J. Sleep Res. (2000) **9,** 399–401

Letter to the Editor

The multiple sleep latency test and Epworth sleepiness scale in the assessment of daytime sleepiness

Dr Murray Johns' recent article argues that the Epworth Sleepiness Scale (ESS) not only costs about 1000 times less than the Multiple Sleep Latency Test (MSLT) but also serves as a superior gold standard measure of excessive daytime sleepiness (Johns, 2000). The article provides no new data but relies on published ESS and MSLT results from normal and narcoleptic subjects to calculate sensitivities and specificities of the ESS and MSLT for these states, which are assumed to represent fully alert and pathologically sleepy individuals. Receiver-operator curves then show that the ESS assigns subjects to their correct groups with nearly perfect accuracy, whereas the MSLT does not perform as well.

Unfortunately, these analyses are flawed. The narcoleptics were defined, in large part, by their subjective complaint of excessive daytime sleepiness (American Sleep Disorders Association 1997; Mitler *et al*. 1998; US Modafinil in Narcolepsy Multicentre Study Group 1998). The normal subjects' MSLT data were derived from a study in which no subject with subjective daytime sleepiness could be included (Levine *et al*. 1988). Normal subjects' ESS data (Johns and Hocking 1997) were taken from the 72 (22%) of 331 Australians who were, as described by Johns, 'selected by strict criteria derived from a detailed sleep questionnaire' (Johns 2000); this sample, defined by self-report, could be labeled 'super subjective normals'.

The result is that in John's re-analysis, the normal subjects had no subjective sleepiness and the narcoleptics did have subjective sleepiness, by definition. Subjects were grouped more by virtue of their subjective sleepiness than by objective criteria. Johns then showed that the subjectively derived ESS assigned persons to their subjectively defined groups more accurately than did the objective MSLT. The ESS previously has been shown to correlate well with patients' own perceptions of their sleepiness, and to have weak or no correlation with objective measures of sleepiness (Chervin and Aldrich 1999): Johns' recent analysis only seems to confirm these observations.

Correspondence: Ronald D. Chervin, MD, MS, Michael S. Aldrich Sleep Disorders Laboratory, University Hospital 8D8702, Box 0117, 1500 E. Medical Center Dr, Ann Arbor, MI 48109–0117, USA. Tel.: (1) 734 647 9064; e-mail: chervin@umich.edu

Several additional problems with the ESS are not addressed in Johns' article. Increasing evidence suggests that in the assessment of sleepiness, the ESS is subject to undesirable confounding variables, including gender (Chervin and Aldrich 1999), psychological influences (Olson *et al*. 1998), and subjective perception of fatigue, tiredness, and lack of energy (Chervin, 2000a). Although Johns repeatedly argues, based on face validity, that the ESS measures sleep propensity in eight specific situations rather than just one (like the MSLT) (Johns 1991, Johns 1993; Johns 1994; Johns 1998; Johns 2000) he has provided no criterion validity to substantiate this argument. In one study that did test his hypothesis, subjective responses to the ESS item that asks about 'lying down to rest in the afternoon when circumstances permit' failed to show any robust association with objective measures in this specific situation, namely the afternoon naps of MSLTs (Chervin *et al*. 1997).

Finally, Johns' recent article was somewhat selective in its review of existing literature. The largest existing studies of sleep apneics assessed with both MSLTs and Epworth scales showed no statistically significant associations between the two measures (Chervin and Aldrich 1999; Benbadis *et al*. 1999). Some studies suggest that apnea severity as determined by polysomnography is associated with MSLT-measured sleepiness but not with ESS scores (Chervin *et al*. 1997; Chervin and Aldrich 1999; Kingshott *et al*. 1995). A recent study of 1824 individuals did show a highly significant – but impressively weak – association between apnea severity and ESS scores (Gottlieb *et al*. 1999). For example, the subjects with little or no sleep apnea had a mean ESS score of 7.2 ± 4.3 while those with the most severe apnea had a mean score of 9.3 ± 4.9.

Johns is correct in writing that the MSLT is unlikely to be a perfect gold standard, but in that respect the test is similar to many other medical gold standards. His point is also well taken that the 5 and 10 minute 'rule of thumb' for MSLT interpretation should not be misused. Strict cut-points on a continuous unimodal measure will almost always serve a patient poorly if results are not carefully integrated and weighed with data derived from the patient's medical history (Chervin, 2000b). Neither the MSLT nor the ESS have been well-validated against objective sleepiness-related outcomes of importance to patients, such as motor vehicle crashes or work performance. Until those important data become available,

an ESS may indeed cost 1000 times less than an MSLT, but those who seek this savings may end up getting what they paid for.

REFERENCES

American Sleep Disorders Association. *International Classification of Sleep Disorders, Revised: Diagnostic and Coding Manual*. American Sleep Disorders Association, Rochester, 1997.

Benbadis, S. R., Mascha, E., Perry, M. C., Wolgamuth, B. R., Smolley, L. A. and Dinner, D. S. Association between the Epworth Sleepiness Scale and the Multiple Sleep Latency Test in a clinical population. *Ann. Intern. Med.,* 1999, 16: 289–292.

Chervin, R. D. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. *Chest,* 2000a, 118: 372–379.

Chervin, R. D. Use of clinical tools and tests in sleep medicine. In: M. H. Kryger, T. Roth and W. C. Dement (eds), *Principles and Practice of Sleep Medicine* 3rd edn. W.B. Saunders Company, Philadelphia, 2000b: 535–546.

Chervin, R. D. and Aldrich, M. S. The Epworth Sleepiness Scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology,* 1999, 52: 125–131.

Chervin, R. D., Aldrich, M. S., Pickett, R. and Guilleminault, C. Comparison of the results of the Epworth Sleepiness Scale and the multiple sleep latency test. *J. Psychosom Res.,* 1997, 42: 145–155.

Gottlieb, D. J., Whitney, C. W., Bonekat, W. H., Iber, C., James, G. D., Lebowitz, M., Nieto, F. J. and Rosenberg, C. E. Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. *Am. J. Resp. Crit. Care Med.,* 1999, 159: 502–507.

Johns, M. W. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep,* 1991, 14: 540–545.

Johns, M. W. Daytime sleepiness, snoring, and obstructive sleep apnea: the Epworth Sleepiness Scale. *Chest,* 1993, 103: 30–36.

Johns, M. W. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep,* 1994, 17: 703–710.

Johns, M. W. Rethinking the assessment of sleepiness. *Sleep Med. Reviews,* 1998, 2: 3–15.

Johns, M. W. Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the Epworth sleepiness scale: Failure of the MSLT as a gold standard. *J. Sleep Res.,* 2000, 9: 5–11.

Johns, M. W. and Hocking, B. Daytime sleepiness and sleep habits of Australian workers. *Sleep,* 1997, 20: 844–849.

Kingshott, R. N., Sime, P. J., Engleman, H. M. and Douglas, N. J. Self assessment of daytime sleepiness: patient versus partner. *Thorax,* 1995, 50: 994–995.

Levine, B., Roehrs, T., Zorick, F. and Roth, T. Daytime sleepiness in young adults. *Sleep,* 1988, 11: 39–46.

Mitler, M. M., Walsleben, J., Sangal, R. B. and Hirshkowitz, M. Sleep latency on the maintenance of wakefulness test (MWT) for 530 patients with narcolepsy while free of psychoactive drugs. *Electroencephalogr. Clin. Neurophysiology,* 1998, 107: 33–38.

Olson, L. G., Cole, M. F. and Ambrogetti, A. Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J. Sleep Res.,* 1998, 7: 248–253.

U. S. Modafinil in Narcolepsy Multicenter Study Group. Randomized trial of modafinil for the treatment of pathological somnolence in narcolepsy. *Ann. Neurol.,* 1998, 43: 88–97.

Ronald D. Chervin
Sleep Disorders Center,
Department of Neurology,
University of Michigan,
Ann Arbor, Michigan, U.S.A.

Reply

I welcome the opportunity to respond to the above letter of Dr Chervin about my paper, 'Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the Epworth sleepiness scale: Failure of the MSLT as a gold standard' (Johns 2000). In this paper I reported evidence, based on receiver operator characteristic curves, that the ESS is more accurate than either the MSLT or the MWT in distinguishing normal from abnormal daytime sleepiness in the sense of a subject's sleep propensity in daily life.

Dr Chervin claims that the groups of narcoleptics and normal subjects used in my assessment of the sensitivity and specificity of the ESS were selected mainly on the basis of their levels of 'subjective sleepiness', which is what he thinks the ESS measures. He argues therefore that for me to use the ESS as a measure of 'subjective sleepiness' to distinguish groups of subjects who were selected because they differ in that regard, is not a valid test of the ESS. Of course, he would be right if his assumptions were valid, but they are not.

Dr Chervin fails to distinguish 'subjective sleepiness' from sleep propensity, which is what the MSLT and MWT measure objectively and the ESS measures subjectively. He confuses the method of measurement with the nature of what is to be measured. I believe 'subjective sleepiness' refers to the presence/absence or the intensity of a set of feelings and symptoms that accompany the drowsy state, measured, for example, by the Karolinska sleepiness scale. This is not what the ESS measures (Johns 1998). Use of the term 'sleepiness' to mean any of several different things is, in my opinion, a cause of much of the present confusion in this field. Dr Chervin is also wrong in his claim that my control subjects, who provided normal ESS scores, were selected on the basis of an absence of complaints about 'subjective sleepiness'. In fact, they were selected, post hoc, solely on the basis of their reported sleep characteristics, without reference to their sleepiness, as follows: their usual sleep quality was 'good' or 'very good'; they did not snore or snored only occasionally; they were not reported to stop breathing or to make choking noises during sleep; they seldom had difficulty in falling asleep initially and usually took less than 30 min to do so; they did not recall being awake more than twice per night, if at all, and did not have difficulty going back to sleep. The mean duration of their usual sleep period (which was not a selection criterion) was 7 h 37 ± 54 min (SD), with a range in different subjects from 5 h 25 min to 11 h 30 min (Johns and Hocking 1997). Quite what Dr Chervin means when he calls these subjects 'super subjective normal' I do not know.

The above criticism may be more appropriately directed at the validity of the equivalent MSLT results because the

Correspondence: Murray W. Johns, Epworth Sleep Centre, Epworth Hospital, Melbourne, Vic. 3121, Australia. Tel.: (613) 94271849; fax: (613) 94278522; e-mail: sleepsda@entemet.com.au