16] or on its own in sleep apnea syndrome.[17-27]

The importance of monitoring and quantifying apneas is widely acknowledged. Apneas are one of the two leading causes of Sudden Death in Epilepsy (SUDEP), which only in the UK affects more people than cot death and AIDS together.[28-29] Apneas are also known to be a major problem due to its potentially disastrous consequences in anaesthesia recovery rooms.[30-31] And just sleep apnea may affect between 2% and 10% of the adult population [24] and 1% to 3% of the pediatric population,[18] and is heavily under-diagnosed.[19] The indirect medical costs of under diagnosed adult patients, in the years preceding the diagnosis, is estimated to increase by up to a two-fold, even after correcting for chronic disease status.[22-23] This, added to the potential social consequences, in the form of accidents, increased morbidity and impact on work efficiency makes the condition a major public health issue.[24]

Currently existing techniques for monitoring and quantification of apneas are not satisfactory. In sleep apnea diagnosis, polysomnography is the gold standard but the lack of sleep labs, sleep specialists and the associated cost, either make it difficult for the family physician to confirm the suspicion, or delays diagnosis.[25] The importance of the problem has led Medicare and Medicaid in the USA to recently authorize payment of treatment for adults diagnosed with unattended home sleep monitoring devices.[24] Unfortunately

existing home monitoring devices suffer from one or several of the following limitations: the sensors can be difficult to place resulting in invalid recordings, they still require considerable specialist time in order to interpret the results, automatic interpretation is very inaccurate mainly due to the inability to deal with artifacts, sensors can be cumbersome or intrusive so affecting the quality of sleep.[26-27] Furthermore, there is no portable apnea monitoring system that can detect apneas with high enough sensitivity and specificity in real time to potentially be used to alert carers of life-threatening situations due to acute apnea that can occur in the context of other clinical scenarios such as epilepsy or in anaesthesia. In these scenarios also, the alternative of relying on devices that might be able to detect the sequelae of apnea (for example pulse oximeters to detect hypoxeamia or heart rate monitors to detect bradycardia) might result in fatal consequences due to a delayed response to the apnea.

We present the results of the first clinical study of a new wearable apnea detection device (WADD) specifically designed to overcome the limitations of all other existing technologies.

METHODS

Device

We determined that the strongest externally detectable signal related to breathing corresponded to turbulence in the trachea. This signal was detected with a customized acoustic chamber that optimized the signal transmission. The signal detected by the sensor has components corresponding to both the wanted “signal” (breathing) and undesired “noise” caused by artifacts (cardiac signal, external noise (eg speech, music, wind), movement causing rubbing against the sensor and electromagnetic interference). A novel signal processing algorithm was developed to differentiate “signal” from “noise”. The

algorithm evaluates both the temporal and frequency characteristics of the signal obtained from the sensor. Over 15 different features are analyzed with parametric functions that dynamically adapt over time, to compensate for changes in both the subject and the environment. No pre-calibration or subject specific knowledge or modification is needed for the sensor or the signal processing algorithm.[32-34] Part of the algorithm was implemented on hardware and incorporated into the sensor. This reduces the amount of data that is needed for wireless transmission and consequently the amount of power required from the battery; hence the small size of the device. The wireless receiver and the remaining part of the algorithm were run on a laptop computer.

The WADD was wireless, measured 3·74 by 2·4 by 2·1 cm, weighed 17 grams, and was fixed to the skin on the neck with hydrocolloid colostomy adhesive patches of approximately 4cm diameter (Boots). The preferred location was over the trachea, halfway between the lower margin of the thyroid cartilage and the supra-sternal notch (Figure 1(a)). If the skin in that location was loose, as was common in subjects over 40 years of age, the device was placed antero-laterally, anterior to the sternomastoid muscle. The device was left in place overnight, for approximately 14 hours.

Participants

The study was conducted in a sleep study room of the National Hospital for Neurology and Neurosurgery (UK). We studied 20 healthy controls and 10 patients, as they were sequentially admitted for diagnostic monitoring of sleep-related disorders of breathing, because these were likely to have spontaneous apnea events. The patients and controls were not matched. Patients also had a variety of neurological conditions, including epilepsy, dementia, neuropathy and motor neuron disease. The reasons to recruit patients who had

been referred for diagnosis of possible sleep-related disorders of breathing, as opposed to those who had been already diagnosed, were twofold. First, the purpose of this study was not to evaluate the WADD for sleep apnea diagnosis, but rather to evaluate its ability to detect individual events, both during controlled conditions to assess the robustness to artifact rejection, and during spontaneous sleep. Good performance on individual event identification would however be expected to translate in a good performance in the context of the different clinical applications. Secondly, non-diagnosed patients were recruited because studying diagnosed individuals would have involved either delay or interruption of their treatment. The decision on the number of patients was based on obtaining a large enough number of events that would lead to the study goals of 95% confidence intervals for sensitivity and specificity values (based on clinical experience on the minimum number of apnea events which would be expected per subject referred for sleep apnea diagnosis, per night). A larger number of controls were included to be able to assess specificity amongst those who were most likely to be disease free, and also in the presence of artifacts. The patient group comprised 2 females and 8 males with: a median age of 44·5 years of age (range 25-82); a median weight of 74 Kg (range 41-187); a median height of 177 cm (range 160-188); a median body mass index (BMI) of 23 Kg/cm^2 (range 17-61); and a median neck circumference of 40 cm (range 30-43). The control group comprised 3 females and 17 males with: a median age of 33·5 years of age (range 23-63); a median weight of 81.5 Kg (range 60-120); a median height of 176 cm (range 145-185); a median body mass index (BMI) of 26·5 Kg/cm2 (range 20-36); and a median neck circumference of 38 cm (range 34-48). Overall 40% of the subjects were overweight and 24% were obese. The study was approved by the Medicine and Healthcare products Regulatory Agency (MHRA) and the Research Ethics Committee of the UK National Hospital for Neurology and Neurosurgery.

Procedure

All subjects also had simultaneous, clinically standard respiratory monitoring comprising: finger oximetry; oro-nasal airflow sensors; thoracic and abdominal expansion bands; and ECG; using the SOMNO polysomnography system (SOMNOscreen ™ RC kombi . SOMNO Medics, Germany)- Figure 1 (b). Additionally, to further facilitate expert interpretation of polysomnography data, a second pulse oximeter (Pulsox-300i, Konica Minolta, Japan) was attached to the free hand. After attachment of the WADD and the SOMNOmedics polysomnography system, controls subjects participated in a series of exercises, comprising:

T1. Normal breathing for 5 minutes.

T2. Shallow breathing for 5 minutes.

T3. Normal breathing for 45 secs alternating with 15 secs instructed breath holds for 5 min.

T4. Normal breathing for 30 secs alternating with 30 secs instructed breath holds for 10min.

T5. As in 4 but with loud music in the background.

T6. Normal breathing while walking for 5 minutes.

T7. Normal breathing for 30 secs alternating with 30 secs instructed breath holds while lying prone for 5 min.

These exercises were designed to be representative of the worse case of artifact situations affecting the WADD following previous, very exhaustive, lab based research and testing. Following the exercises subjects were allowed to prepare for sleep and were left undisturbed overnight.

Data analysis

The breathing exercises data were analyzed by the automated WADD software and the automated SOMNO software. Instructed apneas were considered to be the "true events". The last six hours of sleep were blindly analyzed by: the automated WADD software, the automated SOMNO software, and by the experienced clinician who reviewed the raw signals from all SOMNO sensors, and had no knowledge about how WADD had been designed or worked. The reason to evaluate the last six hours of sleep was to try to keep the same amount of sleep data in as many subjects as possible in order to prevent biasing of the results. The pulse oximeter was also used by the clinician to support the diagnostic decisions and also individual event classification mostly in those cases in which the signals from the other SOMNO pulse oximeter was corrupted by artifacts. After the separate classification of WADD and SOMNO data, a further investigator compared the results.

The breathing exercises data were analyzed in 15 seconds epochs because this was the shortest duration of an instructed apnea. The sleep data was analyzed in 10 second epochs.

Two assessments were carried out of the sleep data. In the first assessment there was no pre-assumption of a gold standard, and the three systems (WADD, SOMNO and expert marker) were put under test and treated indistinctively. An epoch would be classified as true positive apnea or true positive hypopnea if at least two out of the three systems concurred on the classification. In the second assessment the final classification of these epochs would be that of the expert market, or in other words the expert marker was considered to be the gold standard deciding, and the performance of both SOMNO and WADD system was evaluated. The SOMNO was evaluated as well as the WADD, as there is

little or no quantitative information about the accuracy of automated polysomnography systems.

In both assessments epochs could be classified as:

a) True Positive Apnea (cessation of breathing signal, with correspondent absence of respiratory airflow)
b) True Positive Hypopnea (over 50% reduction in oronasal signal and in thoracoabdominal movement together with over 2% decrease in oxygen saturation).
c) False Positive Hypopneas (if a system had classified a breathing epoch as a hypopnea).
d) False Positive Apnea (if a system had classified a breathing epoch as an apnea)
e) False Classification Apnea as Hypopnea (if a system had classified an apnea epoch as hypopnea).
f) False Classification Hypopnea as Apnea (if a system had classified hypopnea as apnea).
g) False Negative Apnea (if a system classifies an apnea as breathing).
h) False Negative Hypopnea (if a system classifies a hypopnea as breathing).

The breathing exercises data were analyzed in the same way, but the instructed apneas and breathing sections were considered the absolute truth and hence there was no independent expert review.

The performance of the three systems was evaluated, using the following metrics:

Sensitivity=(TP)/(TP+FN)

Specificity= (FP)/(TN+FP)

(TP=True Positive, TN=True Negative, FP=False Positive,FN=False Negative).

The analysis was carried out assuming that all apnea events were independent, since it was observed that the characteristics of the breathing signal changed as much within the same subject (depending on timing, position, external artefacts, etc.), as between different subjects. This was further verified by taking three random 10 minutes sections of the sensed breathing signals in the 30 different subjects and obtaining the different correlation coefficients (2700 in total). The maximum correlation coefficient obtained from signals within the same subject was 0.05. The maximum correlation coefficient obtained from different subjects was 0.067.

For each one of the two assessments (i.e. not presuming a gold standard, and considering the expert to be the gold standard), two different analyses were carried out. Firstly only apneas were considered to be true positives. Hence any hypopnea would be regarded as breathing (true negative); False Classification of Hypopneas as Apneas were re-classified as false positives; and False classification of Apneas as Hypopneas were re-classified as false negatives. Secondly, apneas and hypopneas were considered indistinctively, and hence true events of both variety would be also considered together.

These two analyses were carried out as they would be relevant to different clinical scenarios. For example, high sensitivity for detecting apnea would be crucial for identification of sudden apnea if monitoring those with epilepsy; whereas for diagnosis of sleep-related breathing dysfunction, which generally relies on the Apnea Hypopnea Index, the differentiation between apnea and hypopnea might be clinically less important.

RESULTS

Breathing exercises data

Data were available in 3956 15 second epochs for the controls performing the breathing exercises (132 in total). Table 1 summarizes the performance of WADD and SOMNO in the seven breathing exercises. Figure 2 illustrates examples of the signals obtained from the different sensors. Table 1 is divided in three parts. Part (a) and (b) quantify performance considering different scenarios for wrongly classified hypopneas. Although the real events were apneas, both systems had the ability to indicate hypopneas too. This resulted in some real apnea and breathing epochs being wrongly marked as hypopneas. In order to account for these, Table 1 (a) shows the sensitivity and specificity when only apneas are considered as events (i.e. hypopneas would be regarded as breathing). In part (b) of the table hypopneas and apneas are indistinctively considered. Part (c) illustrates the total number of epochs that fall into a specific classification for both systems. The combined sensitivity and specificity for all the exercises across all the subjects for the WADD was 97·7% and 99·6% (considering hypopneas as breathing); or 99·2% and 99·5% (considering hypopneas as events). With the same criteria the sensitivity and specificity for the SOMNO was only 37·8% sensitivity, 96·5% specificity; or 62·8% sensitivity, 90·5% specificity.

WADD versus Somno performance in instructed exercises with hypopneas NOT considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99.2
T2	NA	NA	100	90.6
T3	94·6	38	99	96·9
T4	98·9	38·8	99·7	94·5
T5	99·2	31·4	99·2	99·7
T6	NA	NA	100	96·5
T7	94·2	48·2	98·5	99
Total	**97·7**	**37·8**	**99·6**	**96·5**

(a)

WADD versus Somno performance in instructed exercises with hypopneas and apneas indistinctively considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99
T2	NA	NA	100	81.4
T3	96·7	66.3	99	87·5
T4	100	64·6	99·7	87·9
T5	99·2	59.2	99·2	95·1
T6	NA	NA	100	89·8
T7	99	64·4	97	93.4
Total	**99·2**	**62·8**	**99·5**	**90·5**

(b)

Summary of classification of the different epochs										
Exercise	TP		TN		FC		FP		FN	
	WADD	Somno	WADD	Somno	WADD	Somno	WADD	Somno	WADD	Somno
T1	0	0	380	376	0	0	0	4	0	0
T2	0	0	360	293	0	0	0	67	0	0
T3	89	61	285	252	2	26	3	36	3	31
T4	356	230	363	320	4	92	1	44	0	126
T5	357	213	365	350	0	100	3	18	3	147
T6	0	0	400	359	0	0	0	41	0	0
T7	189	123	191	184	9	31	6	13	2	68
Total	**991**	**627**	**2344**	**2134**	**15**	**249**	**13**	**223**	**8**	**372**

(c)

Table 1: Summary of performance for the WADD and SOMNO across the seven breathing exercises (as detailed in Procedure). TP= true positive (apnea), TN=true negative (breathing), FC= false classification, FP= False Positive, FN=False Negative. Part (a) of the table shows the sensitivity and specificity not considering hypopneas as events (i.e. all hypopneas are considered breathing). Based on this all False Classifications, FC (apneas wrongly classified as hypopneas) are considered False Negatives (FN); and all False Positives hypopneas are considered True Negatives (TN). Part (b) shows the sensitivity and specificity considering apnea and hypopnea as indistinctive events. Based on this all False Classifications are re-classified as True Positives (TP); and all False Positives hypopneas are False Positives (FP). Part (c) details the number of epochs corresponding to a particular classification.

Sleep data

For the sleep data 62,727 10 second epochs were analyzed in total. 34 true apnea epochs and 40 true hypopnea epochs were identified for the controls (36 and 37 if the clinician scorer was considered to be the gold standard); and 312 apnea epochs and 181 hypopnea epochs for the patients (342 and 200 if the clinician scorer was considered to be the gold standard). The average number of apnea epochs for the patient group throughout the night was 32. All patients had episodes of apnea or hypopnea. There was only two patients who did not have any episode of apnea. For one control, only 3·2 hours of data were recorded, because of an ICT error. For one patient, only 3 hours were analyzed as more than one

SOMNO sensor including the nasal cannula and the pulse oximeters detached prematurely. The results in terms of sensitivity and specificity, for the control group, patient group and overall are presented in Table 2. Table 2 is divided in four parts: the first and second evaluate the performance for apnea and apnea/hypopnea combined detection respectively without assuming a gold standard (i.e. the consensus of the majority determines a true event); and the third and fourth parts present the same evaluation but considering the expert as the gold standard.

Apnea detection (% sensitivity and specificity)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**47·1**	**97·1**	**94·1**	**99·3**	**99·8**	**100**
(95% CI)	(30.3-63.8)	(91.4-100)	(86.2-100)	(99.2-99.3)	(99.7/99.8)	(100-100)
Patients	**14·7**	**99·4**	**98·1**	**99·5**	**99·5**	**99·9**
(95% CI)	(10.8-18.7)	(98.5-100)	(96.6-99.6)	(99.5-99.6)	(99.4-99.6)	(99.8-99.9)
All	**17·9**	**99·1**	**97·7**	**99·4**	**99·7**	**99·9**
(95% CI)	(13.9-22.0)	(98.2/100)	(96.1-99.3)	(99.3-99.4)	(99.6-99.7)	(99.9-100)

Apnea and Hypopnea combined detection (%)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**87·8**	**58·1**	**94·6**	**98·6**	**99·7**	**100**
(95% CI)	(80.4-95.3)	(46.9-69.4)	(89.4-99.8)	(98.5-98.7)	(99.6-99.7)	(100-100)
Patients	**53·3**	**88·2**	**98·8**	**97·9**	**99·5**	**99·8**
(95% CI)	(48.9-57.8)	(85.4-91.1)	(97.8-99.8)	(97.7-98.1)	(99.4-99.6)	(99.8-99.9)
All	**57·8**	**84·1**	**98·2**	**98·4**	**99·5**	**99·9**
(95% CI)	(53.8-61.9)	(81.1-87.1)	(97.2-99.3)	(98.3-98.5)	(99.5-99.6)	(99.9-100)

Apnea Detection with the clinician scorer as Gold Standard reference (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**38·9**	**86·1**	**99·2**	**99·7**
(95% CI)	(23.0-54.8)	(74.8-97.4)	(99.1-99.3)	(99.7-99.8)
Patients	**11·7**	**88·9**	**99·5**	**99·4**
(95% CI)	(8.3-15.1)	(85.6-92.2)	(99.4-99.6)	(99.3-99.5)
All	**14·3**	**88·6**	**99·3**	**99·6**
(95% CI)	(10.8-17.8)	(85.4-91.8)	(99.2-99.4)	(99.6-99.7)

Apnea and Hypopnea combined detection with clinician scorer as Gold Standard (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**86·3**	**54·8**	**98·6**	**99·6**
(95% CI)	(63/73)	(40/73)	(41539/42139)	(41987/42139)
Patients	**49·6**	**80·1**	**98·4**	**100**
(95% CI)	(45.4-53.8)	(75.2-82.1)	(98.2-98.6)	(99.9-100)
All	**54·0**	**77·1**	**98·5**	**99·7**
(95% CI)	(50.0-57.9)	(73.8-80.5)	(98.4-98.6)	(99.7-99.8)

Table 2: Summary of performance for the WADD, SOMNO and clinician scorer systems for detection of apnea and hypopnea in 15 second epochs of overnight recordings.

The WADD also had the added feature of being able to differentiate between central and obstructive apnea. 90% of the central apneas were rightly marked as central. 96% were rightly marked as obstructive. Approximately 60% of the total apneas were obstructive in origin.

Device comfort

After the overnight study, the devices were detached and the subjects scored the comfort of the devices and quality of sleep (rating 1 to 5, with 5 representing maximum comfort and quality). Skin irritation caused by the WADD's adhesive was also rated from 1 to 5 (5 representing no irritation, 4 mild transient, redness, and 1 severe irritation). The median rating for WADD comfort was 5 (range 4-5). The median rating for SOMNO comfort was 3 (range 1-5 for controls and 2-5 for patients). The median rating for irritation caused by the WADD plaster on the neck was 5 (range 5-5 for controls and 4-5 for patients).

DISCUSSION

Main findings

WADD had very high sensitivity and specificity for detecting apnea in 15 second epochs in a series of breathing and breath-holding exercises in a variety of conditions, including the presence of external background noise, movement and posture. The tolerability of WADD was superior to the portable polysomnography system (SOMNO) during overnight recordings.

WADD had 97·7-99·2% sensitivity to detect instructed apneas and 88.6-99·1% for 10 seconds spontaneous apneas during natural sleep, with similar performance in controls and patients. The WADD also detected all apneas over 30 seconds and there were only 3 over 30 seconds false positives. For short apneas, in most cases, disagreement between the clinician scorer and the WADD were caused by the WADD identifying as apnea epochs that the expert classified as hypopneas.

As expected, the WADD performance was less good when apneas and hypopneas were considered together (minimum sensitivity 77·1%).This is not surprising since the WADD was designed to identify apnea, not hypopneas, and the latter were detected from the transmitted signal which had already been pre-processed for apnea detection. From the table, it can be observed that the degradation of performance was more evident in the controls because the controls had a large number of shorter hypopneas (under 22.5 seconds) which the WADD did not detect properly. In the patients, who often demonstrated apneas, the hypopnea events were longer and these were detected by WADD. Although the lower sensitivity in hypopnea detection might in principle seem problematic if the WADD

was to be used in the context of sleep apnea diagnosis (hypopneas are very common events in sleep labs), it is worth noting that: 1) there is no other reported automatic system that gets anywhere close to this with similar specificity and apnea detection performance; 2) the variations between different sleep labs due to the non-uniform definition of hypopneas already leads to much larger diagnostic variations than the limitation in sensitivity of the WADD;[35-37] 3) assuming the worse case scenario for the WADD, this is that a patient only had hypopneas throughout the night, this reduced sensitivity would be a problem that would translate to non-diagnosis of sleep apnea for patients that with 100% sensitivity would have had a sleep apnea hypopnea index (AHI) between 5 and 6 (i.e very mild cases of sleep apnea). Patients with no sleep apnea, moderate sleep apnea, severe sleep apnea and those with mild sleep apnea with AHI between 6 and 15 would have been rightly diagnosed.

The median difference between the WADD calculated Apnea Hypopnea Index (AHI) and the one obtained by the gold standard was 0 (average=0.7).

SOMNO Performance

The automatic analysis of the SOMNO apparatus, an FDA approved and clinically accepted system, based on assessing apnea from a variety of different sensors, significantly differed from that of the expert marker, with an average sensitivity value of around 14%. The results obtained from the instructed apneas tests also showed that even in the absence of artifacts, apneas were not well detected by the SOMNO system, with an average sensitivity of 37·8%. This demonstrates the need for caution if relying on current automated assessment methods for diagnosing apnea. Whilst performance might be improved by optimizing parameters for individual patients, this is not practical for single overnight recordings or use

as an alerting monitor. The WADD does not require any parameter optimization or subject specific calibration.

The SOMNO system performance improved in the event of indistinct classification of apneas and hypopneas, but was still poor compared with the clinician scorer (54% sensitivity). This sensitivity was at the cost of reduced specificity: for every true hypopnea detected there were approximately four false detections. Overall, the performance of the WADD in hypopnea/apnea combined detection was significantly better than the SOMNO, in sensitivity (77·1% vs 54% if considering the clinician scorer as a gold standard, and 84·1% vs 57·8% otherwise), but also in specificity, as the WADD only detected one false hypopnea epoch for every four true events.

Limitations. Future improvements

The study described in this paper is a small pilot study and hence further more comprehensive clinical evaluation of the technology will be necessary before it can be used. The size of the study was however adequate to assess the potential of the technology; to determine whether the initial performance results in controlled conditions were equivalent to those obtained in real scenarios; and to inform a clinical trial. Based on these positive results it is expected that a fully powered clinical trial, focused on diagnosis rather than on individual event identification, will follow in the future.

The calculation of the sensitivity and specificity has assumed that all apnea events were independent, which for some might not be completely correct. If the data had not been pooled, and taking the expert marker as the gold-standard, in 67% of the subjects the individual apnoea detection sensitivity was 100%. In 77% it was over 90%. In the remaining cases, the drop in sensitivity corresponded always to just one non-detected apnoea shorter

than 15 seconds in the 6 hour night, which is clinically insignificant. The average from the individual sensitivities was 2% higher than the value obtained pooling the data. In terms of specificity 90% of the subjects had values higher than 99%. Two thirds of them were over 99.9%. The average of the individual specificities was identical to the specificity obtained pooling the data.

A different statistical analysis, possibly comparing pooled with non-pooled data will be the subject of investigation when the technology undergoes a larger clinical trial.

The WADD is obviously no substitute to a full night study in a sleep clinic, since it does not provide all the information that a full polysomnography system would. There are advantages and disadvantages to this device with respect to full polysomnography. The WADD can be used to determine the Apnea Hypopnea Index (AHI), which is used in sleep apnea diagnosis to ascertain whether a patient has sleep apnea and to score the severity of the condition. The main advantage is that it can be used for at home assessment or monitoring, and from that point of view it is clearly superior to any of the other existing devices (highly resilient to artifacts, very easy to attach and durable in position, low cost, much more comfortable, and accurate). Considering the restricted resources for sleep clinic referral this device could be a very useful tool to determine at very low cost who should be referred to a specialist centre for full polysomnography. The disadvantage is that there are other parameters that could be used for extra assessment that the device does not measure, such as microarousal or full cardiac activity. Furthermore, the WADD does not allow to assess the hypoxic load or autonomic activation and therefore impact the cardiovascular or stroke risk associated with OSA syndrome.

The WADD device used in the current study relied on wireless transmission to a PC. However changing the PC to a dedicated mobile phone sized receiver poses no technological challenge. A subsequent version that is being developed is smaller (2.4 by 2.4 by 1.2 cm, weighing 7.5 grams) and can operate continuously on hearing aid batteries for over 48 hours. It has a separate dedicated receiver of comparable size to a mobile phone which can be located up to 10 metres from the subject.

Acknowledgments: Eduardo Aguilar-Pelaez for assisting with the data collection at the initial phases of the clinical study and proof reading one of the versions of the manuscript; Pauline Rogers (Statistician, UCL/H Biomedical Research Unit) for carrying out the statistical analysis leading to the number of tested subjects; UCL Business; Epilepsy Society; and Staff on MITU at NHNN.

Sources of Support. UCL Business and Imperial Innovations.

Authors' contributions: John Duncan and Esther Rodriguez- Villegas had the initial idea of developing a wearable breathing monitoring device. Esther Rodriguez-Villegas and Guangwei Chen were the main creators of the novel device carrying out the electronic design which includes both hardware and software. John Duncan contributed through the process helping to define the engineering specifications to meet the clinical need. John Duncan designed the protocol for the clinical study. Esther Rodriguez-Villegas created the documentation for MHRA approval. Jeremy Radcliffe led the clinical studies and blindly marked all the data using the sensors of the SOMNO system as well as the Konica Minolta pulse oximeter. Esther Rodriguez-Villegas and Guangwei Chen compared the results of the data scored by the WADD, SOMNO and blinded expert, and carried out the data analysis. All the authors contributed to the interpretation of data. John Duncan, Jeremy Radcliffe and Esther Rodriguez- Villegas wrote the paper.

Conflict of interest: There are no competing interests.

Data sharing: No additional data available

Figure Legends:

Figure 1: (a) WADD worn by one of the investigators. (b) Subject wearing an existing state of the art ambulatory apnea monitoring system (SOMNO), comprising finger oximetry; oro-nasal flow sensors; thoracic and abdominal expansion bands; and ECG.

Figure 2: Illustration of the SOMNO and WADD output signals showing an apnea event: (top) raw signals from the different SOMNO sensors, (middle) processed WADD signal, (bottom) WADD output signal.

References

1 Shorvon S, Tomnson T. Sudden unexpected death in epilepsy. *Lancet* 2011;378(9808):2028–38.

2 Mace SE. Acute bacterial meningitis. *Emerg Med Clin North Am* 2008;**26**(2):281–317.

3 Cordero L, Rau R, Taylor D, et al. Enteric gram-negative bacilli bloodstream infections: 17 years' experience in a neonatal intensive care unit. *Am J Infect Control* 2004;**32**(4):189–95.

4 Powers, KS. Acute pulmonary infections. In: Lucking SE, Maffei, FA, Tamburro, RF and Thomas, NJ, editors. Pediatric critical care study guide. London: Springer 2012; 514–534.

5 Moruzzi P, Sarzi-Braga S, Rossi M, et al. Sleep apnea in ischaemic heart disease: differences between acute and chronic coronary syndromes. *Heart* 1999;**82**(3):343–7.

6 Van Vliet G, Deal CL, Crock PA, et al. Sudden death in growth hormone-treated children with Prader-Willi syndrome. *J Pediatr* 2004;**144**(1):129–31.

7 Taylor MB, Grounds RM, Mulrooney PD, et al. Ventilatory effects of propofol during induction of anaesthesia. *Anaesthesia* 1986;**41**(8):816–20.

8 Murphy JJ, Swanson T, Ansermino M, et al. The frequency of apneas in premature infants after inguinal hernia repair: do they need overnight monitoring in the intensive care unit? *J Pediatr Surg* 2008;**43**(5):865-8.

9 Naeem N, Montenegro H. Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation* 2005;**67**(1): 13-23.

10 Moon RY, Horne RS, Hauck FR. Sudden infant death syndrome. *Lancet* 2007;**370**(9598):1578–87.

11 Gami AS, Howard DE, Olson EJ, et al. Day-night pattern of sudden death in obstructive sleep apnea, *N Engl J Med* 2005;**352**(12):1206–14.

12 Julien JY, Martin JG, Ernst P, et al. Prevalence of obstructive sleep apnea–hypopnea in severe versus moderate asthma. *J Allergy Clin Immunol* 2009;**124**(2):371–6.

13 Emilsson O, Janson C, Benediktsdóttir B,et al. Nocturnal gastroesophageal reflux, lung function and symptoms of obstructive sleep apnea: Results from an epidemiological survey. *Respir Med* 2012;**106**(3):459–66.

14 Reichmuth K, Austin D, Skatrud JB, et al. Association of sleep apnea and type II diabetes. A population-based study. *Am J Respir Crit Care Med* 2005;**172**(12):1590–5.

15 Hanly PJ. Mechanisms and management of central sleep apnea. *Lung* 1992;**170**(1):1–17.

16 Culebras A. Sleep and neuromuscular disorders. *Neurol Clin* 2005;**23**(4):1209–23.

17 Young T, Palta M, Dempsey J, et al. The occurrence of sleep disordered breathing among middle aged adults. *N Engl J Med* 1993;**328**(17):1230–5.

18 Faroux B. What's new in pediatric sleep? *Pediatr Respir Rev* 2007;**8**(1):85–9

19 Pang KP, Terris DJ. Screening for obstructive sleep apnea; an evidence-based analysis. *Am J Otolaryngol* 2006;**27**(2):112–8.

20 Redline S, Quan SF. Sleep apnea: a common mechanism for the deadly triad--cardiovascular disease, diabetes, and cancer? *Am J Respir Crit Care Med* 2012;**186**(2):123–124.

21 Hirshkowitz M. The clinical consequences of obstructive sleep apnea and associated excessive sleepiness. *J Fam Pract* 2008;**57**(8 Suppl):S9–16.

22 Wittmann V, Rodenstein DO. Health care costs and the sleep apnea syndrome. *Sleep Medicine Rev* 2004;**8**(4):269–279

23 Smith R, Ronald J, Delaive K, et al. What are obstructive sleep apnea patients being treated for prior to this diagnosis? *Chest* 2002;**12**(1):164-172.

24 Leger D, Bayon V, Laaban JP, et al. Impact of sleep apnea on economics. *Sleep Med Rev* 2012;**16**(5):455–62.

25 Flemons WW, Douglas NJ, Kuna ST, et al. Access to diagnosis and treatment of patients with suspected sleep apnea. *Am J Respir Crit Care Med* 2004;**169**(6):668–72.

26 Oliveira MG, Nery LE, Santos-Silva R, et al. Is portable monitoring accurate in the diagnosis of obstructive sleep apnea syndrome in chronic pulmonary obstructive disease? *Sleep Med*. 2012;**13**(8):1033-8.

27 Flemons WW, Littner MR; Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest* 2003;**124**(4):1543–79

28 McDonald F. Avoidable Epilepsy-Related Deaths: The case for prevention. *Epilepsy Bereaved Report*, 2011.

29 Surges R, Sander JW. Sudden unexpected death in epilepsy: mechanisms, prevalence and prevention. *Curr Opin Neurol* 2012;**2**:201–207.

30 Lofsky A. Sleep apnea and narcotic postoperative pain medication: a morbidity and mortality risk. *Anaesthesia Patient Safety Foundation Newsletter*. Summer 2002;**17**:24–25.

31 Bhananker SM, Posner KL, Cheney FW, et al. Injury and liability associated with monitored anesthesia care: A closed claims analysis. *Anesthesiology* 2006;**104**:228–34.

32 Rodriguez-Villegas E, Corbishley P, Duncan JS, et al. A method and apparatus for monitoring respiratory activity. Patent *WIPO;* WO/2007/099314.

33 Corbishley PG, Rodriguez-Villegas E. Breathing detection: towards a miniaturized wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering* 2008; **55**(1):196–204.

34 Rodriguez-Villegas E, Chen G, Aguilar-Pelaez E. Feature characterization for breathing monitor. Patent *WIPO*; WO/2012/007719.

35 Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. *Medicina* 2003;**63**(2):119–24.

36 Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med* 2000;**161** (2):369–74.

37 Ruehland WR, Rochford PD, O'Donoghue FJ, et al. The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index, *Sleep* 2009;**32**(2):150–157.

A PILOT STUDY OF A WEARABLE APNEA DETECTION DEVICE

Rodriguez-Villegas Esther[1], PhD, SMIEEE; Gwangwei Chen[1]; Radcliffe Jeremy[2], FRCA, Duncan John[3], FRCP

[1]Dept of Electrical and Electronic Engineering,
Imperial College London, Exhibition Road
London SW7 2AZ, UK

[2]Dept of Neuroanaesthesia
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

[3]Dept of Clinical and Experimental Epilepsy
UCL Institute of Neurology
National Hospital for Neurology and Neurosurgery, Queen Square
London WC1N 3BG, UK

Corresponding author:
Esther Rodriguez-Villegas
Circuits and Systems
Department of Electrical and Electronic Engineering
Imperial College London
Exhibition Road
London, SW7 2AZ, UK
e.rodriguez@imperial.ac.uk, +4475946261, Fax: +44 207 581 4419

Keywords: Apnea, Breathing, Low Power Electronics, Medical Devices, Physiological Monitoring

Word Count. 3500

ABSTRACT

Rationale. Current techniques for monitoring patients for apnea suffer from significant limitations. These include insufficient availability to meet diagnostic needs, cost, accuracy of results in the presence of artifacts, or difficulty of use in unsupervised conditions.

Objectives. We created and clinically tested a novel miniature medical device that targets to overcome these limitations.

Methods. We studied 20 healthy control subjects and 10 sleep apnea patients. The performance of the new system and also of the FDA approved SOMNO clinical system, conventionally used for sleep apnea diagnosis was evaluated in the same conditions. Both systems were tested during a normal night of sleep in both controls and patients. Their performance was quantified in terms of detection of apnea and hypopnea in individual 10 second epochs, which were compared with scoring of signals by a blinded clinician.

Main Results. For spontaneous apneas during natural sleep and considering the clinician scorer as the gold standard the new wearable apnea detection device had 88.6% sensitivity and 99.6% specificity. In comparison the SOMNO system had 14·3% sensitivity and 99·3% specificity. The novel device had been specifically designed to detect apnea, but if both apnea and hypopneas during sleep were considered in the assessment, the sensitivity and specificity were 77·1% and 99·7% respectively; versus 54% and 98.5% for the SOMNO.

Conclusions. The performance of the novel device compares very well to the scoring by an experienced clinician even in the presence of breathing artifacts, in this small pilot study. This can potentially make it a real solution for apnea home monitoring.

Word Count: 259

SUMMARY

Strenghts:

- We present the smallest, least intrusive technology to automatically detect apneas/hypopneas
- Performance characterization in normal signal conditions and with signal artefacts, showing excellent agreement with expert- 60,000 epochs assessed in controls and patients
- Sensitivity six times better than a state of the art commercial system, and excellent scoring in terms of user acceptance.

Limitations:

- The size of the study is limited. This is however justified by the fact that it was an initial pilot study to prove the strength of this novel technology to detect individual events even in the presence of artefacts (study goals of 95% confidence intervals for sensitivity and specificity values).

- The technology is still not optimized for hypopnea detection.

INTRODUCTION

Apnea may occur acutely in the context of infectious, respiratory, cardiac and neurological disease;[1-5] be caused by medication;[6-8] and on occasion death may be averted with urgent intervention.[1, 6, 9-11] Apnea may also occur recurrently either as a co-morbidity in chronic conditions including asthma, gastro-oesophageal reflux, neuromuscular disorders and diabetes;[12-16] or on its own in sleep apnea syndrome.[17-27]

The importance of monitoring and quantifying apneas is widely acknowledged. Apneas are one of the two leading causes of Sudden Death in Epilepsy (SUDEP), which only in the UK affects more people than cot death and AIDS together.[28-29] Apneas are also known to be a major problem due to its potentially disastrous consequences in anaesthesia recovery rooms.[30-31] And just sleep apnea may affect between 2% and 10% of the adult population [24] and 1% to 3% of the pediatric population,[18] and is heavily under-diagnosed.[19] The indirect medical costs of under diagnosed adult patients, in the years preceding the diagnosis, is estimated to increase by up to a two-fold, even after correcting for chronic disease status.[22-23] This, added to the potential social consequences, in the form of accidents, increased morbidity and impact on work efficiency makes the condition a major public health issue.[24]

Currently existing techniques for monitoring and quantification of apneas are not satisfactory. In sleep apnea diagnosis, polysomnography is the gold standard but the lack of sleep labs, sleep specialists and the associated cost, either make it difficult for the family physician to confirm the suspicion, or delays diagnosis.[25] The importance of the problem has led Medicare and Medicaid in the USA to recently authorize payment of treatment for adults diagnosed with unattended home sleep monitoring devices.[24] Unfortunately

existing home monitoring devices suffer from one or several of the following limitations: the sensors can be difficult to place resulting in invalid recordings, they still require considerable specialist time in order to interpret the results, automatic interpretation is very inaccurate mainly due to the inability to deal with artifacts, sensors can be cumbersome or intrusive so affecting the quality of sleep.[26-27] Furthermore, there is no portable apnea monitoring system that can detect apneas with high enough sensitivity and specificity in real time to potentially be used to alert carers of life-threatening situations due to acute apnea that can occur in the context of other clinical scenarios such as epilepsy or in anaesthesia. In these scenarios also, the alternative of relying on devices that might be able to detect the sequelae of apnea (for example pulse oximeters to detect hypoxeamia or heart rate monitors to detect bradycardia) might result in fatal consequences due to a delayed response to the apnea.

We present the results of the first clinical study of a new wearable apnea detection device (WADD) specifically designed to overcome the limitations of all other existing technologies.

METHODS

Device

We determined that the strongest externally detectable signal related to breathing corresponded to turbulence in the trachea. This signal was detected with a customized acoustic chamber that optimized the signal transmission. The signal detected by the sensor has components corresponding to both the wanted “signal” (breathing) and undesired “noise” caused by artifacts (cardiac signal, external noise (eg speech, music, wind), movement causing rubbing against the sensor and electromagnetic interference). A novel signal processing algorithm was developed to differentiate “signal” from “noise”. The

algorithm evaluates both the temporal and frequency characteristics of the signal obtained from the sensor. Over 15 different features are analyzed with parametric functions that dynamically adapt over time, to compensate for changes in both the subject and the environment. No pre-calibration or subject specific knowledge or modification is needed for the sensor or the signal processing algorithm.[32-34] Part of the algorithm was implemented on hardware and incorporated into the sensor. This reduces the amount of data that is needed for wireless transmission and consequently the amount of power required from the battery; hence the small size of the device. The wireless receiver and the remaining part of the algorithm were run on a laptop computer.

The WADD was wireless, measured 3·74 by 2·4 by 2·1 cm, weighed 17 grams, and was fixed to the skin on the neck with hydrocolloid colostomy adhesive patches of approximately 4cm diameter (Boots). The preferred location was over the trachea, halfway between the lower margin of the thyroid cartilage and the supra-sternal notch (Figure 1(a)). If the skin in that location was loose, as was common in subjects over 40 years of age, the device was placed antero-laterally, anterior to the sternomastoid muscle. The device was left in place overnight, for approximately 14 hours.

Participants

The study was conducted in a sleep study room of the National Hospital for Neurology and Neurosurgery (UK). We studied 20 healthy controls and 10 patients, as they were sequentially admitted for diagnostic monitoring of sleep-related disorders of breathing, because these were likely to have spontaneous apnea events. The patients and controls were not matched. Patients also had a variety of neurological conditions, including epilepsy, dementia, neuropathy and motor neuron disease. The reasons to recruit patients who had

been referred for diagnosis of possible sleep-related disorders of breathing, as opposed to those who had been already diagnosed, were twofold. First, the purpose of this study was not to evaluate the WADD for sleep apnea diagnosis, but rather to evaluate its ability to detect individual events, both during controlled conditions to assess the robustness to artifact rejection, and during spontaneous sleep. Good performance on individual event identification would however be expected to translate in a good performance in the context of the different clinical applications. Secondly, non-diagnosed patients were recruited because studying diagnosed individuals would have involved either delay or interruption of their treatment. The decision on the number of patients was based on obtaining a large enough number of events that would lead to the study goals of 95% confidence intervals for sensitivity and specificity values (based on clinical experience on the minimum number of apnea events which would be expected per subject referred for sleep apnea diagnosis, per night). A larger number of controls were included to be able to assess specificity amongst those who were most likely to be disease free, and also in the presence of artifacts. The patient group comprised 2 females and 8 males with: a median age of 44·5 years of age (range 25-82); a median weight of 74 Kg (range 41-187); a median height of 177 cm (range 160-188); a median body mass index (BMI) of 23 Kg/cm^2 (range 17-61); and a median neck circumference of 40 cm (range 30-43). The control group comprised 3 females and 17 males with: a median age of 33·5 years of age (range 23-63); a median weight of 81.5 Kg (range 60-120); a median height of 176 cm (range 145-185); a median body mass index (BMI) of 26·5 Kg/cm2 (range 20-36); and a median neck circumference of 38 cm (range 34-48). Overall 40% of the subjects were overweight and 24% were obese. The study was approved by the Medicine and Healthcare products Regulatory Agency (MHRA) and the Research Ethics Committee of the UK National Hospital for Neurology and Neurosurgery.

Procedure

All subjects also had simultaneous, clinically standard respiratory monitoring comprising: finger oximetry; oro-nasal airflow sensors; thoracic and abdominal expansion bands; and ECG; using the SOMNO polysomnography system (SOMNOscreen ™ RC kombi . SOMNO Medics, Germany)- Figure 1 (b). Additionally, to further facilitate expert interpretation of polysomnography data, a second pulse oximeter (Pulsox-300i, Konica Minolta, Japan) was attached to the free hand. After attachment of the WADD and the SOMNOmedics polysomnography system, controls subjects participated in a series of exercises, comprising:

T1. Normal breathing for 5 minutes.

T2. Shallow breathing for 5 minutes.

T3. Normal breathing for 45 secs alternating with 15 secs instructed breath holds for 5 min.

T4. Normal breathing for 30 secs alternating with 30 secs instructed breath holds for 10min.

T5. As in 4 but with loud music in the background.

T6. Normal breathing while walking for 5 minutes.

T7. Normal breathing for 30 secs alternating with 30 secs instructed breath holds while lying prone for 5 min.

These exercises were designed to be representative of the worse case of artifact situations affecting the WADD following previous, very exhaustive, lab based research and testing. Following the exercises subjects were allowed to prepare for sleep and were left undisturbed overnight.

Data analysis

The breathing exercises data were analyzed by the automated WADD software and the automated SOMNO software. Instructed apneas were considered to be the "true events". The last six hours of sleep were blindly analyzed by: the automated WADD software, the automated SOMNO software, and by the experienced clinician who reviewed the raw signals from all SOMNO sensors, and had no knowledge about how WADD had been designed or worked. The reason to evaluate the last six hours of sleep was to try to keep the same amount of sleep data in as many subjects as possible in order to prevent biasing of the results. The pulse oximeter was also used by the clinician to support the diagnostic decisions and also individual event classification mostly in those cases in which the signals from the other SOMNO pulse oximeter was corrupted by artifacts. After the separate classification of WADD and SOMNO data, a further investigator compared the results.

The breathing exercises data were analyzed in 15 seconds epochs because this was the shortest duration of an instructed apnea. The sleep data was analyzed in 10 second epochs.

Two assessments were carried out of the sleep data. In the first assessment there was no pre-assumption of a gold standard, and the three systems (WADD, SOMNO and expert marker) were put under test and treated indistinctively. An epoch would be classified as true positive apnea or true positive hypopnea if at least two out of the three systems concurred on the classification. In the second assessment the final classification of these epochs would be that of the expert market, or in other words the expert marker was considered to be the gold standard deciding, and the performance of both SOMNO and WADD system was evaluated. The SOMNO was evaluated as well as the WADD, as there is

little or no quantitative information about the accuracy of automated polysomnography systems.

In both assessments epochs could be classified as:

a) True Positive Apnea (cessation of breathing signal, with correspondent absence of respiratory airflow)
b) True Positive Hypopnea (over 50% reduction in oronasal signal and in thoracoabdominal movement together with over 2% decrease in oxygen saturation).
c) False Positive Hypopneas (if a system had classified a breathing epoch as a hypopnea).
d) False Positive Apnea (if a system had classified a breathing epoch as an apnea)
e) False Classification Apnea as Hypopnea (if a system had classified an apnea epoch as hypopnea).
f) False Classification Hypopnea as Apnea (if a system had classified hypopnea as apnea).
g) False Negative Apnea (if a system classifies an apnea as breathing).
h) False Negative Hypopnea (if a system classifies a hypopnea as breathing).

The breathing exercises data were analyzed in the same way, but the instructed apneas and breathing sections were considered the absolute truth and hence there was no independent expert review.

The performance of the three systems was evaluated, using the following metrics:

Sensitivity=(TP)/(TP+FN)

Specificity= (FP)/(TN+FP)

(TP=True Positive, TN=True Negative, FP=False Positive,FN=False Negative).

The analysis was carried out assuming that all apnea events were independent, since it was observed that the characteristics of the breathing signal changed as much within the same subject (depending on timing, position, external artefacts, etc.), as between different subjects. This was further verified by taking three random 10 minutes sections of the sensed breathing signals in the 30 different subjects and obtaining the different correlation coefficients (2700 in total). The maximum correlation coefficient obtained from signals within the same subject was 0.05. The maximum correlation coefficient obtained from different subjects was 0.067.

For each one of the two assessments (i.e. not presuming a gold standard, and considering the expert to be the gold standard), two different analyses were carried out. Firstly only apneas were considered to be true positives. Hence any hypopnea would be regarded as breathing (true negative); False Classification of Hypopneas as Apneas were re-classified as false positives; and False classification of Apneas as Hypopneas were re-classified as false negatives. Secondly, apneas and hypopneas were considered indistinctively, and hence true events of both variety would be also considered together.

These two analyses were carried out as they would be relevant to different clinical scenarios. For example, high sensitivity for detecting apnea would be crucial for identification of sudden apnea if monitoring those with epilepsy; whereas for diagnosis of sleep-related breathing dysfunction, which generally relies on the Apnea Hypopnea Index, the differentiation between apnea and hypopnea might be clinically less important.

RESULTS

Breathing exercises data

Data were available in 3956 15 second epochs for the controls performing the breathing exercises (132 in total). Table 1 summarizes the performance of WADD and SOMNO in the seven breathing exercises. Figure 2 illustrates examples of the signals obtained from the different sensors. Table 1 is divided in three parts. Part (a) and (b) quantify performance considering different scenarios for wrongly classified hypopneas. Although the real events were apneas, both systems had the ability to indicate hypopneas too. This resulted in some real apnea and breathing epochs being wrongly marked as hypopneas. In order to account for these, Table 1 (a) shows the sensitivity and specificity when only apneas are considered as events (i.e. hypopneas would be regarded as breathing). In part (b) of the table hypopneas and apneas are indistinctively considered. Part (c) illustrates the total number of epochs that fall into a specific classification for both systems. The combined sensitivity and specificity for all the exercises across all the subjects for the WADD was 97·7% and 99·6% (considering hypopneas as breathing); or 99·2% and 99·5% (considering hypopneas as events). With the same criteria the sensitivity and specificity for the SOMNO was only 37·8% sensitivity, 96·5% specificity; or 62·8% sensitivity, 90·5% specificity.

WADD versus Somno performance in instructed exercises with hypopneas NOT considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99.2
T2	NA	NA	100	90.6
T3	94·6	38	99	96·9
T4	98·9	38·8	99·7	94·5
T5	99·2	31·4	99·2	99·7
T6	NA	NA	100	96·5
T7	94·2	48·2	98·5	99
Total	**97·7**	**37·8**	**99·6**	**96·5**

(a)

WADD versus Somno performance in instructed exercises with hypopneas and apneas indistinctively considered as events				
Exercise	Sensitivity (%)		Specificity (%)	
	WADD	Somno	WADD	Somno
T1	NA	NA	100	99
T2	NA	NA	100	81.4
T3	96·7	66.3	99	87·5
T4	100	64·6	99·7	87·9
T5	99·2	59.2	99·2	95·1
T6	NA	NA	100	89·8
T7	99	64·4	97	93.4
Total	**99·2**	**62·8**	**99·5**	**90·5**

(b)

Summary of classification of the different epochs										
Exercise	**TP**		**TN**		**FC**		**FP**		**FN**	
	WADD	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**	**WADD**	**Somno**
T1	0	0	380	376	0	0	0	4	0	0
T2	0	0	360	293	0	0	0	67	0	0
T3	89	61	285	252	2	26	3	36	3	31
T4	356	230	363	320	4	92	1	44	0	126
T5	357	213	365	350	0	100	3	18	3	147
T6	0	0	400	359	0	0	0	41	0	0
T7	189	123	191	184	9	31	6	13	2	68
Total	**991**	**627**	**2344**	**2134**	**15**	**249**	**13**	**223**	**8**	**372**

(c)

Table 1: Summary of performance for the WADD and SOMNO across the seven breathing exercises (as detailed in Procedure). TP= true positive (apnea), TN=true negative (breathing), FC= false classification, FP= False Positive, FN=False Negative. Part (a) of the table shows the sensitivity and specificity not considering hypopneas as events (i.e. all hypopneas are considered breathing). Based on this all False Classifications, FC (apneas wrongly classified as hypopneas) are considered False Negatives (FN); and all False Positives hypopneas are considered True Negatives (TN). Part (b) shows the sensitivity and specificity considering apnea and hypopnea as indistinctive events. Based on this all False Classifications are re-classified as True Positives (TP); and all False Positives hypopneas are False Positives (FP). Part (c) details the number of epochs corresponding to a particular classification.

Sleep data

For the sleep data 62,727 10 second epochs were analyzed in total. 34 true apnea epochs and 40 true hypopnea epochs were identified for the controls (36 and 37 if the clinician scorer was considered to be the gold standard); and 312 apnea epochs and 181 hypopnea epochs for the patients (342 and 200 if the clinician scorer was considered to be the gold standard). The average number of apnea epochs for the patient group throughout the night was 32. All patients had episodes of apnea or hypopnea. There was only two patients who did not have any episode of apnea. For one control, only 3·2 hours of data were recorded, because of an ICT error. For one patient, only 3 hours were analyzed as more than one

SOMNO sensor including the nasal cannula and the pulse oximeters detached prematurely. The results in terms of sensitivity and specificity, for the control group, patient group and overall are presented in Table 2. Table 2 is divided in four parts: the first and second evaluate the performance for apnea and apnea/hypopnea combined detection respectively without assuming a gold standard (i.e. the consensus of the majority determines a true event); and the third and fourth parts present the same evaluation but considering the expert as the gold standard.

Apnea detection (% sensitivity and specificity)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**47·1**	**97·1**	**94·1**	**99·3**	**99·8**	**100**
(95% CI)	(30.3-63.8)	(91.4-100)	(86.2-100)	(99.2-99.3)	(99.7/99.8)	(100-100)
Patients	**14·7**	**99·4**	**98·1**	**99·5**	**99·5**	**99·9**
(95% CI)	(10.8-18.7)	(98.5-100)	(96.6-99.6)	(99.5-99.6)	(99.4-99.6)	(99.8-99.9)
All	**17·9**	**99·1**	**97·7**	**99·4**	**99·7**	**99·9**
(95% CI)	(13.9-22.0)	(98.2/100)	(96.1-99.3)	(99.3-99.4)	(99.6-99.7)	(99.9-100)

Apnea and Hypopnea combined detection (%)						
	SOMNO Sensitivity	WADD Sensitivity	Clinician Sensitivity	SOMNO Specificity	WADD Specificity	Clinician Specificity
Controls	**87·8**	**58·1**	**94·6**	**98·6**	**99·7**	**100**
(95% CI)	(80.4-95.3)	(46.9-69.4)	(89.4-99.8)	(98.5-98.7)	(99.6-99.7)	(100-100)
Patients	**53·3**	**88·2**	**98·8**	**97·9**	**99·5**	**99·8**
(95% CI)	(48.9-57.8)	(85.4-91.1)	(97.8-99.8)	(97.7-98.1)	(99.4-99.6)	(99.8-99.9)
All	**57·8**	**84·1**	**98·2**	**98·4**	**99·5**	**99·9**
(95% CI)	(53.8-61.9)	(81.1-87.1)	(97.2-99.3)	(98.3-98.5)	(99.5-99.6)	(99.9-100)

Apnea Detection with the clinician scorer as Gold Standard reference (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**38·9**	**86·1**	**99·2**	**99·7**
(95% CI)	(23.0-54.8)	(74.8-97.4)	(99.1-99.3)	(99.7-99.8)
Patients	**11·7**	**88·9**	**99·5**	**99·4**
(95% CI)	(8.3-15.1)	(85.6-92.2)	(99.4-99.6)	(99.3-99.5)
All	**14·3**	**88·6**	**99·3**	**99·6**
(95% CI)	(10.8-17.8)	(85.4-91.8)	(99.2-99.4)	(99.6-99.7)

Apnea and Hypopnea combined detection with clinician scorer as Gold Standard (%)				
	SOMNO Sensitivity	WADD Sensitivity	SOMNO Specificity	WADD Specificity
Controls	**86·3**	**54·8**	**98·6**	**99·6**
(95% CI)	(63/73)	(40/73)	(41539/42139)	(41987/42139)
Patients	**49·6**	**80·1**	**98·4**	**100**
(95% CI)	(45.4-53.8)	(75.2-82.1)	(98.2-98.6)	(99.9-100)
All	**54·0**	**77·1**	**98·5**	**99·7**
(95% CI)	(50.0-57.9)	(73.8-80.5)	(98.4-98.6)	(99.7-99.8)

Table 2: Summary of performance for the WADD, SOMNO and clinician scorer systems for detection of apnea and hypopnea in 15 second epochs of overnight recordings.

The WADD also had the added feature of being able to differentiate between central and obstructive apnea. 90% of the central apneas were rightly marked as central. 96% were rightly marked as obstructive. Approximately 60% of the total apneas were obstructive in origin.

Device comfort

After the overnight study, the devices were detached and the subjects scored the comfort of the devices and quality of sleep (rating 1 to 5, with 5 representing maximum comfort and quality). Skin irritation caused by the WADD's adhesive was also rated from 1 to 5 (5 representing no irritation, 4 mild transient, redness, and 1 severe irritation). The median rating for WADD comfort was 5 (range 4-5). The median rating for SOMNO comfort was 3 (range 1-5 for controls and 2-5 for patients). The median rating for irritation caused by the WADD plaster on the neck was 5 (range 5-5 for controls and 4-5 for patients).

DISCUSSION

Main findings

WADD had very high sensitivity and specificity for detecting apnea in 15 second epochs in a series of breathing and breath-holding exercises in a variety of conditions, including the presence of external background noise, movement and posture. The tolerability of WADD was superior to the portable polysomnography system (SOMNO) during overnight recordings.

WADD had 97·7-99·2% sensitivity to detect instructed apneas and 88.6-99·1% for 10 seconds spontaneous apneas during natural sleep, with similar performance in controls and patients. The WADD also detected all apneas over 30 seconds and there were only 3 over 30 seconds false positives. For short apneas, in most cases, disagreement between the clinician scorer and the WADD were caused by the WADD identifying as apnea epochs that the expert classified as hypopneas.

As expected, the WADD performance was less good when apneas and hypopneas were considered together (minimum sensitivity 77·1%).This is not surprising since the WADD was designed to identify apnea, not hypopneas, and the latter were detected from the transmitted signal which had already been pre-processed for apnea detection. From the table, it can be observed that the degradation of performance was more evident in the controls because the controls had a large number of shorter hypopneas (under 22.5 seconds) which the WADD did not detect properly. In the patients, who often demonstrated apneas, the hypopnea events were longer and these were detected by WADD. Although the lower sensitivity in hypopnea detection might in principle seem problematic if the WADD

was to be used in the context of sleep apnea diagnosis (hypopneas are very common events in sleep labs), it is worth noting that: 1) there is no other reported automatic system that gets anywhere close to this with similar specificity and apnea detection performance; 2) the variations between different sleep labs due to the non-uniform definition of hypopneas already leads to much larger diagnostic variations than the limitation in sensitivity of the WADD;[35-37] 3) assuming the worse case scenario for the WADD, this is that a patient only had hypopneas throughout the night, this reduced sensitivity would be a problem that would translate to non-diagnosis of sleep apnea for patients that with 100% sensitivity would have had a sleep apnea hypopnea index (AHI) between 5 and 6 (i.e very mild cases of sleep apnea). Patients with no sleep apnea, moderate sleep apnea, severe sleep apnea and those with mild sleep apnea with AHI between 6 and 15 would have been rightly diagnosed.

The median difference between the WADD calculated Apnea Hypopnea Index (AHI) and the one obtained by the gold standard was 0 (average=0.7).

SOMNO Performance

The automatic analysis of the SOMNO apparatus, an FDA approved and clinically accepted system, based on assessing apnea from a variety of different sensors, significantly differed from that of the expert marker, with an average sensitivity value of around 14%. The results obtained from the instructed apneas tests also showed that even in the absence of artifacts, apneas were not well detected by the SOMNO system, with an average sensitivity of 37·8%. This demonstrates the need for caution if relying on current automated assessment methods for diagnosing apnea. Whilst performance might be improved by optimizing parameters for individual patients, this is not practical for single overnight recordings or use

as an alerting monitor. The WADD does not require any parameter optimization or subject specific calibration.

The SOMNO system performance improved in the event of indistinct classification of apneas and hypopneas, but was still poor compared with the clinician scorer (54% sensitivity). This sensitivity was at the cost of reduced specificity: for every true hypopnea detected there were approximately four false detections. Overall, the performance of the WADD in hypopnea/apnea combined detection was significantly better than the SOMNO, in sensitivity (77·1% vs 54% if considering the clinician scorer as a gold standard, and 84·1% vs 57·8% otherwise), but also in specificity, as the WADD only detected one false hypopnea epoch for every four true events.

Limitations. Future improvements

The study described in this paper is a small pilot study and hence further more comprehensive clinical evaluation of the technology will be necessary before it can be used. The size of the study was however adequate to assess the potential of the technology; to determine whether the initial performance results in controlled conditions were equivalent to those obtained in real scenarios; and to inform a clinical trial. Based on these positive results it is expected that a fully powered clinical trial, focused on diagnosis rather than on individual event identification, will follow in the future.

The calculation of the sensitivity and specificity has assumed that all apnea events were independent, which for some might not be completely correct. If the data had not been pooled, and taking the expert marker as the gold-standard, in 67% of the subjects the individual apnoea detection sensitivity was 100%. In 77% it was over 90%. In the remaining cases, the drop in sensitivity corresponded always to just one non-detected apnoea shorter

than 15 seconds in the 6 hour night, which is clinically insignificant. The average from the individual sensitivities was 2% higher than the value obtained pooling the data. In terms of specificity 90% of the subjects had values higher than 99%. Two thirds of them were over 99.9%. The average of the individual specificities was identical to the specificity obtained pooling the data.

A different statistical analysis, possibly comparing pooled with non-pooled data will be the subject of investigation when the technology undergoes a larger clinical trial.

The WADD is obviously no substitute to a full night study in a sleep clinic, since it does not provide all the information that a full polysomnography system would. There are advantages and disadvantages to this device with respect to full polysomnography. The WADD can be used to determine the Apnea Hypopnea Index (AHI), which is used in sleep apnea diagnosis to ascertain whether a patient has sleep apnea and to score the severity of the condition. The main advantage is that it can be used for at home assessment or monitoring, and from that point of view it is clearly superior to any of the other existing devices (highly resilient to artifacts, very easy to attach and durable in position, low cost, much more comfortable, and accurate). Considering the restricted resources for sleep clinic referral this device could be a very useful tool to determine at very low cost who should be referred to a specialist centre for full polysomnography. The disadvantage is that there are other parameters that could be used for extra assessment that the device does not measure, such as microarousal or full cardiac activity. Furthermore, the WADD does not allow to assess the hypoxic load or autonomic activation and therefore impact the cardiovascular or stroke risk associated with OSA syndrome.

The WADD device used in the current study relied on wireless transmission to a PC. However changing the PC to a dedicated mobile phone sized receiver poses no technological challenge. A subsequent version that is being developed is smaller (2.4 by 2.4 by 1.2 cm, weighing 7.5 grams) and can operate continuously on hearing aid batteries for over 48 hours. It has a separate dedicated receiver of comparable size to a mobile phone which can be located up to 10 metres from the subject.

Acknowledgments: Eduardo Aguilar-Pelaez for assisting with the data collection at the initial phases of the clinical study and proof reading one of the versions of the manuscript; Pauline Rogers (Statistician, UCL/H Biomedical Research Unit) for carrying out the statistical analysis leading to the number of tested subjects; UCL Business; Epilepsy Society; and Staff on MITU at NHNN.

Sources of Support. UCL Business and Imperial Innovations.

Authors' contributions: John Duncan and Esther Rodriguez- Villegas had the initial idea of developing a wearable breathing monitoring device. Esther Rodriguez-Villegas and Guangwei Chen were the main creators of the novel device carrying out the electronic design which includes both hardware and software. John Duncan contributed through the process helping to define the engineering specifications to meet the clinical need. John Duncan designed the protocol for the clinical study. Esther Rodriguez-Villegas created the documentation for MHRA approval. Jeremy Radcliffe led the clinical studies and blindly marked all the data using the sensors of the SOMNO system as well as the Konica Minolta pulse oximeter. Esther Rodriguez-Villegas and Guangwei Chen compared the results of the data scored by the WADD, SOMNO and blinded expert, and carried out the data analysis. All the authors contributed to the interpretation of data. John Duncan, Jeremy Radcliffe and Esther Rodriguez- Villegas wrote the paper.

Conflict of interest: There are no competing interests.

Data sharing: No additional data available

References

1 Shorvon S, Tomnson T. Sudden unexpected death in epilepsy. *Lancet* 2011;378(9808):2028–38.

2 Mace SE. Acute bacterial meningitis. *Emerg Med Clin North Am* 2008;**26**(2):281–317.

3 Cordero L, Rau R, Taylor D, et al. Enteric gram-negative bacilli bloodstream infections: 17 years' experience in a neonatal intensive care unit. *Am J Infect Control* 2004;**32**(4):189–95.

4 Powers, KS. Acute pulmonary infections. In: Lucking SE, Maffei, FA, Tamburro, RF and Thomas, NJ, editors. Pediatric critical care study guide. London: Springer 2012; 514–534.

5 Moruzzi P, Sarzi-Braga S, Rossi M, et al. Sleep apnea in ischaemic heart disease: differences between acute and chronic coronary syndromes. *Heart* 1999;**82**(3):343–7.

6 Van Vliet G, Deal CL, Crock PA, et al. Sudden death in growth hormone-treated children with Prader-Willi syndrome. *J Pediatr* 2004;**144**(1):129–31.

7 Taylor MB, Grounds RM, Mulrooney PD, et al. Ventilatory effects of propofol during induction of anaesthesia. *Anaesthesia* 1986;**41**(8):816–20.

8 Murphy JJ, Swanson T, Ansermino M, et al. The frequency of apneas in premature infants after inguinal hernia repair: do they need overnight monitoring in the intensive care unit? *J Pediatr Surg* 2008;**43**(5):865-8.

9 Naeem N, Montenegro H. Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest. *Resuscitation* 2005;**67**(1): 13-23.

10 Moon RY, Horne RS, Hauck FR. Sudden infant death syndrome. *Lancet* 2007;**370**(9598):1578–87.

11 Gami AS, Howard DE, Olson EJ, et al. Day-night pattern of sudden death in obstructive sleep apnea, *N Engl J Med* 2005;**352**(12):1206–14.

12 Julien JY, Martin JG, Ernst P, et al. Prevalence of obstructive sleep apnea–hypopnea in severe versus moderate asthma. *J Allergy Clin Immunol* 2009;**124**(2):371–6.

13 Emilsson O, Janson C, Benediktsdóttir B,et al. Nocturnal gastroesophageal reflux, lung function and symptoms of obstructive sleep apnea: Results from an epidemiological survey. *Respir Med* 2012;**106**(3):459–66.

14 Reichmuth K, Austin D, Skatrud JB, et al. Association of sleep apnea and type II diabetes. A population-based study. *Am J Respir Crit Care Med* 2005;**172**(12):1590–5.

15 Hanly PJ. Mechanisms and management of central sleep apnea. *Lung* 1992;**170**(1):1–17.

16 Culebras A. Sleep and neuromuscular disorders. *Neurol Clin* 2005;**23**(4):1209–23.

17 Young T, Palta M, Dempsey J, et al. The occurrence of sleep disordered breathing among middle aged adults. *N Engl J Med* 1993;**328**(17):1230–5.

18 Faroux B. What's new in pediatric sleep? *Pediatr Respir Rev* 2007;**8**(1):85–9

19 Pang KP, Terris DJ. Screening for obstructive sleep apnea; an evidence-based analysis. *Am J Otolaryngol* 2006;**27**(2):112–8.

20 Redline S, Quan SF. Sleep apnea: a common mechanism for the deadly triad--cardiovascular disease, diabetes, and cancer? *Am J Respir Crit Care Med* 2012;**186**(2):123–124.

21 Hirshkowitz M. The clinical consequences of obstructive sleep apnea and associated excessive sleepiness. *J Fam Pract* 2008;**57**(8 Suppl):S9–16.

22 Wittmann V, Rodenstein DO. Health care costs and the sleep apnea syndrome. *Sleep Medicine Rev* 2004;**8**(4):269–279

23 Smith R, Ronald J, Delaive K, et al. What are obstructive sleep apnea patients being treated for prior to this diagnosis? *Chest* 2002;**12**(1):164-172.

24 Leger D, Bayon V, Laaban JP, et al. Impact of sleep apnea on economics. *Sleep Med Rev* 2012;**16**(5):455–62.

25 Flemons WW, Douglas NJ, Kuna ST, et al. Access to diagnosis and treatment of patients with suspected sleep apnea. *Am J Respir Crit Care Med* 2004;**169**(6):668–72.

26 Oliveira MG, Nery LE, Santos-Silva R, et al. Is portable monitoring accurate in the diagnosis of obstructive sleep apnea syndrome in chronic pulmonary obstructive disease? *Sleep Med*. 2012;**13**(8):1033-8.

27 Flemons WW, Littner MR; Rowley JA, et al. Home diagnosis of sleep apnea: a systematic review of the literature. An evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *Chest* 2003;**124**(4):1543–79

28 McDonald F. Avoidable Epilepsy-Related Deaths: The case for prevention. *Epilepsy Bereaved Report*, 2011.

29 Surges R, Sander JW. Sudden unexpected death in epilepsy: mechanisms, prevalence and prevention. *Curr Opin Neurol* 2012;**2**:201–207.

30 Lofsky A. Sleep apnea and narcotic postoperative pain medication: a morbidity and mortality risk. *Anaesthesia Patient Safety Foundation Newsletter*. Summer 2002;**17**:24–25.

31 Bhananker SM, Posner KL, Cheney FW, et al. Injury and liability associated with monitored anesthesia care: A closed claims analysis. *Anesthesiology* 2006;**104**:228–34.

32 Rodriguez-Villegas E, Corbishley P, Duncan JS, et al. A method and apparatus for monitoring respiratory activity. Patent *WIPO;* WO/2007/099314.

33 Corbishley PG, Rodriguez-Villegas E. Breathing detection: towards a miniaturized wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering* 2008; **55**(1):196–204.

34 Rodriguez-Villegas E, Chen G, Aguilar-Pelaez E. Feature characterization for breathing monitor. Patent *WIPO*; WO/2012/007719.

35 Nigro CA, Rhodius EE. Effect of the definition of hypopnea on apnea/hypopnea index. *Medicina* 2003;**63**(2):119–24.

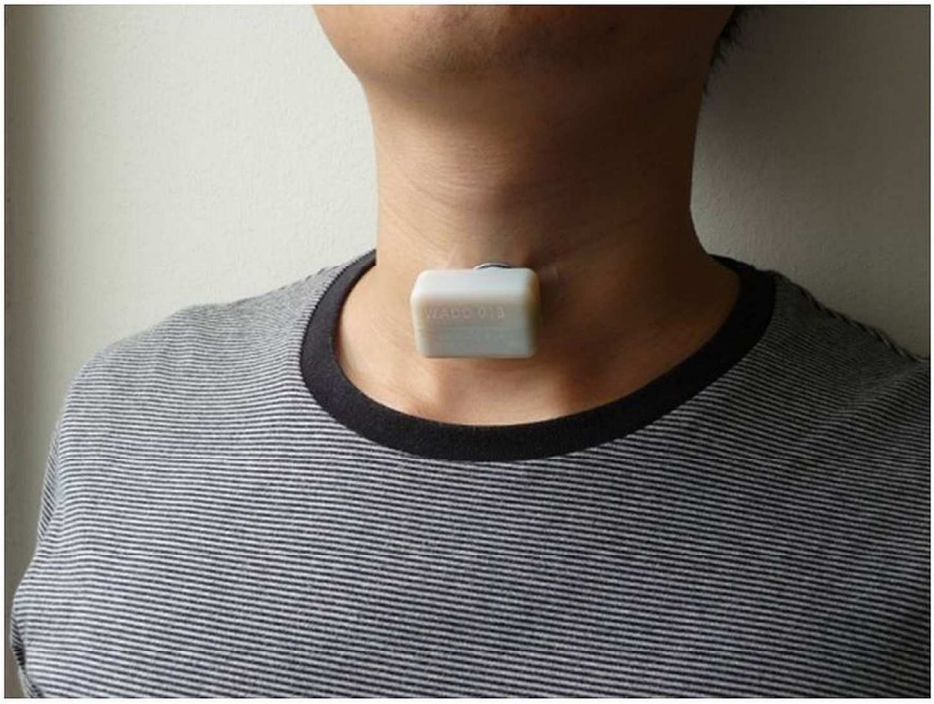
36 Redline S, Kapur VK, Sanders MH, et al. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med* 2000;**161** (2):369–74.

37 Ruehland WR, Rochford PD, O'Donoghue FJ, et al. The New AASM Criteria for Scoring Hypopneas: Impact on the Apnea Hypopnea Index, *Sleep* 2009;**32**(2):150–157.

Figure Legends:

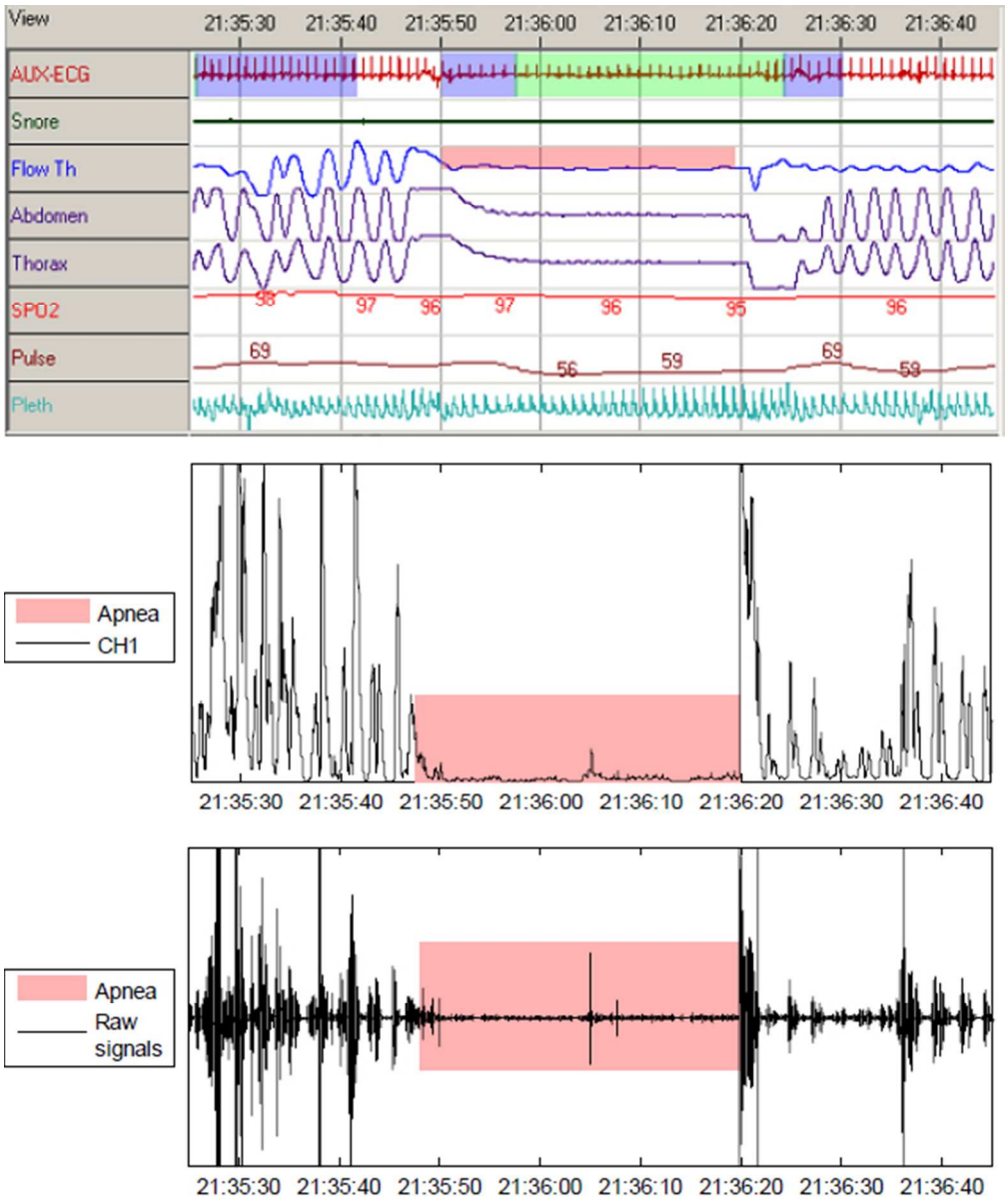
Figure 1: (a) WADD worn by one of the investigators. (b) Subject wearing an existing state of the art ambulatory apnea monitoring system (SOMNO), comprising finger oximetry; oro-nasal flow sensors; thoracic and abdominal expansion bands; and ECG.

Figure 2: Illustration of the SOMNO and WADD output signals showing an apnea event: (top) raw signals from the different SOMNO sensors, (middle) processed WADD signal, (bottom) WADD output signal.



90x69mm (300 x 300 DPI)

90x115mm (300 x 300 DPI)



53x64mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(included, page 1)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(included, page 1)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(included, pages 3-4)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(included, 3-4)**
Methods		
Study design	4	Present key elements of study design early in the paper **(included, pages 5-7)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(included, pages 5-6)**
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants **(included, pages 8-9)**
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case **(NA)**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(included, pages 8-9)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(included, pages 8-10)**
Bias	9	Describe any efforts to address potential sources of bias **(included, pages 8, 10)**
Study size	10	Explain how the study size was arrived at **(included, page 6)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(included, pages 8-10)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(included, pages 6, 14)**
		(*b*) Describe any methods used to examine subgroups and interactions **(NA)**
		(*c*) Explain how missing data were addressed **(page 12)**
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy **(NA)**
		(*e*) Describe any sensitivity analyses **(included)**

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(included, page 6)**
		(b) Give reasons for non-participation at each stage **(NA)**
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(included, page 6)**
		(b) Indicate number of participants with missing data for each variable of interest **(NA)**
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) **(NA)**
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time (NA)
		Case-control study—Report numbers in each exposure category, or summary measures of exposure **(NA)**
		Cross-sectional study—Report numbers of outcome events or summary measures**(NA)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **(included, page 14)**
		(*b*) Report category boundaries when continuous variables were categorized **(included, pages 8-10)**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **(NA)**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **(NA)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(included, pages 11-15)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(included, page 17)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(included, page 17)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(included, page 17)**
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **(included, 19)**

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.