ORIGINAL ARTICLE

Test–Retest Reliability of the Multiple Sleep Latency Test in Central Disorders of Hypersomnolence

Régis Lopez, MD, PhD[1,2,3]; Anis Doukkali, MD[2]; Lucie Barateau, MD[1,2,3]; Elisa Evangelista, MD[2,3]; Sofiene Chenini, MD[2]; Isabelle Jaussent, PhD[3]; Yves Dauvilliers, MD, PhD[1,2,3]

[1]Department of Neurology, National Reference Centre for Orphan Diseases, Narcolepsy, Idiopathic Hypersomnia and Kleine-Levin Syndrome, Gui-de-Chauliac Hospital, Montpellier F-34000, France; [2]Department of Neurology, Sleep Disorders Center, Gui-de-Chauliac Hospital, Montpellier F-34000, France; [3]Inserm U1061, Montpellier F-34000, France

Study Objectives: To assess the test–retest reliability of the polysomnography–multiple sleep latency test (PSG–MSLT) diagnostic classification and measures and to study the determinants of its variability in patients with narcolepsy type 1 (NT1) or with noncataplectic central disorders of hypersomnolence (NCHS): type 2 (NT2), idiopathic hypersomnia (IH), and unspecified hypersomnolence (unspecified excessive daytime sleepiness [UnsEDS]).

Methods: PSG–MSLT in drug-free conditions was administered twice (median interval of 1.9 years) in 22 patients with NT1 (10 males, median age 31.2 years) and 75 patients with NCHS (32 males, median age 25.7 years).

Results: At the first PSG–MSLT, patients with NCHS were classified as having NT2 (22.7%), IH (26.7%), or UnsEDS (50.6%). A positive PSG–MSLT was confirmed in 72.7% of NT1. The classification consistency at retesting was significantly lower for the NT2 (47.1%), IH (25.0%), and UnsEDS (42.1%) categories than NT1 (81.3%). The between-test mean sleep latency (MSL) variability was significantly different in NT1 and NCHS, with higher changes in NT2 and lower in NT1. A longer test–retest interval was associated with improved MSL and MSLT normalization. Between-test variations in SOREMP number were associated with changes in nocturnal REM sleep parameters and MSL. No association was found with the clinical decision to repeat the evaluation or the disease clinical course.

Conclusion: The PSG–MSLT measures and classification are not stable in patients with NCHS, with frequent diagnostic changes, particularly for NT2 and IH, compared with NT1. MSLT needs to be repeated at regular intervals to confirm a stable hypersomnia and provide an accurate diagnosis of NT2 and IH.

Keywords: narcolepsy, idiopathic hypersomnia, hypersomnolence, multiple sleep latency test, test–retest reliability.

Statement of Significance

Although the multiple sleep latency test (MSLT) is the gold standard for diagnosing central disorders of hypersomnolence, we found a poor test–retest replication of MSLT results in patients with initial diagnosis of narcolepsy type 2 (NT2) and idiopathic hypersomnia (IH) compared with NT1. This lack of repeatability is associated with the variability of the MSLT measures at the second testing, whereas the clinical decision to repeat the MSLT and the disease course do not have any significant influence. Clinical features need to be reassessed and the MSLT needs to be repeated at regular intervals to confirm a stable hypersomnia and to identify homogeneous groups of patients with a sufficiently accurate diagnosis of NT2 or IH for the identification of potential NT2 and IH biomarkers that are currently missing.

INTRODUCTION

Central disorders of hypersomnolence are rare and disabling sleep pathologies that include narcolepsy type 1 (NT1), type 2 (NT2), and idiopathic hypersomnia (IH).[1] NT1 is characterized by excessive daytime sleepiness (EDS) and cataplexy, frequently associated with hypnagogic hallucinations and sleep paralysis, and is defined by cerebrospinal fluid (CSF) hypocretin-1 (Hcrt-1)/orexin-A deficiency.[2] NT1 is usually easy to diagnose, because cataplexy is a pathognomonic symptom of the disease[3] and Hcrt-1 is a highly specific and sensitive diagnostic biomarker. Currently, CSF Hcrt-1 level below 110 pg/mL is the most accurate measurement to confirm NT1 diagnosis, especially in the case of doubtful cataplexy. In contrast, there is no pathognomonic symptom or biomarker for NT2 and IH.[4,5] Neurochemical studies on the CSF levels of monoamine metabolites in noncataplectic central disorders of hypersomnolence (NCHS) remain inconclusive, and the CSF Hcrt-1 levels are consistently normal, as well as those of histamine and tele-methylhistamine, in most cases.[6,7] A recent study found that CSF from patients with NCHS (mainly IH) acted as a positive allosteric modulator of GABA-A receptors;[8] however, the bioactive CSF component is still unknown and we could not replicate these results.[9] Therefore, NT2 and IH diagnoses are challenging and are based on the clinical history, polysomnography (PSG), and multiple sleep latency test (MSLT) results.[1,10]

The standardized MSLT procedure was established 30 years ago, with several revisions concerning its interpretation and cutoffs for the diagnosis of narcolepsy.[11–14] No specific cutoff was established for IH, and similar mean sleep latency (MSL) thresholds are used for NT1, NT2, and IH. MSLT test–retest reliability has been evaluated in small samples of healthy subjects and in patients with narcolepsy, with high correlations found between MSL and number of sleep onset REM sleep periods (SOREMPs) in both populations, after a short between-test time interval.[15,16] Conversely, a low repeatability (10%–20%) of positive MSLT results was found in a large sample from the population-based Wisconsin cohort study, with a four-year-interval between tests.[17] A recent clinical cohort study on 29 patients with NT2 or IH and 7 patients with unspecified EDS (MSL above 8 minutes) confirmed the poor test–retest MSLT reliability after four years, with a change in diagnosis in more than 50% of patients.[18]

Considering the limited knowledge on MSLT reliability in clinical-based populations of patients with central disorders of hypersomnolence, we performed a retrospective study to assess the test–retest reliability of (1) the MSLT diagnostic classification, (2) the MSLT measures (MSL and number of SOREMPs), and (3) the variability determinants in patients with NT1, NT2, IH, and unspecified hypersomnolence complaint who underwent two MSLTs in drug-free conditions.

METHODS

Patients

We identified 83 patients with a primary complaint of hypersomnolence, defined by an Epworth Sleepiness Scale (ESS) score > 10/24 hours and/or reported total sleep time > 11/24 hours, who underwent two PSG recordings and MSLTs in drug-free conditions at the Reference National Centre for Narcolepsy of Montpellier, France, between 2003 and 2016. The hypersomnolence complaint was present for 3 months or more, with a main sleep period of at least 7 hours to exclude sleep deprivation, with no shift work schedule, no cataplexy, and no major depression. We excluded eight patients because of body mass index (BMI) > 35 kg/m^2 (n = 2), apnea–hypopnea index (AHI) > 15 per hour (n = 4), and periodic limb movement during sleep (PLMS) index > 15 per hour (n = 2) at the first assessment. All subjects had a standardized clinical interview and neurological examination to rule out the use or withdrawal of medication, substance, and neurological, medical, or psychiatric disorders that may be associated with excessive sleepiness (i.e., mood disorder, myotonic dystrophy type 1, Parkinson's disease, posttraumatic, and metabolic conditions). The final sample included 75 patients (32 males, median age at the first MSLT: 25.7 years, range: 14.2–64.4) with EDS symptoms that could not be explained by another cause of sleepiness. We measured CSF Hcrt-1 levels in 31 patients. All showed levels > 200 pg/mL, but for four patients with levels between 110 and 200 pg/mL. HLA DQB1* genotyping in 56 patients indicated that 28.6% carried the HLA DQB1*06:02 allele.

We also included 22 adults with NT1 (10 males, median age at the first MSLT: 31.2 years, range: 21.2–80.5) with two PSG tests and MSLTs in drug-free conditions performed at the same reference center. NT1 diagnosis was established on the basis of the ICSD-3 criteria: EDS that persists for 3 months or more, with history of clear-cut cataplexy and MSL ≤ 8 minutes and ≥2 SOREMPs, and/or CSF Hcrt-1 deficiency (≤110 pg/mL).[1] CSF Hcrt-1 levels were measured in 13 patients (all <110 pg/mL, including six with undetectable levels). All patients with NT1 were HLA DQB1*06:02-positive.

Evaluation

All patients underwent twice a full-night PSG recording followed the next day by the MSLT according to the standard guidelines, with a median between-test interval of 1.9 years (range: 0.02–11.9), with 1.6 years (range: 0.02–11.9) for the NCHS and 3.8 years (range: 0.1–7.7) for the NT1 group. All patients were drug-free for at least 2 weeks prior to each evaluation. Seventy-seven patients (65 NCHS and 12 NT1) were treatment-naïve at the first test and 29 (26 NCHS and 3 NT1) at both. The MSLT consisted of five scheduled naps separated by a 2-hour interval, between 9:00 am and 5:00 pm. Each time, patients had 20 minutes to fall asleep. If they did fall asleep, they were awakened after 15 minutes. All the MSLT were rescored by two authors (AD and RL). Sleep onset was determined by the first epoch of any stage of sleep according to the latest guidelines.[14] The MSL of the five naps was computed for each patient. A SOREMP was defined by the occurrence of REM sleep within 15 minutes of sleep onset during the MSLT and PSG recordings. PSG assessed total sleep time, sleep efficiency, REM sleep percentage and latency, microarousal index, AHI, and PLMS indexes.

Sex, age at EDS onset, and family history of EDS (first and second degree) were recorded at baseline. Age at MSLT, BMI, clinical assessment of depressive symptomatology, daytime sleepiness quantified with the ESS and presence of cataplexy, hypnagogic/hypnopompic hallucinations, and sleep paralysis were recorded at both sleep evaluations. Based on medical charts, we classified the decision to repeat the sleep study in four categories: (1) systematic reassessment (including re-evaluation for a research protocol, for a new treatment, post-pregnancy), (2) uncertain diagnosis at baseline, (3) spontaneous clinical improvement, or (4) clinical aggravation. We also reported the physician's clinical judgment on sleepiness severity variations (improvement, stability, or worsening) at the second evaluation.

According to the ICSD-3 criteria, the PSG (SOREMP) and MSLT results (MSL and SOREMP number) classified patients as follows: (1) MSL ≤ 8 minutes and ≥2 SOREMPs: narcolepsy phenotype, (2) MSL ≤ 8 minutes and <2 SOREMPs: hypersomnia phenotype, (3) MSL > 8 minutes and ≥2 SOREMPs: REM dysregulation, and (4) MSL > 8 minutes and <2 SOREMPs: normal phenotype.[1]

Statistical Analysis

Demographic characteristics and clinical data were described using percentages or frequencies for categorical variables and medians with ranges [minimal and maximal values] for continuous variables. Distributions were mostly skewed, according to the Shapiro–Wilk's test. To compare categorical variables between groups, the chi square or Fisher's exact test was used. The Mann–Whitney and Kruskall–Wallis tests were employed to compare continuous variables with two and three groups or more, respectively. When a significant relationship was found between more than two groups, two-by-two comparisons were performed to determine whether groups within the study were significantly different. A correction for multiple comparisons with the Bonferroni method was then used. A change of continuous variables between two MSLTs was defined as the difference between the second and first assessment. The Wilcoxon signed-rank test was run to compare differences of continuous variables between assessments. Statistical significance was set at $p < .05$. Statistical analyses were performed using SAS, version 9.4 (SAS Institute, Cary, NC, USA).

RESULTS

Patients' Baseline Characteristics, First PSG–MSLT Results, and Diagnostic Classification

At first assessment, among 75 patients with NCHS, 17 (22.7%) filled the PSG–MSLT diagnostic criteria for NT2, 20 (26.7%) for IH, 22 (29.3%) for REM dysregulation, and 16 (21.3%) had normal MSLT results despite their EDS complaint. We grouped the patients with REM dysregulation and normal results (total n = 38) in the "unspecified EDS" (UnsEDS) category. Twenty patients (9 with IH and 11 with UnsEDS) also performed a 24-hour continuous PSG. Seven patients with IH and nine with UnsEDS had a total sleep time higher than 11/24 hours (data not shown).

Among 22 patients with NT1, 16 (72.7%) were classified as having narcolepsy based on the PSG–MSLT criteria ("positive PSG–MSLT," hereafter). Among six others, five had ≥2 SOREMPs but MSL > 8 minutes, and the other one had MSL > 8 minutes and only one PSG SOREMP; however, all six had clear-cut cataplexy and five had a lumbar puncture that confirmed Hcrt-1 deficiency (<110 pg/mL).

The demographic, clinical, and neurophysiological characteristics according to the diagnosis at first assessment (NT1, NT2, IH, and UnsEDS) are shown in Table 1. Age was significantly different among groups; however, post hoc comparisons indicated that age was significantly lower only in the UnsEDS group compared with the NT1 group. As expected, patients with NT1 had lower CSF Hcrt-1 levels and were more frequently HLA DQB1*06:02-positive than the other groups. Patients with NT1 also had higher ESS scores and microarousal index value and reported more frequently sleep paralysis than patients with UnsEDS. Similarly, patients with NT1 reported more frequently hypnagogic hallucinations and had a shorter REM sleep latency than the other groups, and less sleep efficiency than the IH and UnsEDS groups.

Test–Retest Reliability of the PSG–MSLT Diagnostic Classification

At the second MSLT evaluation, among patients with NCHS, 14 (18.7%) filled the PSG–MSLT criteria for NT2, 19 (25.3%) for IH, 9 (12.0%) for REM dysregulation, and 33 (44.0%) had normal MSLT results (Table 2). None of them developed cataplexy during the between-test time interval. Only 8 of 17 patients (47.1%) who were initially classified as having NT2 remained in this diagnostic group at the second evaluation, whereas eight were classified as normal and one as having REM dysregulation. Only 5 of 20 patients (25.0%) with an initial diagnosis of IH remained in this diagnostic category, whereas five were reclassified as having NT2, one REM dysregulation and nine a normal phenotype. Seven of 22 patients (31.8%) with REM dysregulation remained in this group at retesting, whereas five were classified as having NT2, three IH, and seven a normal phenotype. Finally, among 16 patients (56.2%) with normal phenotype, nine remained in this group at retesting, whereas six were classified as having IH and one as having NT2. Among 16 patients with NT1 and baseline positive MSLT, 13 (81.3%) remained in the narcolepsy group at the second evaluation, whereas two were classified as having REM dysregulation and one hypersomnia phenotype (Table 3).

The repeatability of the PSG–MSLT diagnostic classification in the NT1 group (81.3%) was significantly higher than in the NT2 (47.1%, $p = .04$), IH (25.0%, $p = .008$), both NT2 and IH (35.1%, $p = .002$), and UnsEDS groups (42.1%, $p = .008$), as well as when compared with all patients with NCHS (38.7%, $p = .002$).

The inclusion or not of the nocturnal SOREM in the total number of SOREMPs (e.g., ICSD-2 criteria) did not significantly change the baseline and re-testing classifications. Similarly, taking into consideration only the daytime SOREMPs did not change the baseline and retesting diagnostic classifications, because the frequency of nocturnal SOREMPs was low in patients with NCHS (five patients at baseline and seven at the second evaluation). By applying these criteria, only one patient with NT2 was classified as having IH at both evaluations, whereas it did not affect the number of patients with NT1 with positive MSLT.

Test–Retest Reliability of the 8-minute MSL Cutoff

Among 37 patients with NCHS and a pathological MSL (≤8 minutes) at the first assessment, a MSL ≤ 8 minutes was confirmed in 18 (48.7%) at the second testing. Compared with the first assessment, patients with normalized MSL at the second MSLT reported increased PSG total sleep time at the second assessment (median change 51.0, range [−107.0; 191.0], $p = .01$). They also had a lower percentage of slow wave sleep (median change −4.2 range: [−14.4; 11.8], $p = .04$) and fewer SOREMPs (median change −1, range [−4; 2], $p = .006$) at the second test. Compared with patients with normal MSL at the second assessment, patients with pathological MSL at both tests showed a significantly higher baseline median ESS score (18.5, range: [11.0–23.0] vs 15.0, range: [9.0–22.0], $p = .04$), lower median change in total sleep time (−51.0 range: [−191.0; 107.0] vs 4.5 [−89.0; 157.0], $p = .02$), and higher median change in REM sleep latencies (0.0, range: [−90.0; 57.0] vs 30.0, range: [−163.0; 129.0], $p = .01$). The test–retest interval was longer in patients with normalized MSL at the second MSLT than in patients with pathological MSL at both tests (5.02 years, range: [0.21–11.88] vs 0.99, range: [0.02–6.72], $p = .003$). The rate of MSL normalization was 9.1% within the first year, 33.3% between one and two years, 60.0% between two and four years, and 80% after four years. No significant between-group difference was found concerning the physician's clinical impression of sleepiness severity variations between evaluations and the clinical decision to repeat the MSLT.

Among 38 patients with normal MSL at baseline, 15 had abnormal MSL at the second test and 23 patients remained within the normal range. Compared with the first assessment, patients with normal MSL at both evaluations had reduced percentage of REM sleep (median change = −3.7 range: [−20.7; 6.8], $p = .03$) and fewer SOREMPs (median change = 0 range: [−4.0; 1.0], $p = .02$) at the second assessment. Conversely, no significant PSG change was found between assessments in patients with abnormal MSL at the second test. Concerning the between-group comparisons, higher changes in REM sleep percentages ($p = .005$) and REM sleep latency ($p = .04$) were found in patients with normal MSL results at both tests compared with those with abnormal results at retesting.

MSL Value Variability Between Tests

Overall, MSL values were significantly different among diagnostic groups at both assessments ($p < .0001$ at the first assessment and $p = .004$ at the second assessment). The median MSL value was lower in patients with NT1 than in those with NCHS both at the first (NT1: 5.4-minute range: [0.5; 13.4] vs NCHS: 8.2-minute range: [0.4; 18.2], $p = .01$) and second assessment (NT1: 5.8-minute range: [0.4; 13.6] vs NCHS: 8.6 range: [0.6; 19.8], $p = .002$). Post hoc comparisons revealed higher MSL values in the UnsEDS group than in the three other groups at the first assessment and in the NT1 group at the second assessment.

Table 1—Demographic, Clinical, and Neurophysiological Characteristics of Patients With Narcolepsy Type 1, Narcolepsy Type 2, Idiopathic Hypersomnia, and Unspecified Hypersomnolence at Baseline.

Variable	NT1 *n* = 22		IH *n* = 20		NT2 *n* = 17		UnsEDS *n* = 38		Global *p*	Post hoc comparisons
	n	%	*n*	%	*n*	%	*n*	%		
Demographic characteristics										
Sex, male	10	45.4	7	35.0	8	47.1	17	44.7	.87	
Age (years)[a]	31.2 [21.2–80.5]		34.4 [17.2–54.5]		28.6 [17.3–54.0]		22.9 [14.2–64.4]		.008	UnsEDS < NT1
Clinical and biological characteristics										
BMI (kg/m^2)[a]	26.7 [18.4–39.6]		22.7 [17.9–31.7]		24.5 [20.7–34.6]		22.4 [16.7–30.4]		.006	UnsEDS < NT1
Age at EDS onset (years)[a]	20.5 [10.0–46.0]		19.0 [14.0–46.0]		19.0 [10.0–47.0]		16.5 [5.0–50.0]		.09	
Family history of EDS, yes	7	31.8	5	25.0	5	29.4	8	22.2	.86	
ESS[a]	19.0 [12.0–24.0]		15.5 [9.0–23.0]		18.0 [9.0–22.0]		14.5 [5.0–23.0]		<.001	UnsEDS, IH < NT1
Sleep paralysis, yes	13	59.1	4	20.0	4	23.5	4	10.5	.0005	UnsEDS < NT1
Hypnagogic hallucinations, yes	19	86.4	9	45.0	7	41.2	14	36.8	.002	IH, NT2, UnsEDS < NT1
Depressive symptomatology, yes	5	23.8	6	31.6	3	17.7	5	13.2	NA	
CSF hypocretin-1 level, *n*; (ng/L)[a]	13; 25.0 [3.0–108.0]		9; 335.5 [238.5–784]		9; 338.0 [179.0–436.0]		13; 309.63 [119.0–539.0]		<.001	NT1 < IH, NT2, UnsEDS
HLA DQB1*0602, present, *n*	22/22	100.0	6/19	31.6	5/15	33.3	5/22	22.7	<.0001	IH, NT2, UnsEDS < NT1
Neurophysiological characteristics										
PSG sleep latency (min)[a]	5.0 [0.0–32.0]		6.0 [0.0–24.0]		9.0 [0.0–31.0]		9.5 [0.0–106.0]		.06	
Total sleep time (min)[a]	437.5 [215.0–527.0]		423.0 [316.0–524.0]		422.0 [301.0–478.0]		452.5 [299.0–535.5]		.20	
Sleep efficiency (%)[a]	84.1 [69.4–94.9]		91.6 [68.1–97.2]		91.0 [62.8–98.9]		90.6 [66.4–98.1]		.02	NT1 < IH, UnsEDS
REM (%)[a]	21.7 [11.8–36.3]		22.8 [8.8–28.1]		20.8 [13.3–34.2]		22.1 [9.6–33.8]		.99	
REM latency (min)[a]	3.5 [0.0–152.0]		83.5 [40.0–241.0]		66.0 [5.0–191.0]		69.5 [0.0–164.0]		<.001	NT1 < IH, NT2, UnsEDS
PSG SOREMP	15	68.2	0	0.0	1	5.9	4	10.5	NA	
Microarousal index[a]	13.3 [4.0–40.1]		10.1 [6.9–17.7]		8.3 [2.2–21.3]		8.5 [2.4–22.1]		.004	UnsEDS < NT1
PLMS index (per h)[a]	0.9 [0.0–31.5]		0.6 [0.0–13.0]		0.5 [0.0–11.0]		0.6 [0.0–13.0]		.96	
AHI (per h)[a]	5.1 [0.0–32.1]		1.9 [0.0–13.5]		2.7 [0.1–13.2]		1.3 [0.0–10.5]		.12	
PSG and MSLT SOREMP[a]	4.0 [1.0–6.0]		0.0 [0.0–1.0]		3.0 [2.0–5.0]		2.0 [0.0–5.0]		<.001	UnsEDS < NT1; IH < NT2, UnsEDS
MSLT mean sleep latency (min)[a]	5.4 [0.5–13.4]		6.0 [1.8–8.0]		4.2 [0.4–7.8]		11.4 [8.2–18.2]		<.001	NT1, NT2, IH < UnsEDS

[a]Continuous variables were expressed as medians [minimum value–maximum value].
AHI = apnea-hypopnea index; BMI = body mass index; CSF = cerebrospinal fluid; EDS = excessive daytime sleepiness; ESS = Epworth Sleepiness Scale; IH = idiopathic hypersomnia; MSLT = multiple sleep latency test; NA = not applicable; NT1 = narcolepsy type 1; NT2 = narcolepsy type 2; PLMS = periodic limb movement during sleep; PSG = polysomnography; SOREMP = sleep onset REM sleep period; UnsEDS = unspecified hypersomnolence.

Table 2—Concordance Between the Diagnostic Classification of the First and the Second MSLT in Patients With Noncataplectic Central Disorders of Hypersomnolence.

Noncataplectic central disorders of hypersomnolence		MSLT #2				Total
		Hypersomnia phenotype	Narcolepsy phenotype	REM dysregulation phenotype	Normal phenotype	
MSLT #1	Hypersomnia phenotype	5 (25.0%)[a]	5 (25.0%)	1 (5.0%)	9 (45.0%)	20
	Narcolepsy phenotype	0 (0.0%)	8 (47.1%)	1 (5.9%)	8 (47.1%)	17
	REM dysregulation phenotype	3 (13.6%)	5 (22.7%)	7 (31.8%)	7 (31.8%)	22
	Normal phenotype	6 (37.5%)	1 (6.2%)	0 (0.0%)	9 (56.2%)	16
Total		14	19	9	33	75

[a]Row percentages.
MSLT = multiple sleep latency test.

Table 3—Concordance Between the Diagnostic Classification of the First and the Second MSLT in Patients With Narcolepsy Type 1.

Narcolepsy type 1		MSLT #2				Total
		Hypersomnia phenotype	Narcolepsy phenotype	REM dysregulation phenotype	Normal phenotype	
MSLT #1	Hypersomnia phenotype	0 (0.0%)[a]	0 (0.0%)	0 (0.0%)	0 (0.0%)	0
	Narcolepsy phenotype	1 (6.2%)	13 (81.3%)	2 (12.5%)	0 (0.0%)	16
	REM dysregulation phenotype	0 (0.0%)	1 (20.0%)	3 (60.0%)	1 (20.0%)	5
	Normal phenotype	0 (0.0%)	1 (100.0%)	0 (0.0%)	0 (0.0%)	1
Total		1	15	5	1	22

[a]Row percentages.
MSLT = multiple sleep latency test.

Between-test MSL changes were significant in the IH (median = +2.2 minutes, range: [−2.0; +9.0], p = .002), NT2 (median = +3.2 minutes, range: [−2.6; +11.8], p = .002), and UnsEDS groups (median = −1.6 minutes, range: [−12.2; +8.6], p = .04), but not in the NT1 group (median = −0.2 minutes, range: [−7.8; +5.4]). Intergroup MSL changes (p = .0005) were significant only between the UnsEDS and the NT2 and IH groups in post-host comparisons (Figure 1).

We then classified patients with NCHS in three subgroups according to the tertiles of MSL changes between MSLTs (Table 4): (1) worsening MSL (from −12.2 to −2.4 minutes; median = −6.4 minutes; n = 19 patients); (2) stable MSL (from −2.0 to +2.0 minutes; median = −0.2 minutes; n = 26); (3) improving MSL (from +2.2 to +11.8 minutes; median = 4.6 minutes; n = 30). Compared with the other subgroups, patients with improving MSL had a longer test–retest interval, higher weight gain, reduced sleep efficiency, and fewer SOREMPs at the second assessment. Patients with stable MSL reported more frequently family history of sleepiness. Only in patients with improving MSL, BMI (p = .03), sleep efficiency (p = .04), and SOREMP number (p = .0006) significantly changed between tests.

SOREMP Number Variability Between Tests

At the first assessment, 39 (52.0%) patients with NCHS had ≥2 SOREMPs and 28 (37.3%) at the second assessment. Compared with patients with ≥2 SOREMP only at the first assessment (n = 18), those with ≥2 SOREMPs at both evaluations (n = 21) were more often males (66.7% vs 33.3%; p = .04), with shorter REM sleep latency (−14.0 minutes vs +8.0 minutes; p = .04) and MSL (−0.8 minutes vs +3.9 minutes; p = .02) at the second evaluation. The test–retest interval was longer in the group with <2 SOREMPs than in patients with ≥2 SOREMPs at the second evaluation (2.8 years vs 0.7; p = .004). Among 36 patients with NCHS and <2 SOREMPs at baseline, only seven had ≥2 SOREMPs at retesting. The PSG REM sleep latency was reduced (−40.0 minutes vs +5.0 minutes; p = .003) and the REM sleep percentage increased (+6.7% vs −1.4%; p = .002) in patients with ≥2 SOREMP at the second evaluation.

Effect of the Clinical Decision to Repeat the PSG–MSLT Evaluation

Sleepiness severity improved in 14.9% (six NT2, two IH, and three UnsEDS), worsened in 8.1% (one NT2, three IH, and two

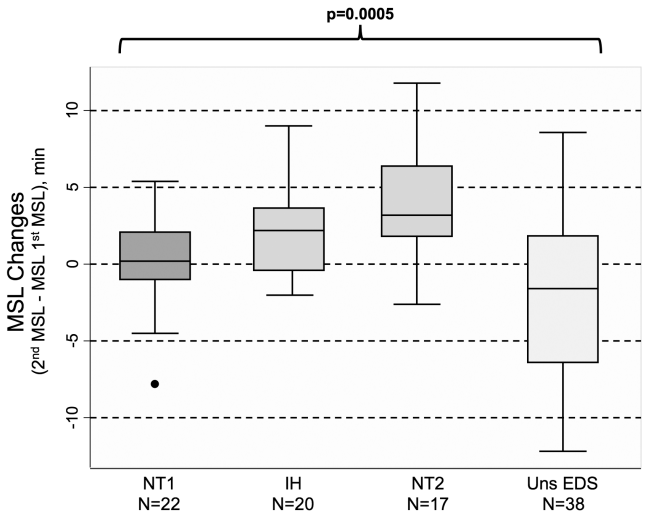

Figure 1—Changes in mean sleep latency (MSL) between the second and first MSLT in patients with narcolepsy type 1 (NT1), narcolepsy type 2 (NT2), idiopathic hypersomnia (IH), and unspecified hypersomnolence (unspecified excessive daytime sleepiness [UnsEDS]).

UnsEDS), and remained stable in 77.0% of patients with NCHS, according to the physician's clinical impression of change at the second evaluation. The second PSG–MSLT recording was decided for a systematic reassessment (n = 26), because of uncertain baseline diagnosis (n = 38) or due to clinically significant changes in the symptom severity (n = 11). Demographic, clinical, and neurophysiological characteristics were not significantly different in these three groups, but for a longer test–retest interval in patients with a systematic reassessment.

DISCUSSION

We examined the test–retest reliability of the MSLT (median between-test time interval of 1.9 years) in two well-characterized clinical populations of patients with NT1 and NCHS. Our findings highlight the lack of stability of the PSG–MSLT based-classification in NCHS, but not in NT1, with frequent diagnostic changes in the NT2 and IH groups. The MSL and SOREMP variability showed high interindividual heterogeneity that was associated with the time interval between tests, but not with the spontaneous course of the clinical symptoms.

As a positive control group, we included 22 patients with NT1 who had clear-cut cataplexy, carried the HLA DQB1*06:02 allele, and had low CSF Hcrt-1 levels (available for 13 patients). At the first test, 27.3% of patients with NT1 had a MSL > 8 minutes and/or <2 SOREMPs (i.e., negative MSLT). A lower rate (6.6%) of false negative MSLT results was recently reported in NT1.[19] This discrepancy could be explained by the clinical setting of our study, in which most patients had a second MSLT evaluation due to a specific reason, including nonconclusive baseline findings. On the other hand, in our study, the test–retest reliability for NT1 was high (83.1% of patients had a positive MSLT at both evaluations). Similarly, in a previous study, 28/30 narcoleptic patients (83% with cataplexy) had positive MSLT results at both evaluations within a median interval of 4.8 years.[16] NT1 is a life-long disease, without spontaneous remission, but with highly heterogeneous disease course, Hcrt-1 levels, and MSLT results.[20] The MSLT results are generally robust, except in elderly patients in whom MSL increases and the number of SOREMPs decreases progressively.[21] Although an age effect cannot be formally ruled out in our study, the test–retest MSLT variability in NT1 mainly reflects the test intrinsic performances without any effect of the spontaneous disease course.

In the NCHS population and differently from that observed in the NT1 group, we confirmed previous preliminary findings on the poor reliability of the MSLT based-diagnoses of NT2 and IH. During an interval of almost 2 years, the diagnosis changed in more than 50% of patients who were initially classified as having NT2. None of the patients with initial NCHS criteria developed cataplexy between the MSLT1 and MSLT2. Trotti et al. found similar rates of diagnostic instability (only 5 of 15 patients in the NT2 category had similar MSLT results at follow-up).[18] Another study reported that fewer than half of 18 patients with NT2 had a positive MSLT at both assessments.[19] The reliability of the MSLT-based narcolepsy criteria has also been studied in the general population (823 subjects had two PSG–MSLT at a 4-year interval).[17] Among subjects without shift-work and with habitual nighttime sleep duration >6 hours, the stability of a positive MSLT was very low (10%–20%). A large general population-based cohort showed a high rate of subjects with EDS and positive MSLT (4.1% of men and 0.4% of women) with a strong association with shift work, sleep deprivation, and antidepressant intake.[22] Another study confirmed the frequent number of SOREMPs in the general population, especially in those with short MSL, with significant association with shift work and sleep apnea.[23] Altogether, these results confirm the limited contribution of a single PSG–MSLT assessment for the diagnosis of NT2.[24] Indeed, NT2 is a diagnosis of elimination and only a minority of patients have a stable phenotype with daytime hypersomnolence and high REM sleep propensity, as measured by the MSLT. To establish a genuine diagnosis of NT2, we recommend to reassess the clinical features and to repeat the MSLT (after exclusion of shift work, sleep deprivation, and antidepressant intake) to identify patients with positive MSLT at both evaluations.

Similarly, we found extremely low test–retest stability for the MSLT-based diagnosis of IH. Only 25% of patients with initial criteria for IH were classified as having IH at the second test. This is lower than in a previous study where 8 of 14 patients with IH had stable MSLT results/diagnostic classification.[18] MSLT contribution to the diagnosis of IH seems, thus, to be limited in most patients, and other features of hypersomnolence should also be taken into account for the diagnosis of this NCHS subtype with long sleep time.

The low test–retest MSLT repeatability in NCHS could be explained by the intrinsic features of the test, but also by the spontaneous clinical course of such disorders. Until recently, IH and NT2 were considered as long-lasting diseases; however, several case series reported spontaneous improvement or remissions in 11%–25% of patients with IH.[25–28] Our results are in line with these observations because 14.9% of patients with

Table 4—Variables Associated With MSL Improvement, Worsening, or Stability at the Second MSLT in Patients With Noncataplectic Central Disorders of Hypersomnolence.

	Improving MSL >2 minutes *n* = 30		**Stable MSL [−2; +2] minutes *n* = 26**		**Worsening MSL <−2 minutes *n* = 19**		***p***
Variable	***n***	**%**	***n***	**%**	***n***	**%**	
Demographic characteristics							
Sex, male	21	70.0	15	57.7	7	36.8	.08
Age (years)[a]	28.1 [15.0–53.6]		26.0 [14.2–54.5]		24.7 [16.3–64.4]		.78
Clinical and biological characteristics at baseline							
BMI (kg/m^2)[a]	24.3 [17.5–34.6]		23.5 [17.9–31.7]		22.2 [16.7–34.6]		.84
Age at EDS onset (years)[a]	18.0 [5.0–47.0]		18.0 [10.0–46.0]		19.0 [11.0–50.0]		.88
Family history of EDS	5	17.24	11	42.31	2	11.11	.04
ESS[a]	14.5 [5.0–23.0]		15.0 [9.0–23.0]		17.0 [7.0–23.0]		.67
Sleep paralysis	7	23.3	3	11.5	2	10.5	.38
Hypnagogic hallucinations	13	43.3	11	42.3	6	31.6	.69
Depressive symptomatology	8	27.59	4	15.38	2	10.53	NA
CSF hypocretin-1 level (ng/L)[a]	338.0 [179–615]		312.5 [119–784]		353.1 [232–539]		.51
HLA DQB1*0602, present	8	34.8	6	27.3	2	18.2	.60
Neurophysiological characteristics at baseline							
PSG sleep latency (min)[a]	7.5 [0.0–83.0]		9.0 [0.0–106.0]		8.0 [1.0–59.0]		.73
Total sleep time (min)[a]	425.0 [299.0–518.0]		434.0 [313.0–535.0]		433.0 [342.0–509.0]		.19
Sleep efficiency (%)[a]	91.4 [68.1–97.6]		91.6 [62.8–98.9]		89.7 [66.4–98.1]		.53
REM (%)[a]	22.7 [13.1–34.2]		20.6 [8.8–32.7]		21.9 [9.6–31.9]		.23
REM latency (min)[a]	74.0 [5.0–193.0]		72.5 [12.0–241.0]		62.0 [0.0–164.0]		.37
PSG SOREMP	1	3.33	1	3.85	3	15.79	NA
Micro-arousal index[a]	8.9 [3.8–21.3]		8.7 [2.2–22.0]		8.7 [2.4–14.1]		.97
PLMS index (per h)[a]	0.0 [0.0–11.0]		1.5 [0.0–13.0]		0.0 [0.0–13.0]		.21
AHI (per h)[a]	1.9 [0.0–13.5]		1.4 [0.0–5.3]		2.3 [0.0–10.5]		.22
PSG and MSLT SOREMP[a]	1.5 [0.0–5.0]		1.0 [0.0–5.0]		2.0 [0.0–5.0]		.49
MSLT mean sleep latency (min)[a]	6.5 [1.0–14.8]		6.9 [0.4–16.5]		13.7 [6.6–18.2]		<.001
Between-test clinical and polysomnographic changes							
Time interval between tests (years)[a]	3.6 [0.2–11.9]		0.9 [0.0–8.5]		0.7 [0.1–6.8]		<.001
Between-test interval < 1 year, yes	6	20.0	14	53.9	11	57.9	.009
Reasons of reassessment							
Systematic reassessment	13	43.3	8	30.8	5	27.8	NA
Uncertain diagnosis at baseline	13	43.3	15	57.7	9	50.0	
Change in severity	4	13.3	3	11.5	4	22.2	
Change in BMI (kg/m^2)[a]	0.4 [−2.9–9.8]		−0.3 [−4.7–3.5]		0.0 [−6.4–2.6]		.02
Change in ESS[a]	–2.0 [−14.0–9]		0.0 [−12.0–11.0]		−1.0 [−7.0–4.0]		.81
CGI-c							
Improvement	7	23.3	3	11.5	1	5.6	NA
Stability	20	66.7	21	80.8	16	88.9	
Worsening	3	10.0	2	7.7	1	5.6	

Table 4—*Continued*

	Improving MSL >2 minutes *n* = 30		Stable MSL [−2; +2] minutes *n* = 26		Worsening MSL <−2 minutes *n* = 19		*p*
Variable	***n***	**%**	***n***	**%**	***n***	**%**	
Change in PSG sleep latency (min)[a]	1.0 [−29.0–28.0]		−2.0 [−88.0–192.0]		−2.0 [−55.0–15.0]		.14
Change in total sleep time (min)[a]	13.5 [−107–191]		−4.0 [−157.0–89.0]		0.0 [−100.0–106.0]		.46
Change in sleep efficiency (%)[a]	−4.7 [−23.3−18.7]		0.3 [−19.8–24.3]		2.0 [−16.4–22.4]		.05
Change in REM (%)[a]	−2.3 [−20.7–20.5]		1.8 [−6.5–13.4]		0.9 [−7.6–8.6]		.06
Change in REM latency (min)[a]	3.0 [−95.0–90.0]		10.0 [−129.0−163.0]		1.0 [−141.0–132.0]		.34
Change in microarousal index[a]	1.1 [−17.8–4.6]		−2.1 [−10.6–11.3]		−0.4 [−8.9–9.6]		.42
Change in PLMS index (per h)[a]	0.0 [−5.5–11.4]		0.0 [−7.1–7.4]		0.0 [−9.1–47.5]		.72
Change in AHI (per h)[a]	0.0 [−8.1–36.4]		0.2 [−3.4–38.5]		0.2 [−3.8–9.0]		.75
Change in PSG and MSLT SOREMP[a]	−0.5 [−4.0–2.0]		0.0 [−4.0–3.0]		0.0 [−4.0–4.0]		.01
Change in MSLT, mean sleep latency (min)[a]	4.6 [2.2–11.8]		−0.2 [−2.0–2.0]		−6.4 [−2.2–12.4]		<.001

[a]Continuous variables were expressed as medians [minimum value–maximum value].
AHI = apnea–hypopnea index; BMI = body mass index; CGI-c = Clinical Global Impression of Change; CSF = cerebrospinal fluid; EDS = excessive daytime sleepiness; ESS = Epworth Sleepiness Scale; IH = idiopathic hypersomnia; MSL = multiple sleep latency; MSLT = multiple sleep latency test; NA = not applicable; PLMS = periodic limb movement during sleep; PSG = polysomnography; SOREMP = sleep onset REM sleep period.

NCHS had a spontaneous clinical improvement after a median of 1.6 years. A higher rate of remission was recently reported in patients with NCHS after a 5-year follow-up (32.5% in IH and 44.6% in NT2).[29] Our study also shows that the sleepiness improvement, based on the ESS score or the physician's clinical impression, was not associated with MSLT normalization in patients with pathological baseline MSL. This discrepancy between subjective and objective daytime sleepiness changes in NT2 and IH has been already described in hypersomnia associated with major depression.[30,31] We found here a strong effect of the test–retest interval (80% of normalization of the MSL after 4 years), suggesting that objective daytime sleepiness tends to decrease over time in both NT2 and IH. Interestingly, the duration of the between-test interval was not associated with a pathological MSL at the second assessment in patients with normal MSL at first-assessment.

Two factors contributed to the variability of MSLT-based classifications: sleepiness (i.e., MSL) and the REM sleep propensity (i.e., daytime and nocturnal SOREMPs). We found significant between-test changes in the MSL values, with a large heterogeneity among patients, and also in the total number of SOREMPs. Indeed, the SOREMP number remained stable only in 50 patients at both evaluations (21/75: ≥2 SOREMPs; 29/75: <2 SOREMPs). Moreover, the changes in SOREMP number were associated with variations in the nocturnal REM sleep parameters (i.e., REM sleep latency and REM sleep percentage). The number of SOREMP was also associated with MSL and with sex (male), as previously described in the general population.[22,23]

Several limitations should be taken into consideration when interpreting our results. First, short sleep duration and sleep deprivation have major effect on the MSLT results. Accordingly, we included patients with a reported main sleep period of at least 7 hours, but this was not objectively assessed by actigraphy.[4] Second, overall, only 29 patients were treatment-naïve at both evaluations, whereas the others stopped their stimulant (for patients with NT1 and NHCS) or antidepressant medication (only for patients with NT1) at least 2 weeks before each test. An acute withdrawal effect cannot be totally excluded, although 2 weeks correspond to more than eight half-lives of psychostimulant medications. Third, our clinical population sample was well characterized, but relatively small. Moreover, all patients were referred to a single reference national sleep clinic. We did not use a controlled prospective design to systematically reassess patients; the interval between the test–retest assessments was variable and the decision to repeat the PSG–MSLT not random. The retrospective design could have introduced a bias that might have increased the high test–retest variability of PSG–MSLT results and the unexpected high rate of negative MSLT in the NT1 group. Fourth, HLA typing and CSF Hcrt-1 levels were not available for all participants, with none of the patients being systematically reassessed for Hcrt-1 levels. We did not include children below 16 years of age with central disorders of hypersomnolence to avoid additional bias regarding the age effect on the MSLT results.[21] Finally, due to the retrospective nature of our observational study with a heterogeneous design, our findings need to be confirmed in another prospective study.

In conclusion, the PSG–MSLT measures and diagnostic classification lack stability in central disorders of hypersomnolence, with frequent diagnostic changes for NT2 and IH, in contrast to NT1. Our results challenge the validity of the established MSL cutoff for the diagnosis of NT2 and IH. On the basis of these findings, we highlight the need to reassess the clinical features and to repeat the MSLT at regular intervals to confirm a stable hypersomnia and to provide a sufficiently accurate diagnosis of NT2 and IH. Repeated testing may allow the identification of

homogeneous populations with NT2 or IH and, hopefully, the discovery of specific biomarkers for each condition.

REFERENCES

1. American Academy of Sleep Medicine. International classification of sleep disorders, 3rd edition (ICSD-3). Darien, IL: AASM Resource Library; 2014.
2. Dauvilliers Y, Arnulf I, Mignot E. Narcolepsy with cataplexy. Lancet. 2007; 369(9560): 499–511.
3. Dauvilliers Y, Siegel JM, Lopez R, Torontali ZA, Peever JH. Cataplexy–clinical aspects, pathophysiology and management strategy. Nat Rev Neurol. 2014; 10(7): 386–395.
4. Baumann CR, Mignot E, Lammers GJ, et al. Challenges in diagnosing narcolepsy without cataplexy: a consensus statement. Sleep. 2014; 37(6): 1035–1042.
5. Billiard M, Sonka K. Idiopathic hypersomnia. Sleep Med Rev. 2016; 29: 23–33.
6. Dauvilliers Y, Delallée N, Jaussent I, et al. Normal cerebrospinal fluid histamine and tele-methylhistamine levels in hypersomnia conditions. Sleep. 2012; 35(10): 1359–1366.
7. Faull KF, Thiemann S, King RJ, Guilleminault C. Monoamine interactions in narcolepsy and hypersomnia: reanalysis. Sleep. 1989; 12(2): 185–186.
8. Rye DB, Bliwise DL, Parker K, et al. Modulation of vigilance in the primary hypersomnias by endogenous enhancement of GABAA receptors. Sci Transl Med. 2012; 4(161): 161ra151.
9. Dauvilliers Y, Evangelista E, Lopez R, et al. Absence of γ-Aminobutyric Acid-A receptor potentiation in central hypersomnolence disorders. Ann Neurol. 2016; 80(2): 259–268.
10. American Psychiatric Association. The Diagnostic and Statistical Manual of Mental Disorders, 5th ed (DSM-5). APA. Arlington, VA: 2013.
11. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep. 1986; 9(4): 519–524.
12. Thorpy MJ. The clinical use of the multiple sleep latency test. Sleep 1992; 15(3): 268–276.
13. Arand D, Bonnet M, Hurwitz T, Mitler M, Rosa R, Sangal RB. The clinical use of the MSLT and MWT. Sleep. 2005; 28(1): 123–144.
14. Littner MR, Kushida C, Wise M, et al.; Standards of Practice Committee of the American Academy of Sleep Medicine. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. Sleep. 2005; 28(1): 113–121.
15. Zwyghuizen-Doorenbos A, Roehrs T, Schaefer M, Roth T. Test-retest reliability of the MSLT. Sleep. 1988; 11(6): 562–565.
16. Folkerts M, Rosenthal L, Roehrs T, et al. The reliability of the diagnostic features in patients with narcolepsy. Biol Psychiatry. 1996; 40(3): 208–214.
17. Goldbart A, Peppard P, Finn L, et al. Narcolepsy and predictors of positive MSLTs in the Wisconsin Sleep Cohort. Sleep. 2014; 37(6): 1043–1051.
18. Trotti LM, Staab BA, Rye DB. Test-retest reliability of the multiple sleep latency test in narcolepsy without cataplexy and idiopathic hypersomnia. J Clin Sleep Med. 2013; 9(8): 789–795.
19. Andlauer O, Moore H, Jouhier L, et al. Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency. JAMA Neurol. 2013; 70(7): 891–902.
20. Lopez R, Barateau L, Evangelista E, et al. Temporal changes in the cerebrospinal fluid level of hypocretin-1 and histamine in narcolepsy. Sleep. 2017; 40(1).
21. Dauvilliers Y, Gosselin A, Paquet J, Touchon J, Billiard M, Montplaisir J. Effect of age on MSLT results in patients with narcolepsy-cataplexy. Neurology. 2004; 62(1): 46–50.
22. Mignot E, Lin L, Finn L, et al. Correlates of sleep-onset REM periods during the multiple sleep latency test in community adults. Brain. 2006; 129(Pt 6): 1609–1623.
23. Singh M, Drake CL, Roth T. The prevalence of multiple sleep-onset REM periods in a population-based sample. Sleep. 2006; 29(7): 890–895.
24. Mayer G, Lammers GJ. The MSLT: More objections than benefits as a diagnostic gold standard? Sleep. 2014; 37(6): 1027–1028.
25. Billiard M, Merle C, Carlander B, Ondze B, Alvarez D, Besset A. Idiopathic hypersomnia. Psychiatry Clin Neurosci. 1998; 52(2): 125–129.
26. Anderson KN, Pilsworth S, Sharples LD, Smith IE, Shneerson JM. Idiopathic hypersomnia: a study of 77 cases. Sleep. 2007; 30(10): 1274–1281.
27. Bassetti C, Aldrich MS. Idiopathic hypersomnia. A series of 42 patients. Brain. 1997; 120 (Pt 8): 1423–1435.
28. Bruck D, Parkes JD. A comparison of idiopathic hypersomnia and narcolepsy-cataplexy using self report measures and sleep diary data. J Neurol Neurosurg Psychiatry. 1996; 60(5): 576–578.
29. Kim T, Lee JH, Lee CS, Yoon IY. Different fates of excessive daytime sleepiness: survival analysis for remission. Acta Neurol Scand. 2016; 134(1): 35–41.
30. Plante DT, Finn LA, Hagen EW, Mignot E, Peppard PE. Longitudinal associations of hypersomnolence and depression in the Wisconsin Sleep Cohort Study. J Affect Disord. 2017; 207: 197–202.
31. Dauvilliers Y, Lopez R, Ohayon M, Bayard S. Hypersomnia and depressive symptoms: methodological and clinical aspects. BMC Med. 2013; 11(1): 78.

FUNDING

YD has received funds for speaking, board engagements, and travel to conferences by UCB pharma, Jazz, and Bioprojet. RL has received funds for speaking, and travel to conferences by UCB pharma, Shire, and HAC Pharma.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication January, 2017
Submitted in final revised form June, 2017
Accepted for publication October, 2017

Address correspondence to: Yves Dauvilliers, MD, PhD and Régis Lopez, MD, PhD, Département de Neurologie, Centre national de référence narcolepsie hypersomnies, Inserm U1061, CHU Gui-de-Chauliac, 80, avenue Augustin Fliche, 34295 Montpellier Cedex 5, France. Telephone: (33) 4 67 33 74 78; Fax: (33) 4 67 33 72 85; Email: y-dauvilliers@chu-montpellier.fr, r-lopez@chu-montpellier.fr

DISCLOSURE STATEMENT

None declared.