Nocturnal Rapid Eye Movement Sleep Latency for Identifying Patients With Narcolepsy/Hypocretin Deficiency

Olivier Andlauer, MD; Hyatt Moore, IV, MSc; Laura Jouhier, MD; Christopher Drake, PhD; Paul E. Peppard, PhD; Fang Han, MD; Seung-Chul Hong, MD, PhD; Francesca Poli, MD; Giuseppe Plazzi, MD; Ruth O'Hara, PhD; Emmanuel Haffen, MD, PhD; Thomas Roth, PhD; Terry Young, PhD; Emmanuel Mignot, MD, PhD

**IMPORTANCE** Narcolepsy, a disorder associated with HLA-DQB1*06:02 and caused by hypocretin (orexin) deficiency, is diagnosed using the Multiple Sleep Latency Test (MSLT) following nocturnal polysomnography (NPSG). In many patients, a short rapid eye movement sleep latency (REML) during the NPSG is also observed but not used diagnostically.

**OBJECTIVE** To determine diagnostic accuracy and clinical utility of nocturnal REML measures in narcolepsy/hypocretin deficiency.

**DESIGN, SETTING, AND PARTICIPANTS** Observational study using receiver operating characteristic curves for NPSG REML and MSLT findings (sleep studies performed between May 1976 and September 2011 at university medical centers in the United States, China, Korea, and Europe) to determine optimal diagnostic cutoffs for narcolepsy/hypocretin deficiency compared with different samples: controls, patients with other sleep disorders, patients with other hypersomnias, and patients with narcolepsy with normal hypocretin levels. Increasingly stringent comparisons were made. In a first comparison, 516 age- and sex-matched patients with narcolepsy/hypocretin deficiency were selected from 1749 patients and compared with 516 controls. In a second comparison, 749 successive patients undergoing sleep evaluation for any sleep disorders (low pretest probability for narcolepsy) were compared within groups by final diagnosis of narcolepsy/hypocretin deficiency. In the third comparison, 254 patients with a high pretest probability of having narcolepsy were compared within group by their final diagnosis. Finally, 118 patients with narcolepsy/hypocretin deficiency were compared with 118 age- and sex-matched patients with a diagnosis of narcolepsy but with normal hypocretin levels.

**MAIN OUTCOME AND MEASURES** Sensitivity and specificity of NPSG REML and MSLT as diagnostic tests for narcolepsy/hypocretin deficiency. This diagnosis was defined as narcolepsy associated with cataplexy plus HLA-DQB1*06:02 positivity (no cerebrospinal fluid hypocretin-1 results available) or narcolepsy with documented low (<110 pg/mL) cerebrospinal fluid hypocretin-1 level.

**RESULTS** Short REML (<15 minutes) during NPSG was highly specific (99.2% [95% CI, 98.5%-100.0%] of 516 and 99.6% [95% CI, 99.1%-100.0%] of 735) but not sensitive (50.6% [95% CI, 46.3%-54.9%] of 516 and 35.7% [95% CI, 10.6%-60.8%] of 14) for patients with narcolepsy/hypocretin deficiency vs population-based controls or all patients with sleep disorders undergoing a nocturnal sleep study (area under the curve, 0.799 [95% CI, 0.771-0.826] and 0.704 [95% CI, 0.524-0.907], respectively). In patients with central hypersomnia and thus a high pretest probability for narcolepsy, short REML remained highly specific (95.4% [95% CI, 90.4%-98.3%] of 132) and similarly sensitive (57.4% [95% CI, 48.1%-66.3%] of 122) for narcolepsy/hypocretin deficiency (area under the curve, 0.765 [95% CI, 0.707-0.831]). Positive predictive value in this high pretest probability sample was 92.1% (95% CI, 83.6%-97.0%).

**CONCLUSIONS AND RELEVANCE** Among patients being evaluated for possible narcolepsy, short REML (<15 minutes) at NPSG had high specificity and positive predictive value and may be considered diagnostic without the use of an MSLT; absence of short REML, however, requires a subsequent MSLT.

Published online May 6, 2013.
The association of excessive daytime sleepiness with cataplexy (episodes of muscle weakness triggered by emotions) has traditionally been used to define narcolepsy. In 1960, it was found that during nocturnal polysomnography (NPSG), 50% of cases have rapid transitions from wake to rapid eye movement (REM) sleep (sleep-onset REM periods [SOREMPs]). The Multiple Sleep Latency Test (MSLT), a test that allows 4 to 5 daytime nap opportunities for sleep onset and SOREMPs, was subsequently developed to increase sensitivity. A mean sleep latency (MSL) of 8 minutes or less and 2 or more SOREMPs is considered diagnostic for narcolepsy. Positive MSLT results have been reported in 2% to 4% of the population.

The opportunity for a new gold standard, low concentration of hypocretin-1 (≤110 pg/mL; hypocretin deficiency) in the cerebrospinal fluid (CSF), has recently arisen with the finding of an autoimmune basis of narcolepsy-cataplexy. Almost all cases with low CSF hypocretin-1 levels are HLA-DQB1*06:02 positive (vs 25% in the US general population), making it one of the most tightly HLA antigen–associated disorders known. Positive MSLT results have been reported in 2% to 4% of the population.

Because sleep apnea and sleep deprivation can cause SOREMPs and short MSL during naps, NPSG is systematically performed prior to MSLT. Considering the MSLT takes a whole day in a sleep laboratory, we aimed to determine if short REM sleep latency (REML) during NPSG could identify narcolepsy cases associated with hypocretin deficiency, saving time and money. In this study, the value of a short REML during NPSG was compared with MSLT results and HLA-DQB1*06:02 typing findings was also explored.

Methods

Participants

Descriptions of the different samples are summarized in Table 1. All participants gave written informed consent. The Stanford institutional review board approved the study for patients included in the Stanford and Chinese databases and the controls collected from the Wisconsin Sleep Cohort; the Detroit, Michigan, tricounty area; and China.

Patients With Narcolepsy

Eight hundred eight patients with narcolepsy were retrospectively identified using the Stanford Center for Narcolepsy Research database (85.0%; n = 687) and a similar database available in China (15.0%; n = 121). In accordance with ICSD-2, narcolepsy diagnosis was based on excessive daytime sleepiness associated with either clear-cut cataplexy or a positive MSLT result (defined as MSL ≤8 minutes associated with ≥2 SOREMPs during 4-5 naps) or a low CSF hypocretin-1 level (≤110 pg/mL). All cases had to have NPSG data available, and patients with a diagnosis of secondary narcolepsy (hypothalamic lesion [n = 1], multiple sclerosis [n = 2], and Turner syndrome [n = 1]) were excluded.

Patients had been recruited between May 1976 and January 2011 at 4 sites: United States (67.1%; n = 542), China (15.0%; n = 121), Korea (10.5%; n = 85), and Italy (7.4%; n = 60). Self-identified ethnicity was 63.5% white (n = 513), 26.9% Asian (n = 217), 7.7% African American (n = 62), and 2.0% other (n = 16). Three samples of patients with narcolepsy were defined as follows:

Sample A: 169 patients with a diagnosis of narcolepsy and documented low CSF hypocretin-1 level (≤110 pg/mL) with or without cataplexy or HLA-DQB1*06:02 positivity.

Sample B: 508 patients with a diagnosis of narcolepsy-cataplexy and HLA-DQB1*06:02 positivity but without CSF hypocretin-1 measurement available. Prior studies have estimated that 98% of such cases have a low CSF hypocretin-1 level (≤110 pg/mL).

Sample C: 131 patients with a diagnosis of narcolepsy and normal CSF hypocretin-1 level (>200 pg/mL). Criteria for a normal CSF hypocretin-1 designation were either CSF hypocretin-1 values of more than 200 pg/mL (n = 60; 45.8% of the sample) or HLA-DQB1*06:02 negativity (n = 71; 54.2% of the sample). Prior studies have estimated that 99.8% of subjects with HLA-DQB1*06:02 negativity have a normal CSF hypocretin-1 level.

Controls

One thousand twenty-six random population controls (sample D) were retrospectively identified from 3 sources: a sample drawn from the Wisconsin Sleep Cohort (36.2%; n = 371; data collection between December 1997 and May 2008), a sample

---

Table 1. Description of the Samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>Narcolepsy</th>
<th>Cataplexy</th>
<th>HLA-DQB1*06:02</th>
<th>Low CSF Hypocretin-1 Level ≤8 min and ≥2 SOREMPs</th>
<th>Sample Size, No. Before Matching</th>
<th>After Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>+</td>
<td>+/-</td>
<td>+/-</td>
<td>+</td>
<td>169</td>
<td>89</td>
</tr>
<tr>
<td>B</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Unknown</td>
<td>508</td>
<td>427</td>
</tr>
<tr>
<td>C</td>
<td>+</td>
<td>+/-</td>
<td></td>
<td>-</td>
<td>131</td>
<td>118</td>
</tr>
<tr>
<td>D</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>1026</td>
<td>516 and 118</td>
</tr>
<tr>
<td>E</td>
<td>Suspected</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>749</td>
<td>NA</td>
</tr>
<tr>
<td>F</td>
<td>Suspected</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>254</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviations: CSF, cerebrospinal fluid; MSL, mean sleep latency; MSLT, Multiple Sleep Latency Test; NA, not available or not applicable; SOREMPs, sleep-onset rapid eye movement periods; +, present; -, absent.
Two hundred fifty-four successive patients from the Beijing University People’s Hospital (36.6%; n = 93) and St Vincent’s Hospital in Suwon, Korea (63.4%; n = 161) were recruited as a high pretest probability for narcolepsy sample (sample F) between 2005 and September 2011. High pretest probability was defined as clinical excessive daytime sleepiness that subsequently led to the ICSD-2 diagnosis of either narcolepsy/hypocretin deficiency or narcolepsy without hypocretin deficiency, idiopathic hypersomnia, and Kleine-Levin syndrome (central hypersomnia syndromes). Final primary diagnoses were narcolepsy-cataplexy (33.5%; n = 136), narcolepsy without cataplexy (14.2%; n = 36), idiopathic hypersomnia (31.5%; n = 80), and Kleine-Levin syndrome (0.8%; n = 2). This population included 122 patients with narcolepsy/hypocretin deficiency (48.0%) (118 [96.7%] with cataplexy; 4 without cataplexy; 60 [49.2%] with low CSF hypocretin-1 levels; and 62 with cataplexy and HLA-DQB1*06:02 positivity). Self-identified ethnicity was 100% Asian.

All the described samples are independent, except for 6 patients with low CSF hypocretin-1 levels used in comparison 1a and comparison 4.

Clinical and Polysomnographic Data
Demographic data (age, sex, and self-identified ethnic group, based on predefined categories) were collected through the Stanford Sleep Inventory, a standardized questionnaire for sleep disorders. Self-identified ethnic group is reported for descriptive purpose, because of the multiple sites of inclusion, and the fact that ethnicity has an effect on HLA-DQB1*06:02 prevalence (eg, higher prevalence of HLA-DQB1*06:02 positivity in African American individuals). Body mass index was calculated as weight in kilograms divided by height in meters squared. Clinical features were recorded through clinical interviews, notes, and the Stanford Sleep Inventory. Rapid eye movement sleep latency, total sleep time, and apnea-hypopnea index were measured using a standard overnight PSG, performed for all participants; an epoch was considered stage R or REM sleep if it contained low-amplitude, mixed-frequency electroencephalography activity and low chin electromyography tone that was the lowest level in the study or at least no higher than the other sleep stages and either had REMs or was preceded by stage R sleep. Almost all patients and controls had MSLT following PSG according to standard protocol (missing data: 7 of the 1026 controls, 40 of the 808 patients; no missing data for the 254 successive patients), except for the naturalistic sample E. Mean sleep latency and the number of SOREMPs (occurrence of REM sleep in ≤15 minutes after sleep onset, if sleep onset occurs within 20 minutes after dark onset) were recorded. During MSLT, patients from the United States, China, and Italy and controls from the Detroit tricounty area and China always had 5 nap opportunities. Patients from Korea and controls from the Wisconsin Sleep Cohort only had 4 nap opportunities, a fifth nap being added only if REM sleep was detected during 1 of the first 4 naps.

HLA-DQB1*06:02 Typing and CSF Hypocretin-1 Measurement
Genetic typing of HLA-DQB1*06:02 was performed between October 1999 and September 2011 using a sequence-specific poly-
merase chain reaction, as described by Hallmayer et al.25 The CSF hypocretin-1 concentrations were measured between March 2000 and September 2011 as previously described26 and, in accordance with ICSD-2 criteria, used as the reference (gold) standard.

Although the retrospective design prevented us from as-
suring that readers of the index tests (NPSG REML) and the re-
ference standard (CSF hypocretin-1 levels) were universally
blind to the result of the other test, CSF hypocretin-1 levels were measured in the laboratory facility, usually without the knowl-
dge of any NPSG result. Technicians scoring REML are usu-
ally blind to the diagnosis of the patient under study. Further,
given that lumbar puncture and HLA allele typing are typi-
cally done after the sleep study, REML measures were mostly
done without knowledge of hypocretin deficiency. All partici-
pants underwent NPSG and MSLT. The only group that did not
undergo systematic MSLT was the group of successive pa-
tients coming for a sleep disorder evaluation of any type
(sample E), where only patients with a high probability of hav-
ing narcolepsy underwent the MSLT. Although NPSG and MSLT
REM onset parameters were evaluated locally using variable
polysomnography equipment, multiple technicians, and bed-
room settings, these measures are known to be highly
reliable.6,27-29

Statistical Analyses
Statistical analyses were performed using the program R ver-
sion 2.13.2 (The R Foundation). To control for the effect of age
and sex on REM sleep propensity, we used the nearest-
neighbor matching function from the MatchIt package for R,30
with a case/control ratio of 1. The number of standard devia-
tions of the distance measure within which to draw control
units (caliper) was set to 0.25. To compare means or medians
between samples in the 4 comparison groups, we used the Mann-Whitney U test or t test when relevant and used binom-
ial logistic regression to adjust the results for sex and age
(multivariate analysis). Significance level was set at 5% (P < .05),
2-sided.

To determine the REML associated with MSLT values pro-
viding optimal specificity for narcolepsy, we developed a novel
graphical program aimed to assist clinical researchers and pro-
vide optimal diagnostic configurations and cutoff values ac-
cording to test and quality of the receiver operating charac-
teristic (ROC). Interactive plots make it easy to quickly select
and compare diagnostic configurations and verify the quality
of each selection. Bootstrapping and training/test split vali-
dation methods are available to measure 95% confidence in-
tervals of test cutoffs and obtain unbiased ROC values. Our tool
allows the user to evaluate a selected model’s accuracy via a
training/test set validation split and stability by a bootstrap pro-
cedure. Because there are 2 criteria of interest, sensitivity and
specificity, there is no uniformly best model. To select and vali-
date a single model, these criteria must be aggregated. The soft-
ware aggregates these criteria by using a convex combination
of the 2: (1−α) sensitivity + α specificity, for a user selected α
∈ [0, 1]. Users specify α either explicitly by entering a value or
implicitly by choosing what they would like to be the “opti-
mal” point on the full data ROC curve.

Results for sensitivity, specificity, and 95% confidence in-
tervals have all been secondarily confirmed with the epiR pack-
age for R.31 Definitions and short explanations of sensitivity,
specificity, positive predictive value (PPV), and negative pre-
dictive value (NPV) are shown in the eTable in Supplement.32

Analytical Plan
Our goal was to determine if short REML during the NPSG could
be used as a screening test to predict narcolepsy cases associ-
ated with hypocretin deficiency. Age and sex have already been
shown to influence NPSG REML in patients with narcolepsy and
controls6-33 35; we therefore matched patients and con-
trols for these variables before comparisons. As a conse-
quence, the total number of patients and controls was re-
duced to 516 pairs (some patients did not have a matching
control). Optimal REML and MSLT finding cutoffs were first
estimated by ROC curves in case vs control comparisons.

In comparison 1a, we used 89 narcolepsy cases with docu-
mented low CSF hypocretin-1 levels (sample A) and 89 ran-
don population controls (sample D) matched for age and sex.
Only 89 cases of 169 (52.7%) could be matched because of a
higher number of young adults in sample A vs sample D.

In comparison 1b, we used 427 narcolepsy cases with clear
cataplexy and HLA-DQB1*06:02 positivity (sample B) and 427
random population controls (sample D) matched for age and sex.
Not all (427 of 508; 84.1%) cases could be matched because of a
higher number of young adults in sample B vs sample D.

Because studies have shown that 98% of comparison 1b
cases have low CSF hypocretin-1 levels,32 we then merged com-
parisons 1a and 1b into comparison 1. Therefore, we used 516
narcolepsy cases either with documented low CSF hypocre-	in-1 levels (sample A) or with clear cataplexy and HLA-
DQB1*06:02 positivity (sample B) and 516 random population
controls (sample D) matched for age and sex. Not all (516 of 677;
76.2%) cases could be matched because of a higher number
of young adults in samples A and B vs sample D.

To extend our observation, we next performed compari-
sons in clinical samples independent from each other.

In comparison 2, we used a clinical naturalistic sample of
749 patients (sample E) referred to the sleep clinic and hav-
ing undergone NPSG. From these, we selected narcolepsy cases
either with documented low CSF hypocretin-1 levels or with
clear cataplexy and HLA-DQB1*06:02 positivity, and we used
all other patients as controls. This sample was used to assess
characteristics of our REML criterion in clinical practice in sleep
centers. In this sample, narcolepsy cases were not matched
with patients with other sleep disorders to maintain the preva-
ience of the disease (to calculate PPV) and because this sample
consists of successive patients. In this sample, only patients
suspected for narcolepsy had MSLT findings; therefore, ROC
analysis was only performed for optimal NPSG REML cutoffs,
not MSLT findings.

In comparison 3, we used a sample of 254 successive pa-
tients with hypersomnia with high pretest probability for nar-
colepsy (sample F) referred to the sleep clinic for daytime
sleepiness not due to sleep apnea and having undergone NPSG
and MSLT. Among them, we selected narcolepsy cases either
with documented low CSF hypocretin-1 levels or with clear
cataplexy and HLA-DQB1*06:02 positivity, and we used all other patients as controls. This sample was used to assess characteristics of our REML criterion, and to compare it with the MSLT findings, in clinical practice among patients with hypersomnia. Cases were not matched with controls for the same reasons noted in comparison 2.

In comparison 4, we used 118 patients with narcolepsy with low CSF hypocretin-1 levels (sample A) and 118 patients with narcolepsy with normal CSF hypocretin-1 levels (sample C) matched for age and sex as controls. Therefore, the patients with narcolepsy with low CSF hypocretin-1 levels in this comparison partially overlap with the patients with narcolepsy with low CSF hypocretin-1 levels from comparison 1. Seventy-three (61.9%) of the 118 controls with normal CSF hypocretin-1 levels had narcolepsy without cataplexy; for these patients, the main diagnostic criterion was not the presence of cataplexy or low CSF hypocretin-1 level, but MSLT positivity. This induces an inclusion bias, and therefore, ROC analysis was only performed for optimal NPSG REML cutoffs, not MSLT results (mandated to be positive for diagnostic inclusion).

Therefore, the total numbers of participants included in the analysis were 1749 different patients (6 patients were used twice, in comparisons 1 and 4) and 516 controls. Most patients (99.0%; n = 1731, see Table 2 for details) and 17.2% (n = 99) of controls had genetic typing for HLA-DQB1*06:02. Patients with narcolepsy who would have a pathologically abnormal REML during NPSG,14 we focused on establishing an optimal cutoff with high specificity to use diagnostically, with the aim of avoiding conducting the MSLT the following day. We therefore determined inflection points of the ROC curves for best specificity.32

Results

Short REML During NPSG Is a Specific Marker of Narcolepsy

The NPSG REML distribution in narcolepsy/hypocretin deficiency was more skewed to the left than in age- and sex-matched controls (median NPSG REML = 15 minutes and 93 minutes, respectively) (Figure 1). Specificity and sensitivity for different cutoffs, with 95% confidence intervals and numbers of patients who tested positive over the total, in the various comparisons are reported in Table 3.

Figure 2A shows ROC curves in comparisons 1a, 1b, and 1, with optimal cutoffs of REML of 17 minutes or less, 21 minutes or less, and 18 minutes or less, respectively. These results were virtually identical. Figure 2B shows ROC curves in comparisons 2, 3, and 4, with optimal cutoffs of REML of 16 minutes or less, 17 minutes or less, and 8 minutes or less, respectively. Based on these findings, and considering that SOREMPs on the MSLT are defined as REML of 15 minutes or less, we opted for the same criterion for NPSG REML, ie, 15 minutes or less, as a screening tool for hypocretin deficiency prior to the MSLT. Using the 15-minute cutoff (Table 3), specificity was high (82.2%-99.6%) and sensitivity was low (35.7%-57.4%) in all comparisons. Positive predictive value was high (92.1%) in the relevant high-probability sample (comparison 3).

Short REML During NPSG Plus HLA-DQB1*06:02 Positivity In Narcolepsy

Because almost all patients with narcolepsy and low CSF hypocretin-1 levels (approximately 99%) were HLA-DQB1*06:02 positive, we also assessed whether combining HLA-DQB1*06:02 positivity with short REML on the NPSG improved specificity. This analysis could only be performed in comparison 1a (only the patients with documented low CSF hypocretin-1 levels), since at least some of the patients in all the other comparisons were selected based on HLA allele positivity or negativity (see the Methods section). Combining REML of 15 minutes or less and HLA-DQB1*06:02 positivity gave a specificity of 98.9% and a sensitivity of 43.8% (Table 3). This result indicates that using HLA-DQB1*06:02 positivity further increases specificity from 97.8% to 98.9%.

Short REML During NPSG Plus MSLT Findings In Narcolepsy

The MSLT specificity and sensitivity for narcolepsy/hypocretin deficiency were, respectively, 98.6% and 92.9% in comparison 1 and 71.2% and 93.4% in comparison 3, where PPV and NPV were 75.0% and 92.2%, respectively. Optimal MSLT parameters were determined by multiple ROC curves: MSL of 8 minutes or less associated with 1 or more SOREMPs (specificity, 96.1% and sensitivity, 95.6%) in comparison 1 and MSL of 8 minutes or less associated with 3 or more SOREMPs in comparison 3 (specificity, 85.6%; sensitivity, 86.1%; PPV, 84.7%; and NPV, 86.9%) (Table 3). Combining NPSG REML of 15 minutes or less as a screening test, and MSLT positivity if the screening test is negative, resulted in 97.8% specificity and 94.4% sensitivity in comparison 1 and 67.4% specificity, 94.2% sensitivity, 72.8% PPV, and 92.7% NPV in comparison 3. Combining these results with NPSG REML of 15 minutes or less as a complementary test to the MSLT (ie, using an NPSG REML ≤15 minutes as a supplementary SOREMP and therefore having a positive MSLT finding with MSL ≤8 minutes and only 1 SOREMP associated with NPSG REML ≤15 minutes) gave similar results (Table 3), indicating that adding 1 SOREMP to the MSLT when an NPSG SOREMP (REML ≤15 minutes) has been observed (about 50% of cases) provides little improvement in diagnostic value. Finally, the combination of an NPSG REML of 15 minutes or less associated with HLA-DQB1*06:02 positivity as a screening test and MSLT positivity if the screening test is negative resulted in 98.8% specificity and 90.1% sensitivity in comparison 1a (Table 3), the only setting in which the testing of this combination was possible because of sample selection (see the Methods section).

Discussion

We found the MSLT to be an excellent diagnostic tool for narcolepsy/hypocretin deficiency (cases with low CSF hypocretin-1 levels, plus patients with cataplexy and HLA-DQB1*06:02 positivity), reaching a sensitivity of 92.9% and a specificity of 98.6% vs age- and sex-matched random controls (Table 3). In a sample of 170 narcolepsy cases compared with 1913 patients with other sleep disorders, using clinical interview and MSLT as the gold standard, Aldrich et al13 had found
### Table 2. Demographic, Clinical, Polysomnographic, and Biological Characteristics of Comparisons 1a, 1b, 1, 2, 3, and 4a

<table>
<thead>
<tr>
<th>Comparison 1a</th>
<th>Comparison 1b</th>
<th>Comparison 1c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>Controls</td>
<td>Patients</td>
</tr>
<tr>
<td>Narcolepsy, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With cataplexy</td>
<td>87 (97.8)</td>
<td>427 (100.0)</td>
</tr>
<tr>
<td>Without cataplexy</td>
<td>2 (2.2)</td>
<td>NA</td>
</tr>
<tr>
<td>Narcolepsy, No. (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With cataplexy</td>
<td>21.4 (5.7)</td>
<td>156.5 (150.9)</td>
</tr>
<tr>
<td>Without cataplexy</td>
<td>0.3 (2.4)</td>
<td>0.3 (2.4)</td>
</tr>
<tr>
<td>AHI (n = 89)</td>
<td>14.9 (15.2)</td>
<td>4.9 (5.2)</td>
</tr>
<tr>
<td>SOREMPs (n = 427)</td>
<td>3.2 (1.3)</td>
<td>3.2 (1.3)</td>
</tr>
<tr>
<td>CSF Hcrt level, mean (SD), pg/mL</td>
<td>6.9 (4.6)</td>
<td>2.8 (2.2)</td>
</tr>
<tr>
<td>HLA-DQB1*06:02 positivity, vs controls.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients with narcolepsy with low CSF Hcrt levels, or cataplexy and HLA-DQB1*06:02 positivity vs controls.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Patients with narcolepsy and low CSF Hcrt levels vs patients with narcolepsy without cataplexy and making it essential to interpret these test findings within a specific clinical context.36 37 Therefore, our higher sensitivity can be explained by the use of a biological marker as the gold standard, reducing the heterogeneity of our population sample. However, a 1.4% false-positive rate is still high considering the large number of patients who undergo MSLT for suspicion of narcolepsy, leading to a likely overdiagnosis of narcolepsy without cataplexy and making it essential to interpret these test findings within a specific clinical context.36 37 To our knowledge, our study is the first to report on the sensitivity and specificity of NPSG short REML in a large number of narcolepsy/hypocretin deficiency cases in comparison with a random population sample (Table 3). We also studied...
this biological marker in patients with other sleep disorders, patients with central nervous system hypersomnias who had a high pretest probability for narcolepsy, and patients with narcolepsy based on positive MSLT findings but with normal CSF hypocretin-1 levels (unknown etiology). The choice of using REML of 15 minutes or less as a cutoff was based on our finding that optimal specificity was reached with cutoffs between 8 and 21 minutes, depending on the comparison, and the preexisting definition of an SOREMP on the MSLT as REML of 15 minutes or less. Although longer REML cutoffs may be used to maximize sensitivity, our goal was to select a conservative REML value that would remain robust to changes in clinical context (Table 3). Sensitivity was also less of a concern to us because a negative REML finding was to be followed by a regular MSLT, keeping a similar sensitivity.

The occurrence of NPSG short REML in 30% to 40% of narcolepsy cases has already been described. However, in these studies, the gold standard for narcolepsy was based on clinical and MSLT results. Using cases with documented low CSF hypocretin-1 levels (comparison 1a), we observed that an SOREMP at NPSG was highly specific (97.8%) vs controls. This suggests the likelihood of overdiagnosis is low, but sensitivity was poor (43.8%), meaning that half of the patients would be missed with this test. Very similar findings were reported in cases with clear cataplexy and HLA-DQB1*06:02 positivity (comparison 1b), indicating that these 2 groups are equivalent and that, as expected, almost all cases with cataplexy and HLA-DQB1*06:02 positivity likely have low CSF hypocretin-1 levels if measured.

The comparison with random population controls was helpful but not applicable to the clinical context. We therefore examined how REML behaved diagnostically in successive samples with increasing pretest probability for narcolepsy. In successive patients undergoing a sleep study evaluation (ie, primarily patients with sleep apnea also complaining of sleepiness [comparison 2]), the test was similarly accurate (specificity, 99.6%; sensitivity, 35.7%), consistent with the results of a prior study in a clinic population that did not use CSF hypocretin measurements (specificity, 98.5%; sensitivity, 29%). Because the principal differential diagnoses for narcolepsy/hypocretin deficiency are narcolepsy without hypocretin deficiency, idiopathic hypersomnia, and Kleine-Levin syndrome (all requiring an MSLT for the differential diagnosis), we also tested specificity and sensitivity for a short REML in successive patients with central hypersomnia needing MSLTs for differential diagnosis (comparison 3). In this sample, the most clinically relevant for clinicians, an SOREMP at night was more specific than a positive MSLT finding (95.4% vs 71.2%). It also had a higher PPV (92.1% vs 75.0%) for the diagnosis of narcolepsy with hypocretin deficiency. These findings show
Table 3. Characteristics of Different NPSG REML, MSLT, and HLA-DQB1*06:02 Criteria Combinations to Diagnose Narcolepsy/Hcrt Deficiency in Different Settings

<table>
<thead>
<tr>
<th>NPSG REML or MSLT</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison 1a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPSG REML ≤15 min</td>
<td>0.704 (0.625-0.786)</td>
<td>43.8 (33.3-54.7)</td>
<td>39/89</td>
<td>97.8 (92.1-99.7)</td>
</tr>
<tr>
<td>NPSG REML ≤17 min</td>
<td>0.704 (0.625-0.786)</td>
<td>44.9 (34.4-55.9)</td>
<td>40/89</td>
<td>97.8 (92.1-99.7)</td>
</tr>
<tr>
<td>NPSG REML ≤15 min and HLA-DQB1*06:02</td>
<td>0.933 (0.896-0.968)</td>
<td>43.8 (33.5-54.1)</td>
<td>39/89</td>
<td>98.9 (96.7-100.0)</td>
</tr>
<tr>
<td>NPSG REML ≤15 min and HLA-DQB1*06:02 or MSLT MSL ≤8 min and MSL SOREMPs ≥2</td>
<td>0.986 (0.977-0.996)</td>
<td>90.1 (83.6-96.6)</td>
<td>73/81</td>
<td>98.8 (96.4-100.0)</td>
</tr>
<tr>
<td><strong>Comparison 1b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPSG REML ≤15 min</td>
<td>0.820 (0.789-0.850)</td>
<td>52.0 (47.1-56.8)</td>
<td>222/427</td>
<td>99.5 (98.3-99.9)</td>
</tr>
<tr>
<td>NPSG REML ≤21 min</td>
<td>0.820 (0.789-0.850)</td>
<td>53.9 (49.0-58.7)</td>
<td>230/427</td>
<td>99.5 (98.3-99.9)</td>
</tr>
<tr>
<td><strong>Comparison 1c</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPSG REML ≤15 min</td>
<td>0.799 (0.771-0.826)</td>
<td>50.6 (46.3-54.9)</td>
<td>261/516</td>
<td>99.2 (98.5-100.0)</td>
</tr>
<tr>
<td>NPSG REML ≤18 min</td>
<td>0.799 (0.771-0.826)</td>
<td>51.6 (47.2-55.9)</td>
<td>266/516</td>
<td>99.2 (98.5-100.0)</td>
</tr>
<tr>
<td>MSLT MSL ≤8 min and SOREMPs ≥2</td>
<td>0.928 (0.904-0.946)</td>
<td>92.9 (90.6-95.2)</td>
<td>445/479</td>
<td>98.6 (97.6-99.6)</td>
</tr>
<tr>
<td>MSLT MSL ≤8 min and SOREMPs ≥1</td>
<td>0.928 (0.904-0.946)</td>
<td>95.6 (93.8-97.4)</td>
<td>458/479</td>
<td>96.1 (94.4-97.8)</td>
</tr>
<tr>
<td>NPSG REML ≤15 min or MSLT MSL ≤8 min and MSL SOREMPs ≥2</td>
<td>0.960 (0.946-0.974)</td>
<td>94.4 (92.3-96.4)</td>
<td>452/479</td>
<td>97.8 (96.6-99.1)</td>
</tr>
<tr>
<td>MSLT MSL ≤8 min and SOREMPs ≥2 (when NPSG REML ≤15 min counts as 1 SOREMP)</td>
<td>0.988 (0.984-0.993)</td>
<td>93.3 (91.0-95.5)</td>
<td>447/479</td>
<td>98.6 (97.5-99.5)</td>
</tr>
<tr>
<td><strong>Comparison 2a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPSG REML ≤15 min</td>
<td>0.704 (0.524-0.907)</td>
<td>35.7 (16.9-60.8)</td>
<td>5/14</td>
<td>99.6 (91.1-100.0)</td>
</tr>
<tr>
<td>NPSG REML ≤16 min</td>
<td>0.704 (0.524-0.907)</td>
<td>42.9 (16.9-68.8)</td>
<td>6/14</td>
<td>99.6 (91.1-100.0)</td>
</tr>
<tr>
<td><strong>Comparison 3a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPSG REML ≤15 min</td>
<td>0.765 (0.707-0.831)</td>
<td>57.4 (48.1-66.3)</td>
<td>70/122</td>
<td>95.4 (90.4-98.3)</td>
</tr>
<tr>
<td>NPSG REML ≤17 min</td>
<td>0.765 (0.707-0.831)</td>
<td>58.2 (48.9-67.1)</td>
<td>71/122</td>
<td>95.5 (90.4-98.3)</td>
</tr>
<tr>
<td>MSLT MSL ≤8 min and SOREMPs ≥2</td>
<td>0.904 (0.868-0.904)</td>
<td>93.4 (87.5-97.1)</td>
<td>114/122</td>
<td>71.2 (62.7-81.8)</td>
</tr>
<tr>
<td>MSLT MSL ≤8 min and SOREMPs ≥3</td>
<td>0.904 (0.868-0.904)</td>
<td>86.1 (78.6-91.7)</td>
<td>105/122</td>
<td>85.6 (78.4-91.1)</td>
</tr>
<tr>
<td>NPSG REML ≤15 min or MSLT MSL ≤8 min and MSL SOREMPs ≥2</td>
<td>0.931 (0.896-0.957)</td>
<td>94.2 (88.3-97.7)</td>
<td>115/122</td>
<td>67.4 (58.7-75.3)</td>
</tr>
<tr>
<td>MSLT MSL ≤8 min and SOREMPs ≥2 (when NPSG REML ≤15 min counts as 1 SOREMP)</td>
<td>0.921 (0.887-0.953)</td>
<td>93.4 (87.5-97.1)</td>
<td>114/122</td>
<td>69.7 (61.1-77.4)</td>
</tr>
</tbody>
</table>

(continued)
that even in a sample of patients being evaluated for possible narcolepsy (high pretest probability sample) vs other hypersomnias, a short REML (≤15 minutes) has high specificity and PPV and may be considered diagnostic without the need for an MSLT in approximately half of the cases. Patients with a strong suspicion for narcolepsy but a normal REML at night would continue with a 4- to 5-nap MSLT according to typical diagnostic recommendations.38

Using these combined criteria in the most clinically relevant sample, overall sensitivity and specificity of the combined test (REML ≤15 minutes or MSLT MSL ≤8 minutes and MSLT SOREMPs ≥2) were estimated to be 94.2% and 67.4%, respectively, similar to the MSLT alone (93.4% and 71.2%) (Table 3). Although specificity of NPSG REML of 15 minutes or less was very high (95.4%) in these patients, 4.6% would still receive a false-positive diagnosis of narcolepsy/hypocretin deficiency, a result that has to be put in balance with the potential cost savings of avoiding the MSLT. Although the cause of narcolepsy and low and normal CSF hypocretin-1 levels should be treated more carefully, avoiding strong stimulants and drugs of abuse, because it is not even known if their problem is lifelong.

The prevalence of severe excessive daytime sleepiness is 5% in the general population and incidence of narcolepsy has been estimated to be 1.37 per 100 000 person-years, corresponding to about 4000 new cases per year in the United States.43 This implies that tens of thousands of patients likely undergo MSLT annually, either for the diagnosis of narcolepsy or other conditions such as idiopathic hypersomnia, Kleine-Levin syndrome, and other hypersomnias. The prevalence of narcolepsy has been estimated to be 1.37 per 100 000 person-years, corresponding to about 4000 new cases per year in the United States.43 This implies that tens of thousands of patients likely undergo MSLT annually, either for the diagnosis of narcolepsy or other conditions such as idiopathic hypersomnia, Kleine-Levin syndrome, and other hypersomnias.

### Table 3. Characteristics of Different NPSG REML, MSLT, and HLA-DQB1*06:02 Criteria Combinations to Diagnose Narcolepsy/Hcrt Deficiency in Different Settings* (continued)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPSG REML ≤8 min</td>
<td>0.628 (0.558-0.702)</td>
<td>44.1 (35.1-53.0)</td>
<td>52/118</td>
<td>87.3 (81.3-93.3)</td>
</tr>
<tr>
<td>NPSG REML ≤15 min</td>
<td>0.628 (0.558-0.702)</td>
<td>46.6 (37.6-55.6)</td>
<td>55/118</td>
<td>82.2 (75.3-89.1)</td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the curve; Hcrt, hypocretin-1; MSL, mean sleep latency; MSLT, Multiple Sleep Latency Test; NPSG REML, nocturnal polysomnography rapid eye movement sleep latency; NPV, negative predictive value; PPV, positive predictive value; SOREMPs, sleep-onset rapid eye movement periods.

* The MSLT specificity and sensitivity for narcolepsy/Hcrt deficiency are, respectively, 98.6% and 92.9% in comparison 1 and 71.2% and 93.4% in comparison 3. Although the specificity of the MSLT in comparison 3, where PPV and NPV are, respectively, 75.0% and 92.2%, is lower than that of the MSLT in comparison 1, the sensitivity is similar (97.8% to 98.9%). Combining NPSG REML of 15 minutes or less as a screening test and MSLT positivity if the screening test findings are negative results in 98.8% specificity and 90.1% sensitivity in comparison 1. Using these combined criteria in the most clinically relevant sample, overall sensitivity and specificity of the combined test (REML ≤15 minutes or MSLT MSL ≤8 minutes and MSLT SOREMPs ≥2) were estimated to be 94.2% and 72.8% PPV, and 92.7% NPV in comparison 3. Combining NPSG REML of 15 minutes or less with HLA-DQB1*06:02 positivity as a screening test and MSLT positivity if the screening test findings are negative results in 98.8% specificity and 93.1% sensitivity in comparison 1.

Eighty-nine patients with narcolepsy and low CSF Hcrt levels vs 89 controls.

Four hundred twenty-seven patients with narcolepsy-cataplexy and HLA-DQB1*06:02 positivity vs 427 controls.

Five hundred sixteen patients with narcolepsy with low CSF Hcrt levels, or cataplexy and HLA-DQB1*06:02 positivity, vs 516 controls.

Fourteen patients with narcolepsy with low CSF Hcrt levels, or cataplexy and HLA-DQB1*06:02 positivity, vs 735 patients with other sleep disorders.

One hundred twenty-two patients with narcolepsy with low CSF Hcrt levels, or cataplexy and HLA-DQB1*06:02 positivity, vs 132 patients with other hypersomnias (narcolepsy, idiopathic hypersomnia, and Kleine-Levin syndrome) and normal CSF Hcrt levels.

One hundred eighteen patients with narcolepsy and low CSF Hcrt levels vs 118 patients with narcolepsy and normal CSF Hcrt levels (positive MSLT findings).
lepsy or for the objective evaluation of hypersomnolence in the absence of a nocturnal sleep pathology, in which case a maintenance of wakefulness test could also be advised.39 Considering a US Medicare payment of $410.43 per MSLT as of 2011 (conservative estimate, since Medicare payment for MSLT is low relative to other payers), at least $10 million could be saved in direct costs, not including time loss for the patient.

Our results confirm the clinical utility of systematically analyzing nocturnal REML results in patients undergoing NPSG before an MSLT, when narcolepsy is suspected. Indeed, each year, millions of patients undergo polysomnography for the evaluation of sleep apnea, and in all likelihood, patients with unrecognized hypocretin deficiency would be flagged for reevaluation. The prevalence of nocturnal SOREMPs in the general population and patients with other sleep disorders (mostly sleep apnea) is quite low (0.8% and 0.4%, respectively)42; thus, a nocturnal SOREMP should alert the clinician to the possibility of a missed narcolepsy/hypocretin deficiency diagnosis. Reevaluation of cataplexy (which can be missed during clinical interviews), confounding factors (sleep deprivation, circadian misalignment, and shift work), and, if needed, NPSG/MSLT testing, HLA-DQB1:0602 typing, and measurement of CSF hypocretin-1 level would be sequentially considered until the situation is clarified. The prevalence of this syndrome is approximately 1 in 3000, and many cases are obese and with a codiagnosis of sleep-disordered breathing (9 of 14 patients with narcolepsy/hypocretin deficiency had comorbid mild/moderate sleep-disordered breathing in our naturalistic sample).44 Also, a recent study has found that many African American patients have narcolepsy with hypocretin deficiency but no cataplexy, making the diagnosis more difficult in this population.45
Our study has a number of important limitations. First, our control sample was mainly adults, and matching excluded a significant number of pediatric cases. Therefore, the reported findings cannot be extended beyond the adult population. Second, the study is retrospective, and scoring was thus not entirely blind to diagnosis. Technicians may have known when a patient with narcolepsy underwent NPSG and could have been more attentive to NSPG REML. We, however, believe this to be unlikely because short REML is not currently a diagnostic criterion, and scoring often occurs independent of the clinical visit. Third, our financial analysis is approximate, and costs or savings have likely not all been considered. A full 70% of patients suspected for narcolepsy had an NPSG REML of more than 15 minutes and therefore would still have to undergo an MLST, reducing the potential cost saving of avoiding the MLST in these cases. Savings might also be partially reduced because detecting short REML during NSPG requires night technologists scoring “on the fly” and/or an interpreting physician (familiar with the patient) available to decide on whether to proceed with an MLST. Finally, health insurance carriers may still require for a time a positive MSLT finding to provide treatment. Nonetheless, in addition to the evidence provided herein, the potential savings will help to recognize a short REML during NSPG as a diagnostic criterion.

Fourth, environmental laboratory conditions, polysomnography equipment, and scorers varied from site to site, and this could have affected some of the comparisons. The fact that REML scoring is known to be highly reliable across scorers6,24-29 and that optimal REML cutoffs were similar across a diverse set of comparisons argues against clinically meaningful variances. One could also argue that REML can only be optimally measured after a prior night of habituation or in a comfortable environment at home. Indeed, REML is often shorter the second laboratory night after habituation46 or at home vs in the laboratory.47 This effect is, however, very small (a few minutes) and not known to apply to the very short REML range (<15 minutes) observed in patients with narcolepsy/hypocretin deficiency. It is also almost certainly smaller than the effect of an uncomfortable laboratory night on subsequent MSLT findings.47 Future studies aimed at studying REML at home using ambulatory polysomnography for several successive days may, however, be an interesting avenue to develop cheaper diagnostic alternatives.

Fifth, MSLT was performed according to the standard protocols published in 1986, or in 2005, depending on when the patient had been diagnosed in the sleep clinics.14,24 The MSLT guidelines of 1986 state that “Drugs known to affect sleep latency...should be withdrawn for 2 weeks before MSLT testing.”24(p520) These drugs include a number of over-the-counter agents and herbal products with anticholinergic properties, as well as alcohol and prescription drugs. In the community, withdrawing these drugs is the major obstacle to obtaining a valid MSLT result, and these drugs were not withdrawn in the Wisconsin Sleep Cohort sample. The new Practice Parameters for Clinical Use of the Multiple Sleep Latency Test published in 2005 address this difficulty by changing the wording to “REM suppressing medications should ideally be stopped 2 weeks before the MSLT.”41(p1019) Our data on NSPG REML, as well as on MSLT, might therefore be biased for some patients and controls for whom it was impossible to control if the REM-suppressing medication withdrawal recommendation was followed. This would be particularly true when comparing patients with controls (comparison 1), although data from the Wisconsin Sleep Cohort do not indicate a major effect on SOREMP prevalence (unlike shift work) in controls. Further, in our 2 independent naturalistic samples, ie, patients eventually diagnosed with narcolepsy and patients with other diagnoses, both groups were likely to have the same type of medications before NSPG or MSLT. Overall, the stability of the finding across groups makes it very unlikely that medication had a significant effect on the diagnostic validity of short NSPG REML in narcolepsy.

In summary, our findings show that among patients being evaluated for possible narcolepsy, short REML (<15 minutes) at NSPG had high specificity and PPV and may be considered diagnostic for narcolepsy/hypocretin deficiency without the use of an MSLT; however, the absence of short REML requires a subsequent MSLT.
REFERENCES


17. Young T. Rationale, design and findings from the Wisconsin Sleep Cohort Study: toward understanding the total societal burden of sleep disorders and breathing. Sleep Med Clin. 2009;4(1):37-46.


