Assessing the sleeping habits of patients in a sleep disorder centre: a review of sleep diary accuracy

Geoffrey Lawrence[1], Rexford Muza[2]

[1]GKT School of Medical Education, King's College London, London, UK; [2]Sleep Disorders Centre, St Thomas' Hospital, London, UK

Contributions: (I) Conception and design: All authors; (II) Administrative support: R Muza; (III) Provision of study materials or patients: R Muza; (IV) Collection and assembly of data: G Lawrence; (V) Data analysis and interpretation: G Lawrence; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Rexford Muza. Sleep Disorders Centre, St Thomas' Hospital, Guy's Hospital Sleep Disorders Centre 3rd floor, Nuffield House Great Maze Pond, London SE1 9RT, UK. Email: Rexford.Muza@gstt.nhs.uk.

Background: Excessive daytime sleepiness (EDS) is a complaint common to many aspects of medicine. There are primary and secondary causes for EDS, with secondary causes including a large number of common conditions. Primary causes, such as narcolepsy, are much rarer. When assessing for primary hypersomnia, restricted or fragmented sleep must be ruled out. This process involves assessment of sleeping habits using a sleep diary and/or actigraphy. Clinicians are suspicious of the accuracy with which patients use the former. This review aims to evaluate the accuracy of a sleep diary study against the 'objective gold standard' actigraphy report.

Methods: Data from 35 patients at a Sleep Disorder Centre who underwent both a sleep diary and actigraphy study for suspected primary hypersomnia in 2016 was collected. Mean values of four variables were calculated: 'time of lights out', 'time to fall asleep', 'time of waking' and 'sleep time'. The 'similarity' was assessed. This was a term defined in three different ways: if sleep diary values are accurate to within 20, 30 and 60 min respectively. Percentage 'similarity', mean time differences and standard deviations (SDs) were calculated for each variable. A paired *t*-test was also performed to assess the significance of the time differences between the two modalities.

Results: Least accurate was 'sleep time', with 14.7%, 23.5% and 58.8% of patients within 20, 30 and 60 min of the actigraphy respectively. Mean time difference for this variable was 66 min (versus 33, 15 and 22). 'Time to fall asleep' was most accurate, with 76.5%, 82.4% and 100% 'similarity' respectively.

Conclusions: The clinically acceptable accuracy has no universal definition, so clinicians must use experience and reasoning to determine this level to interpret this data. The review suggests that some variables are entered with high accuracy, and the diary is low cost and adds subjective information that cannot be gathered from actigraphy. Therefore, use is recommended to continue alongside actigraphy.

Keywords: Excessive daytime sleepiness; sleep diary; actigraphy

Introduction

Excessive daytime sleepiness (EDS) refers to an increased propensity and subjective compulsion to sleep, as well as a tendency to take involuntary naps and suffer sleep attacks when sleep is inappropriate (1). EDS is a common complaint in many areas of medicine. Around 20% of the general population describe being so sleepy that it interferes with their daily activities on at least a few days per week or more; this figure rises to 43% when relating to a few days a month or more (2).

Sleepiness during the day can have a significant effect on

an individual and a society. On the individual level, it leads to reduced levels of performance and graduation rates at school or work, as well as impairments of concentration, memory and mood (2-4). At a societal level, more insidious consequences can occur. One source reported that annually in the US over 50,000 motor vehicle incidents can be attributed to sleepiness while driving (2), while another source puts this figure at over 100,000 (4). These incidents result in over 71,000 personal injuries and around 1,550 deaths each year (5,6). This is not surprising when it is considered that a 1999 study found alcohol-impaired individuals at the legal limit for driving and sleep-disordered individuals performed equally badly in reaction time tests (2,7).

Primary hypersomnias have a central origin, and include conditions such as narcolepsy and idiopathic hypersomnia. Secondary hypersomnias, as the name suggests, are characterised by EDS secondary to another condition affecting sleep, such as obstructive sleep apnoea hypopnoea syndrome (OSAHS), periodic limb movement disorder (PLMD) or circadian rhythm disorders (CRD). Sleep deprivation is the most common cause of secondary hypersomnia (1,2,4). A large study in 2000 found that 1/3rd of Americans slept on average less than 6.5 h each night (2). Commonly, EDS resulting from sleep deprivation can be described as a behaviourally-induced insufficient sleep syndrome (BIISS). This is a syndrome of chronic, voluntarily restricted sleep often found in individuals who work long and varied shifts in their job, or restrict their sleep in favour of another activity (e.g., watching television) (8). A vast number of sleep, medical and psychiatric conditions can lead to EDS through a combination of sleep deprivation and fragmentation. Examples include chronic pain, arthritis and urinary dysfunction (2,9). Furthermore, 63% of mental health patients report significant sleep loss (10,11); notably, around three quarters of depressed patients will have sleep symptoms (12). In the assessment of primary hypersomnias it is therefore imperative that secondary hypersomnias are excluded. A number of assessments and tests are usually performed in specialist sleep centres in order to help make a diagnosis of a primary hypersomnia; these include use of the Epworth sleepiness scale (ESS) (10,13), nocturnal polysomnography (PSG), the multiple sleep latency test (MSLT), and measurement of the hypocretin level in the cerebrospinal fluid (14). A sleep diary asks the patient to record information about their sleeping habits, but many sleep physicians are concerned that the diary does not provide an accurate reflection of this. Actigraphy involves the patient wearing an 'actigraph' or 'actiwatch' on the wrist. This will continuously measure limb movement and light intensity, from which a computerised scoring system is used to determine periods of sleep and wake (15-17). The patient presses a button on the actiwatch that indicates the time they are going to bed and waking. Carrying out a 2-week sleep diary with or without actigraphy is routine prior to PSG/MSLTs.

Aim

The aim of this review is to determine the accuracy of the sleep diary in assessing an individual's sleep habits. Sleep diary data will be compared to that from actigraphy, which is taken here as an objective measure and 'gold standard'.

Methods

The data from 35 patients who had completed both a sleep diary and actigraphy was collected. A number of different variables of the diary and the matching or similar variables from the actigraphy report were collected. All patients had used actigraphy in 2016, and were patients from a tertiary referral sleep clinic in a metropolitan area. A total of 353 adult patients were assessed against the following exclusion criteria:

   (I)   No/incomplete sleep diary;
   (II)   Diagnosis of condition other than idiopathic hypersomnia or narcolepsy;
   (III) No/incomplete actigraphy report;
   (IV)   Patient not on EPR;
   (V)   Unsure diagnosis from information on electronic patient record (EPR).

Three hundred and eighteen patients were excluded based on these criteria. The vast majority were excluded due to either the lack of a complete sleep diary on the EPR system or local computer files, or due to non-availability of diagnosis or indication.

Four variables were specifically identified as being easily compared. These were:

   ❖   Time of lights out ('lights out');
   ❖   Time to fall asleep;
   ❖   Time of waking up ('woke up');
   ❖   Total sleep time ('sleep time').

The times of lights out and waking up were variables in both the sleep diary and the actigraphy, and therefore directly comparable. The time taken to fall asleep was estimated in the sleep diary, and was taken from the

Table 1 A table showing the 'similarity' between a patient's sleep diary and actigraphy mean values for the 4 different variables

Similarity	Lights out (%)	Time to fall asleep (%)	Woke up (%)	Sleep time (%)
Number 'similar' within 20 min	21 (60.0)	26 (76.5)*	26 (74.3)	5 (14.7)*
Number 'similar' within 30 min	27 (77.1)	28 (82.4)*	30 (85.7)	8 (23.5)*
Number 'similar' within 60 min	28 (80.0)	34 (100.0)*	33 (94.3)	20 (58.8)*

*, for these variables, one patient did not fill in the sleep diary section, and therefore this measure is out of 34 rather than 35.

Table 2 A table displaying the mean, SD and P values of the time differences between sleep diary and actigraphy report for each variable

Analysis	Lights out	Time to fall asleep	Woke up	Sleep time
Mean time differences (min)	32.5	14.9	21.6	66.3
SD	61.1	12.1	46.8	47.3
P value	0.411	0.001	0.716	0.001

SD, standard deviation.

actigraphy report by calculating the difference between the 'lights out' and 'fell asleep' times. The sleep time was estimated by the patient in the sleep diary, while the actigraphy recorded both 'estimated' and 'actual' sleep times; in this review the comparison is made with the 'actual' sleep time. Depending on the length of the sleep diary and actigraphy studies, a number of values were collected for each variable. For example, in a 1 week study there are seven 'lights out' times noted. These were all recorded and a mean value calculated. These mean values were then compared against those from the actigraphy study to assess similarity.

Mean values were calculated for the times of each of the four variables compared. For each of the 35 patients included in the data collection, the mean value from the sleep diary was compared on 'similarity' with the equivalent mean from the actigraphy report. 'Similarity' was a concept that was given three different definitions in the evaluation of this data. The value from the sleep diary was defined as 'similar' if the:

(I)   Sleep diary value was within 20 min of the actigraphy value;
(II)   Sleep diary value was within 30 min of the actigraphy value;
(III)   Sleep diary value was within 60 min of the actigraphy value.

Out of the 35 patients, the number of times that the mean value from the sleep diary was 'similar' to that of the actigraphy report was noted, and a 'percentage similarity' calculated.

As well as this, the time differences between the sleep diary and actigraphy for the four variables were compared; a mean value and SD was calculated for each variable and a paired t-test used to assess the statistical significance of the time differences.

Results

The datasets were evaluated for each of the variables based on each definition of 'similarity', therefore giving three similarity percentages for each variable that was compared. This is shown in *Table 1*.

The data presented in *Table 1* shows the similarity of the data from the sleep diaries to that of the actigraphy reports of 35 patients, based on each different definition of 'similar'. The results indicate the number of times that a value from the sleep diary is within a certain time of the correct value; the correct value in this case assumed to be that from the actigraphy report. For example, in 77.1% of cases a value recorded by a patient in their sleep diary for the time of lights out is within 30 min of the correct time of lights out.

The time difference between these mean values was also calculated for each patient for each variable, along with a paired t-test. The mean, SDs and P values for the time differences were then calculated, and are displayed in *Table 2* correct to 3 significant figures.

The results show a high level of variability in the time differences recorded by different patients within the 'lights out', 'woke up' and 'sleep time' categories, with SDs of

S180

Lawrence and Muza. A review of sleep diary accuracy

Table 3 A table displaying the mean and SD of the time differences for all 4 categories having excluded the data from patient 3

Analysis	Lights out	Time to fall asleep	Woke up	Sleep time
Mean time differences (min)	23.4	14.5	14.1	65.67
SD	29.4	12.0	15.3	47.9
P value	0.219	0.001	0.177	<0.001

SD, standard deviation.

Table 4 A table summarising the advantages and disadvantages of using a sleep diary and actigraphy to assess patients' sleeping habits

Comparison	Sleep diary	Actigraphy
Advantages	Cheap and widely available; easy to use and fill out; allows patients to give extra information in the form of 'other comments'	More objective than sleep diary; less reliant on patient compliance; easy to use
Disadvantages	Subjective; if patient non-compliant then not useful; reliant on accuracy of patient memory and therefore high interpatient variability in usefulness; often needs actigraphy report to validate data	More expensive and less widely available; still some level of subjectivity/reliance on patient compliance due to need to press button; actigraphy report can be difficult to interpret, e.g., movement and light data

61.1, 46.8 and 47.3 respectively. However, in the 'time to fall asleep' category a much lower level of SD around the mean was found (12.1), as well as a mean time difference of only 14.9 min. Interestingly, this time difference was found to be statistically significant, while the larger differences in mean time had P values greater than 0.05. The mean time difference in the 'sleep time' variable is over double that of the next-highest variable, standing at over 66 min. The P value of 0.001 indicates that the difference between the sleep diary and actigraphy is statistically significant. This implies that this variable is generally not accurately entered into the sleep diary by patients, as well as this being much less accurate than the other variables assessed.

One patient included in the data presented in *Tables 1,2* was deemed an outlier. The dataset was also calculated after removing this patient's data in order to see the effect this had on the results. This is shown in *Table 3*.

Discussion

Current practice

There are a number of advantages and disadvantages of the use of either method of assessment of an individual's sleep habits. These are summarised in the *Table 4*.

Clearly, neither technique is perfect. The current practice at the sleep centre in this review is for individuals to undergo both sleep diary and actigraphy studies simultaneously. When used together, some potential gaps in the data collected can be filled, such as additional comments added by patients in the sleep diary. These may be invaluable to making a properly informed diagnosis.

Another possible reason for this practice is the general suspicion felt by many sleep clinicians that patients do not accurately fill in a sleep diary in many cases. As reported by a number of sleep clinicians, patients seem particularly inaccurate in estimating their length of sleep time. In some cases it is suspected that a patient may fill in the entire diary on the train to an appointment, leading to inaccuracies due to problems with recall and a tendency to just put rough guesses to complete the document. In these cases, the data in the sleep diary is unlikely to be useful and clinical time and resources are wasted

The use of actigraphy together with the sleep diary allows the information in the sleep diary to be validated to a certain extent if it matches up well. However, there is an argument that the actigraphy report provides all the information required, and at a higher level of accuracy; thus, use of the sleep diary may be unnecessary, especially if patients are unlikely to fill in the sleep diary properly.

Interpretation of results

This study of 35 patients found evidence to suggest that the majority of these patients had a relatively high level

of concurrence between the two modalities in 3 of the 4 categories assessed. Using the test of 'similarity', as defined in patients and methods, it was found that when the acceptable level of similarity was taken as within 30 min, 81.7% of patient's data was 'similar' across these three categories. When the acceptable time difference was increased to 60 min, this figure rises to 91.3%. Interestingly, despite the fact that 'time to fall asleep' had the smallest mean time difference, it was found to be statistically significant (P=0.001) using a paired *t*-test, while the larger mean differences of 'lights out' and 'woke up' were not significant (P=0.219 and 0.177 respectively). This may be due to a greater effect of the outlier on the latter two variables, and the likely non-normal distribution of this data hinted at by the large SDs.

However, the fourth variable—sleep time—was entered to a much poorer level of accuracy. Only 23.5% of patients had sleep diary values within 30 min of actigraphy, and this rose only to 58.8% within 60 min. The mean time of 66 min was over double that of the other variables and a high SD (47.3) suggests a wide spread of values around this mean. The presence of an outlier helps to show this was not due to a small number of patients skewing these values. *Table 3* shows the mean time difference, SD and P values following the removal of this data.

The removal of this outlier and the subsequent reduction in the mean and SD values in the 'lights out' and 'woke up' variables shows that this data contributed significantly to the high values in these two columns. However, the values for 'sleep time' have changed a negligible amount, suggesting that this low level of accuracy was common amongst many of the patients in this review. From the raw data it is also seen that in the majority of cases (24 out of 34), the patients overestimated their sleep time, often by a significant amount.

The interpretation of the data in this review is entirely dependent on what level of accuracy is clinically acceptable. There is a wide variation in the levels of similarity depending on the definition used, and the judgement of what level is clinically acceptable should be made by experienced sleep clinicians. For example, taking 30 min as the acceptable level of accuracy across three of the four categories (excluding 'sleep time') over 80% of data is similar. This drops to around 70% if 20 min is acceptable, and is over 90% if 60 min is. However, for 'sleep time' the similarity levels are much lower: at 14.7%, 23.5% and 58.8% for each increasingly lenient definition.

Clearly, patients are able to fill in different variables to widely fluctuating levels of accuracy. As shown, patients were significantly less accurate in entering sleep time values than any other category, and *Table 3* indicates that this was across all patients rather than limited to one of two within the sample. The actigraphy report variable of 'actual sleep time' calculates this value using movements recorded, which are translated to sleep-wake scores using a computerised scoring system (17). It can then distinguish between states of sleep and wakefulness on the basis of movement during periods that the patient may assume to be sleep. This is likely to be a reason contributing to the very low coherence between the sleep diaries and actigraphy reports for this variable.

There were a number of limiting factors to this review. The first is that patients who did not complete the sleep diary were excluded from the review. Therefore, the review excluded possibly the most challenging patient group for the use of sleep diaries, so it is difficult to make a fully informed judgement on the overall efficacy of their use. Furthermore, the exclusion criteria used meant that there was a small sample size, limiting the range and validity of conclusions that can be drawn.

Another issue was that the variables measured by the sleep diary and the actigraphy were in a number of cases slightly different from one another. This gave limited scope for direct comparison between the two modalities, and again limits the conclusions that can be drawn.

Subjectivity was inherently present in the collection of the data, since patients differed slightly in how they filled in the sleep diary. For example, one patient may fill in that it took them 15–30 min to get to sleep, while another may put 20 min. For the first patient, the middle value of 22.5 min would be taken, despite the fact that these two patients may have fallen asleep in the same time. In this way, error was introduced into the data.

Finally, actigraphy as a method of assessing sleeping habits is taken in this review to be the 'gold standard'. This assumes that the actigraphy values are correct, giving the exact times that a patient fell asleep, for example. There is a level of subjectivity in the use of actigraphy, since the patient must press a button to indicate when they went to bed and woke up, and therefore scope for error. Furthermore, Sadeh in 2011 noted that while actigraphy had a good level of validity in assessing sleep in healthy populations, this in '…*individuals with poor sleep or other sleep-related disorders is more questionable*…', and that '…*the most problematic validity issue*

S182

Lawrence and Muza. A review of sleep diary accuracy

is the low specificity of actigraphy in detecting wakefulness within sleep periods reported with certain devices…' (17). This relates directly to the 'actual sleep time' category of the actigraphy, which was used to compare with the 'estimated sleep time' of the diary.

Conclusions

The data collected from this sample of 35 patients suggests that there is a high level of variation in the accuracy of a sleep diary compared to the 'gold standard' of actigraphy. For three of the four variables assessed ('lights out', 'time to fall asleep' and 'woke up'), a reasonable level of accuracy was demonstrated. However, the level of accuracy was much poorer for the 'sleep time' variable, with the difference found to be statistically significant in this and another variable. The level of clinically acceptable accuracy of a sleep diary has no universal guidelines, with much of the assessment down to the clinical judgement of the clinician in each individual case. It is therefore difficult to draw an obvious conclusion from this data without this knowledge and experience, and is important that clinicians review the data presented and interpret it cautiously.

Since the accuracy level in most categories reviewed is reasonable in the opinion of the authors, it is suggested that the use of sleep diaries continues as it has the potential to give additional information to actigraphy which can be important in making a diagnosis, as well as being a very low cost tool kit. When interpreting a diary, clinicians should bear in mind the low level of accuracy patients seem to have in reporting 'sleep time', as well as the limitations of this review.

In the future, it is important to review how regularly a sleep diary is fully filled out relative to how often it is either not filled out or is unusable. This information, along with that in this review, would give a more complete look into the efficacy of sleep diary use in assessing patients' sleeping habits.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Bittencourt LR, Silva RS, Santos RF, et al. Excessive daytime sleepiness. Rev Bras Psiquiatr 2005;27 Suppl 1:16-21.
2. Guilleminault C, Brooks SN. Excessive daytime sleepiness: a challenge for the practising neurologist. Brain 2001;124:1482-91.
3. Pagel JF, Forister N, Kwiatkowki C. Adolescent sleep disturbance and school performance: the confounding variable of socioeconomics. J Clin Sleep Med 2007;3:19-23.
4. Pagel JF. Excessive daytime sleepiness. Am Fam Physician 2009;79:391-6.
5. National Sleep Foundation. State of the states report on drowsy driving. Available online: http://drowsydriving.org/docs/2007%20State%20of%20the%20States%20Report.pdf
6. Mahowald MW. Eyes wide shut. The dangers of sleepy driving. Minn Med 2000;83:25-30.
7. Powell NB, Riley RW, Schechtman KB, et al. A comparative model: reaction time performance in sleep-disordered breathing versus alcohol-impaired controls. Laryngoscope 1999;109:1648-54.
8. Hauri PJ. The sleep disorders (eBook). USA: National Sleep Foundation, 1977.
9. Chokroverty S. Sleep disorders in other medical disorders. In: Chokroverty S, editor. Sleep disorders medicine: Basic science, technical considerations and clinical aspects. 2nd edition. Boston: Butterworth-Heinemann, 1999:587-617.
10. Kryger M, Avidan A, Berry R. Atlas of clinical sleep medicine. 2nd edition. Philadelphia: Elselvier, 2014.
11. McCall C, Shapiro C, McCall W. Sleep and psychiatric disease. In: Kryger M, Avidan A, Berry R, editors. Atlas of clinical sleep medicine. 2nd edition. Philadelphia: Elselvier Saunders 2014:364-9.
12. Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. Dialogues Clin Neurosci 2008;10:329-36.
13. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540-5.
14. NHS Choices. Narcolepsy. Available online: http://www.nhs.uk/Conditions/narcolepsy/Pages/Introduction.aspx
15. Kryger M, Roth T, Dement W. Principles and practice of

sleep medicine. 6th edition. Philadelphia: Elsevier 2016.
16. Stone K, Ancoli-Israel S. Actigraphy. In: Kryger M, Roth T, Dement W, editors. Principles and practice of sleep

medicine. 6th edition. Philadelphia: Elsevier, 2016:1671-8.
17. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 2011;15:259-67.