

Anesthesiology
 1998; 88:579-88
 © 1998 American Society of Anesthesiologists, Inc.
 Lippincott-Raven Publishers

The Sensitivity and Specificity of the Caffeine-Halothane Contracture Test

A Report from the North American Malignant Hyperthermia Registry

Gregory C. Allen, M.D., F.R.C.P.C.,* Marilyn Green Larach, M.D., F.A.A.P.,† Allen R. Kunselman, M.A.,‡ and The North American Malignant Hyperthermia Registry of MHAUS

Background: The caffeine-halothane contracture test (CHCT) is the only recognized laboratory test to diagnose malignant hyperthermia (MH). The authors report the results of their analysis of pooled data from the North American Malignant Hyperthermia Registry database to determine the sensitivity and specificity of the CHCT.

Methods: The MH Clinical Grading Scale was used to identify 32 case subjects who were "almost certain" to be MH susceptible based on clinical criteria alone. Their CHCT results were compared with those of a group of 120 control subjects considered to be at low risk for MH. Diagnostic thresholds of the CHCT were adjusted, and its component tests were combined to generate receiver operating characteristic curves. The maxi-

mal Youden index for each component test was chosen as the diagnostic threshold indicative of MH susceptibility.

Results: The highest sensitivity (97%; 95% CI, 84-100%) was achieved with a two-component test with thresholds of ≥ 0.5 g contracture for 3% halothane, ≥ 0.3 g contracture at 2 mM caffeine, or both, considered positive for MH. The test specificity was 78% (95% CI, 69-85%). The addition of other CHCT component tests did not improve CHCT sensitivity or specificity.

Conclusion: The CHCT achieves high sensitivity and acceptable specificity as a clinical laboratory diagnostic test when it is performed according to published standards. However, it cannot be used as a screening test because of the low prevalence of MH in the general population. (Key words: Epidemiology; measurement techniques; skeletal muscle; susceptibility.)

* Director, The North American Malignant Hyperthermia Registry; Associate Professor of Anesthesia, Department of Anesthesia, College of Medicine, The Pennsylvania State University.

† Senior Research Associate, The North American Malignant Hyperthermia Registry; Associate Professor of Anesthesiology and Critical Care, The Johns Hopkins University School of Medicine.

‡ Research Assistant, Biostatistics Section, Department of Health Evaluation Sciences, Department of Anesthesia, College of Medicine, The Pennsylvania State University.

Received from The North American Malignant Hyperthermia Registry, The Pennsylvania State University, Hershey, Pennsylvania, and Johns Hopkins University School of Medicine, Baltimore, Maryland. Submitted for publication April 15, 1997. Accepted for publication September 2, 1997. Presented in part at the 1994 meeting of the North American Malignant Hyperthermia Group, September 16-18, 1994, Chicago, Illinois. The North American Malignant Hyperthermia Registry received financial support for this study from the Malignant Hyperthermia Association of the United States, and from the Department of Anesthesia, The Pennsylvania State University.

Address reprint requests to Dr. Allen: Department of Anesthesia, College of Medicine, The Pennsylvania State University, P.O. Box 850, Hershey, Pennsylvania, 17033-0850. Address electronic mail to: gallen@psghs.edu

§ Larach MG, Landis JR, Shirk SJ, Diaz M, and the North American Malignant Hyperthermia Registry: Prediction of malignant hyperthermia susceptibility in man: Improving sensitivity of the caffeine halothane contracture test. *ANESTHESIOLOGY* 1992; 77:A1052.

THE caffeine-halothane contracture test (CHCT) is the only generally recognized test for the laboratory diagnosis of malignant hyperthermia (MH). Centers worldwide use one of two protocols—either the European Malignant Hyperthermia Group protocol¹ or the standards published by the North American Malignant Hyperthermia Group.² With a mandate from the North American Malignant Hyperthermia Group, the North American Malignant Hyperthermia Registry was created in 1987 to collect and analyze data from all participating biopsy centers. The Registry published a report on the specificity of the CHCT in 1992.³

In 1992, the Registry published its initial findings on the sensitivity of the North American CHCT§. After more data were collected and analyzed, the North American Malignant Hyperthermia Group met in September 1994 to discuss the results. We now report the analysis of that data, which involves 120 control subjects and 32 case subjects. This represents the first full report of diagnostic thresholds derived from the Registry database, one of the largest disease registries in the world. We also discuss the implications of the results for clinical

Table 1. Diagnostic Centers Included

Cleveland Clinic Foundation
Hahnemann University Hospital
Mayo Clinic
Ottawa Civic Hospital
University of California, Davis
Uniformed Services University of the Health Sciences
University of Manitoba
University of Massachusetts
University of South Florida
University of Texas, Houston

cians who are treating patients thought to be MH susceptible and for future genetics studies of MH.

Materials and Methods

After we received institutional review board approval, we studied all persons reported to the Registry between 15 March 1989 and 19 August 1994 who underwent skeletal muscle biopsy by North American Malignant Hyperthermia diagnostic centers. The diagnostic centers' results were included for analysis if the center had reported at least 10 control subjects and at least 1 case subject (table 1). An expert panel, blinded to CHCT results, excluded subjects according to explicit criteria determined at the study's outset (table 2A and 2B). Later the panel reviewed the results for all muscle strips and excluded those that did not meet specific technical criteria (table 2C). The CHCT results were reviewed only after subjects had been identified.

We searched the entire Registry database and identified case subjects using the MH Clinical Grading Scale as the clinical case definition of MH susceptibility;⁴ the scale does not rely on CHCT results. A case subject was ranked D6 on the scale; *i.e.*, "almost certain" to be MH susceptible. A control subject (D1) was considered to be at low risk for MH susceptibility (negative personal and family history of MH, no history of myopathy) and underwent muscle biopsy during an unrelated surgical procedure, such as total hip arthroplasty.

All diagnostic centers followed the protocol for *in vitro* testing as published in 1989 by the North American Malignant Hyperthermia Group.² The halothane and caffeine contracture tests were done in triplicate; an abnormal response in any muscle strip was considered diagnostic for MH susceptibility. The protocol defined an abnormal muscle contracture response as one of the following: for the halothane contracture test, 0.2–0.7

g contracture after exposure to 3% halothane for as long as 10 min; for the caffeine contracture test, (1) ≥ 0.2 g contracture at 2 mM caffeine, (2) caffeine-specific concentration (CSC) < 4 mM caffeine, or (3) $> 7\%$ increase in tension at 2 mM compared with maximal tension generated at 32 mM caffeine (percentage maximal response). These diagnostic thresholds were set by consensus at the 1987 meeting of the North American Malignant Hyperthermia Group. Each center was to determine its diagnostic threshold or cutpoint within the range agreed on for 3% halothane and to determine what method it would use to interpret the caffeine contracture test.²

We compared the case and control responses to 3% halothane alone and to caffeine alone. This allowed us to calculate the sensitivity and specificity of the CHCT, assuming that the cases were truly MH susceptible and that the controls were not. These calculations used combinations of the halothane contracture test and the three methods of interpreting the caffeine contracture test (2 mM caffeine, CSC, percentage maximal response) to form two-, three-, and four-component test strategies.

Demographic and muscle strip characteristics of case and control groups after exposure to halothane and caffeine were compared. Associations between MH sta-

Table 2. Exclusion Criteria

A. Case subject exclusions	
Probable other medical cause for adverse signs (<i>e.g.</i> , chorioamnionitis)	
Only arterial blood gas results taken during cardiopulmonary resuscitation were available for D6 score	
Achieved D6 only when diagnostic muscle biopsy was obtained in a family member following an adverse event in case subject	
No recorded exposure to MH-triggering anesthetic agents	
Nonvastus muscle biopsy	
Incomplete reporting	
B. Control subject exclusions	
Subject experienced a problem with prior anesthetic exposure	
Family history of MH	
Abnormal muscle histology or histochemistry (<i>e.g.</i> , myopathy)	
Nonvastus muscle biopsy	
Incomplete reporting	
C. Muscle strip exclusions	
Poor sample (<i>e.g.</i> , no twitch to electrical stimulation)	
Technical problem with test	
Muscle strips tore during test	
Nonstandard test (<i>e.g.</i> , use of 0.25 mM caffeine dose)	
Incomplete reporting	

SENSITIVITY AND SPECIFICITY OF CHCT

tus (MH case *vs.* control subjects) and dichotomous variables, controlling for the center where the subject had the biopsy, were assessed using Mantel-Haenszel methods.⁵ Differences between MH status for continuous measures were assessed using mixed-effects analysis of variance models, in which the individual subject was considered a random effect. To accommodate skewed distributions and to meet modeling assumptions, appropriate transformations of the dependent variable were used when necessary. For ease of interpretation, untransformed data are reported as a mean \pm 1 SD.

Demographic and muscle strip characteristics that appeared to differ between case and control subjects ($P < 0.005$ after multiple comparison testing) were further investigated. Characteristic tests were constructed using combinations of these factors to determine if these characteristic tests performed better than the component tests. For multiple-factor combinations of the characteristics, logistic regression was used to model MH status. In addition, logistic regression models were fit to determine if a component test was confounded with MH status in the presence of these demographic and muscle strip characteristics.⁶

Receiver operating characteristic curves were generated for each component and characteristic test, varying the diagnostic thresholds of each test. The diagnostic thresholds for multiple-factor characteristic tests were based on the predicted probabilities that the subjects would have MH based on the logistic regression models. The area under the receiver operating characteristic curve for each test was calculated and compared.⁷ The diagnostic threshold indicative of MH susceptibility for each test was selected based on the maximal Youden index (sensitivity + specificity - 1).⁸ Comparison of sensitivity and specificity estimates between tests at selected diagnostic thresholds were done using generalized estimating equations in marginal regression models to accommodate unbalanced data and correlation between tests on the same persons.⁹

In this article, the sensitivity of the CHCT is defined as the percentage of positive test results in a population of "almost certain" (D6) MH susceptible subjects. Specificity refers to the percentage of negative CHCT results in a population of control, low-risk (D1) subjects. Thus an MH-susceptible subject who had a positive CHCT result would be described as a "true positive;" if the same MH-susceptible subject had a negative CHCT result, that result would be labeled as a "false negative." The false-negative rate (%) is calculated using the formula: $100 - \text{sensitivity}$. Similarly, if a subject who truly

Table 3. Subject Exclusions

	Control Subjects	Case Subjects	Total
Total reviewed	298 (1,333)	63 (680)	361 (2,013)
Excluded by center	109 (368)	13 (59)	122 (427)
Excluded by subject	69 (318)	18 (79)	87 (397)
Excluded by muscle strip	(40)	(359)	(399)
Subjects included in analysis	120 (607)	32 (183)	152 (790)

Values are number of subjects (number of muscle strips).

was not MH susceptible had a positive CHCT result, that result would be called a "false positive." The false-positive rate (%) of a test is calculated by the formula: $100 - \text{specificity}$.

All statistical analyses were performed with the SAS statistical package (SAS Institute, Cary, NC).

Results

We reviewed data from 19 diagnostic centers. We excluded data from nine centers for the following reasons: fewer than 10 control subjects (five centers), no case subjects (four centers), and use of non-vastus muscle (one center). This excluded data on 122 subjects (109 controls, 13 cases). Then we applied the subject and muscle strip exclusion criteria to the data from the remaining 10 centers (table 3). After these exclusions, data from 120 control subjects (607 muscle strips) and 32 case subjects (183 muscle strips) remained for analysis.

Table 4 lists the characteristics of the two study groups. Case subjects were younger than control subjects and were more likely to be male. There was no difference in subject body build between groups. Muscle strips from case subjects did not differ from control subjects with respect to length, wet weight, or cross-sectional area. Muscle twitch height immediately before testing was higher in case subjects, and the interval from biopsy to the beginning of the CHCT was shorter.

Table 5 lists the sensitivity and specificity of the 3% halothane contracture test and the 2 mm caffeine contracture test at varying diagnostic thresholds. These data are displayed in figures 1A and 1B, where the areas under the curves are 0.908 and 0.859, respectively. For diagnostic testing of a potentially lethal condition, a high degree of sensitivity (no false negatives) and an

Table 4. Demographic and Muscle Strip Characteristics

	Control Subjects (n = 120) [mean (SD)]*	Case Subjects (n = 32) [mean (SD)]*	P Value†
Demographics			
Age (yr)	53 (21)	25 (16)	<0.001
Gender (% male)	51	88	0.003§
Race (% white)	93	78	0.16§
Body build (% normal)‡	41	50	0.54§
Muscle strips exposed to halothane			
Length (cm)	1.77 (0.56)	1.88 (0.55)	0.31
Wet weight (g)	0.10 (0.05)	0.10 (0.05)	0.33
Cross-sectional area (CSA) (cm ²)	0.06 (0.04)	0.05 (0.04)	0.19
Biopsy-test interval (h)	2.41 (1.07)	1.91 (0.84)	0.01
Predrug twitch tension (g weight)	1.62 (1.92)	2.76 (2.99)	0.01
Predrug twitch tension/CSA (g weight/cm ²)	39.25 (52.70)	96.85 (131.27)	0.001
Muscle strips exposed to caffeine			
Length (cm)	1.62 (0.57)	1.79 (0.57)	0.12
Wet weight (g)	0.10 (0.04)	0.10 (0.05)	0.35
Cross-sectional area (CSA) (cm ²)	0.07 (0.05)	0.06 (0.04)	0.09
Biopsy-test interval (h)	2.63 (1.21)	2.22 (1.53)	0.01
Predrug twitch tension (g weight)	1.56 (1.72)	2.74 (3.22)	0.002
Predrug twitch tension/CSA (g weight/cm ²)	29.04 (33.06)	106.13 (246.93)	0.001

* Mean and one SD reported where applicable; otherwise a percentage is reported.

† Model adjusted P values from a mixed effects analysis of variance model unless otherwise noted.

‡ Biopsy centers chose one of four options: normal, lean, muscular, or obese. No strict criteria were applied.

§ P value from a Mantel-Haenszel chi-square test, controlling for center.

acceptable level of specificity (or false positives) are desired. For the 3% halothane test, this occurred at a cutpoint of ≥ 0.5 g, where sensitivity equaled 87% and

specificity 86%. For the 2 mm caffeine test, a cutpoint of ≥ 0.2 g produced a sensitivity of 84% and specificity of 79%. When the results of these two tests were com-

Table 5. CHCT Sensitivity and Specificity

Halothane 3%* (g)	Sensitivity (n = 30)	Specificity (n = 118)	Caffeine 2 mm† (g)	Sensitivity (n = 32)	Specificity (n = 119)	Combined‡ (g)	Sensitivity (n = 32)	Specificity (n = 117)
0.1	0.97	0.44	0.05	0.91	0.53	0.1	0.97	0.40
0.2	0.93	0.60	0.1	0.88	0.61	0.2	0.97	0.55
0.3	0.93	0.72	0.2	0.84	0.79	0.3	0.97	0.64
0.4	0.87	0.77	0.3	0.69	0.87	0.4	0.97	0.69
0.5	0.87	0.86	0.4	0.63	0.89	0.5	0.97	0.78
0.6	0.80	0.87	0.5	0.59	0.89	0.6	0.94	0.79
0.7	0.73	0.89	0.6	0.59	0.94	0.7	0.88	0.81
0.8	0.73	0.89	0.7	0.56	0.96	0.8	0.88	0.81
0.9	0.73	0.93	0.8	0.50	0.96	0.9	0.88	0.84
1.0	0.70	0.93	0.9	0.50	0.96	1.0	0.88	0.84
1.1	0.67	0.94	1.0	0.50	0.96	1.1	0.84	0.84
1.2	0.67	0.95				1.2	0.84	0.85
1.3	0.67	0.95				1.3	0.84	0.85
1.4	0.67	0.95				1.4	0.84	0.85
1.5	0.63	0.96				1.5	0.84	0.85

* Halothane 3% refers to the contracture response during exposure to 3% halothane alone.

† Caffeine 2 mm refers to the contracture response measured at 2 mm during serial caffeine exposures.

‡ Combined refers to the response to 3% halothane and assumes that the contracture response to halothane and/or caffeine may be abnormal, using a cutpoint for 2 mm caffeine ≥ 0.3 g.

SENSITIVITY AND SPECIFICITY OF CHCT

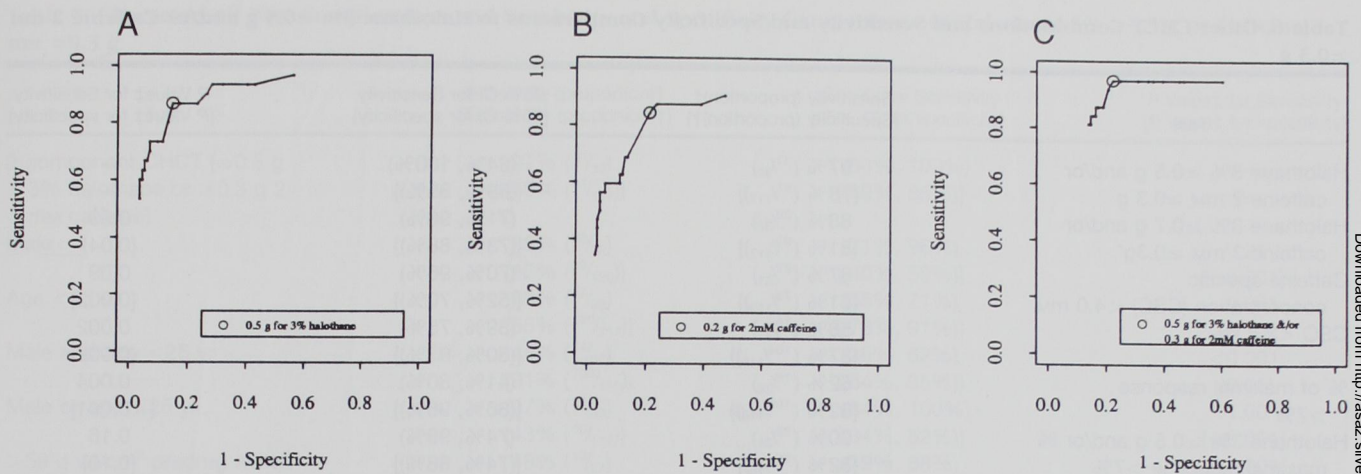


Fig. 1. (A) Receiver operating characteristic (ROC) curve for 3% halothane based on 118 controls and 30 D6 cases. (B) The ROC curve for 2 mM caffeine based on 119 controls and 32 D6 cases. (C) The ROC curve for two-component caffeine-halothane contracture test (3% halothane, 0.3 g 2 mM caffeine) based on 117 controls and 32 D6 cases. Circled values on the ROC curves are the cutpoints chosen using the maximal Youden index.

bined, the cutpoints that produced the highest test sensitivity (97%; 95% CI, 84–100%) and specificity (78%; 95% CI, 69–85%) were ≥ 0.5 g for 3% halothane, ≥ 0.3 g for 2 mM caffeine (fig. 1C, area under curve, 0.885), or both.

Table 6 summarizes the results of other combinations of CHCT data interpretation. This includes the two other methods of analyzing the caffeine contracture test (CSC, percentage maximal response)² and several multiple component tests. None matches the sensitivity and specificity of the two-component CHCT with a cutpoint of ≥ 0.5 g for 3% halothane, ≥ 0.3 g for 2 mM caffeine, or both, while involving only two tests. Although the two-component test of 3% halothane ≥ 0.5 g, percentage maximal response $> 7\%$, or both, had similar rates of sensitivity and specificity, data were lost because the muscle strips frequently tore apart when exposed to 32 mM caffeine.

When comparing the area under the receiver operating characteristic curve of the two-component CHCT to the area under the curve of each of the other component tests, only one, the curve for the halothane, CSC < 4 mm test was significantly smaller ($P = 0.01$). This indicates that using the CSC < 4 mm as one arm of a two-component test does not provide as good diagnostic information as using other measurements of caffeine sensitivity.

Table 7 summarizes combinations of demographic and muscle strip characteristic tests. Cutpoints for the multiple-factor characteristic tests were chosen based

on the maximal Youden index from logistic regression models. Only the characteristic test using age, sex, and predrug twitch tension per cross-sectional area of muscle strips exposed to caffeine had sensitivity and specificity rates that did not differ significantly from the two-component CHCT using 3% halothane and 2 mM caffeine. The area under the curve for the two-component CHCT was greater than the individual characteristic tests of sex ($P = 0.001$) and predrug twitch tension per cross-sectional area exposed to halothane or caffeine ($P < 0.001$).

The coefficient for the two-component CHCT in a logistic regression model using only this test was 0.32 (SE, 0.05). The coefficient for the same test in the presence of the factors of age, sex, predrug twitch tension and cross-sectional area for muscle exposed to halothane and caffeine was 0.34 (SE, 0.08). There was no significant difference between the coefficients or the areas under the curves. This suggests that the two-component test was not confounded with MH status in the presence of age, sex, or predrug twitch tension per cross-sectional area for muscle strips exposed to either halothane or caffeine.

Discussion

In 1987, the North American Malignant Hyperthermia Group reached a consensus on a protocol for performing the CHCT.² At that time it was agreed that a

Table 6. Other CHCT Combinations and Sensitivity and Specificity Comparisons to Halothane 3% \geq 0.5 g and/or Caffeine 2 mm \geq 0.3 g

Test	Sensitivity (proportion)† {Specificity (proportion)†}	95% CI for Sensitivity {95% CI for specificity}	P Value‡ for Sensitivity {P Value‡ for specificity}
Halothane 3% \geq 0.5 g and/or caffeine 2 mm \geq 0.3 g	97% (³¹ / ₃₂) {78% (⁹¹ / ₁₁₇)}	(84%, 100%) {(69%, 85%)}	— —
Halothane 3% \geq 0.7 g and/or caffeine 2 mm \geq 0.3g*	88% (²⁸ / ₃₂) {81% (⁹⁵ / ₁₁₇)}	(71%, 96%) {(73%, 88%)}	0.09 {0.04}
Caffeine specific concentration (CSC) <4.0 mm CSC <2.5 mm	87% (²⁷ / ₃₁) {61% (⁷³ / ₁₁₉)}	(70%, 96%) {(52%, 70%)}	0.09 {0.002}
% of maximal response >7%	58% (¹⁸ / ₃₁) {87% (¹⁰⁴ / ₁₁₉)}	(39%, 75%) {(80%, 93%)}	0.002 {0.002}
Halothane 3% \geq 0.5 g and/or % maximal response >7%	62% (¹⁶ / ₂₆) {92% (¹⁰⁹ / ₁₁₈)}	(41%, 80%) {(86%, 96%)}	0.004 {<0.001}
Halothane 3% \geq 0.5 g and/or CSC <4.0 mm	90% (²⁸ / ₃₁) {82% (⁹⁵ / ₁₁₆)}	(74%, 98%) {(74%, 88%)}	0.16 {0.10}
Halothane 3% \geq 0.5 g and/or caffeine 2 mm \geq 0.3 g and/or % maximal response >7%	94% (³⁰ / ₃₂) {54% (⁶³ / ₁₁₆)}	(79%, 99%) {(44%, 63%)}	0.32 {<0.001}
Halothane 3% \geq 0.5 g and/or caffeine 2 mm \geq 0.3 g and/or CSC <2.0 mm	94% (³⁰ / ₃₂) {76% (⁸⁸ / ₁₁₆)}	(79%, 99%) {(67%, 83%)}	0.32 {0.15}
Halothane 3% \geq 0.5 g and/or caffeine 2 mm \geq 0.3 g and/or CSC <2.0 mm and/or % maximal response >7%	94% (³⁰ / ₃₂) {78% (⁹¹ / ₁₁₇)}	(79%, 99%) {(69%, 85%)}	0.32 {0.99}

CI = confidence interval.

* Diagnostic thresholds accepted by North American MH Group as unequivocally positive CHCT.

† Denominators for proportions vary due to missing data.

‡ P values using generalized estimating equation in marginal regression models⁹ to compare sensitivity and specificity to the 2-component CHCT of halothane 3% \geq 0.5 g and/or caffeine 2 mm \geq 0.3 g.

muscle contracture \geq 0.2 g to 2 mm caffeine would be considered abnormal in determining MH susceptibility. In addition, two other methods of interpreting the caffeine contracture test (CSC, percentage maximal response) were accepted. For the 3% halothane contracture test, a diagnostic threshold was not identified. Instead, a range of thresholds (0.2–0.7 g) was recommended, with each center to determine its own threshold after reviewing the results of at least 30 control subjects.

The present analysis refines the decisions made in 1987 by reviewing pooled data from 10 MH diagnostic biopsy centers. This analysis has determined the optimal diagnostic thresholds of \geq 0.5 g for the 3% halothane contracture test, \geq 0.3 g at 2 mm for the caffeine contracture test, or both (table 5). The combination of these thresholds for a two-component test appears to be superior to other combinations of diagnostic thresh-

olds, test components, or both (table 6). However, when we compared the area of the receiver operating characteristic curve for this test with other component tests, no significant differences were found except for the halothane, CSC <4 mm test.

In this study, we determined the sensitivity and specificity of the CHCT by comparing results from control subjects who were unlikely to be MH susceptible with results from case subjects who were "almost certain" to be MH susceptible. We calculated a sensitivity of 97% (95% CI, 84–100%) for the two-component CHCT (response to 3% halothane alone, to 2 mm caffeine alone, or both) with a specificity of 78% (95% CI, 69–85%). The determination of sensitivity and specificity requires two distinct populations of subjects. The MH Clinical Grading Scale was used as a clinical case definition for MH susceptibility; the scale does not rely on CHCT results to score subjects.⁴ Because MH suscepti-

SENSITIVITY AND SPECIFICITY OF CHCT

Table 7. Characteristic Test Combinations and Sensitivity and Specificity Comparisons to Halothane 3% \geq 0.5 g or Caffeine 2 mm \geq 0.3 g

Test	Sensitivity (proportion) [†] {Specificity (proportion) [†] }	95% CI for Sensitivity {95% CI for specificity}	P Value [‡] for Sensitivity {P Value [‡] for specificity}
2 component CHCT (\geq 0.5 g 3% halothane or \geq 0.3 g 2 mm caffeine)	97% (³¹ / ₃₂) {78% (⁹¹ / ₁₁₇)}	(84%, 100%) {(69%, 85%)}	— —
Male	88% (²⁸ / ₃₂) {49% (⁵⁹ / ₁₂₀)}	(71%, 96%) {(40%, 59%)}	0.21 {<0.001}
Age <25 yr	53% (¹⁷ / ₃₂) {86% (¹⁰⁰ / ₁₁₇)}	(35%, 71%) {(78%, 91%)}	0.001 {0.12}
Male and age <25 yr	44% (¹⁴ / ₃₂) {91% (¹⁰⁶ / ₁₁₇)}	(26%, 62%) {(84%, 95%)}	<0.001 {0.01}
Male or age <25 yr	97% (³¹ / ₃₂) {43% (⁵⁰ / ₁₁₇)}	(84%, 100%) {(34%, 52%)}	1.00 {<0.001}
>59 g wt/cm ² predrug twitch tension/cross-sectional area exposed to halothane	48% (¹³ / ₂₇) {79% (⁹¹ / ₁₁₅)}	(29%, 68%) {(71%, 86%)}	<0.001 {0.83}
>33 g wt/cm ² predrug twitch tension/cross-sectional area exposed to caffeine	57% (¹⁷ / ₃₀) {73% (⁸⁵ / ₁₁₆)}	(37%, 75%) {(64%, 81%)}	0.002 {0.44}
Age + gender*	97% (³¹ / ₃₂) {65% (⁷⁶ / ₁₁₇)}	(84%, 100%) {(56%, 73%)}	1.00 {0.03}
Age + gender + predrug twitch tension/cross-sectional area exposed to halothane*	100% (²⁷ / ₂₇) {64% (⁷² / ₁₁₂)}	(87%, 100%) {(55%, 73%)}	— {0.03}
Age + gender + predrug twitch tension/cross-sectional area exposed to caffeine*	83% (²⁵ / ₃₀) {84% (⁹⁵ / ₁₁₃)}	(65%, 94%) {(76%, 90%)}	0.11 {0.23}
Age + gender + predrug twitch tension/cross-sectional area exposed to halothane and caffeine*	100% (²⁷ / ₂₇) {65% (⁷² / ₁₁₁)}	(87%, 100%) {(55%, 74%)}	— {0.03}

CI = confidence interval.

* Logistic regression models with the characteristics listed as predictors of MH susceptibility. The test was dichotomized at the maximal Youden index from this model.

[†] Denominators for proportions vary due to missing data.

[‡] P values using generalized estimating equation in marginal regression models⁹ to compare sensitivity and specificity to the 2-component CHCT.

bility is expressed under general anesthesia with triggering agents, phenotyping subjects accurately is difficult. Before 1994, no consensus definition for MH existed, which made the comparison of results among subjects or biopsy centers highly subjective.

The MH Clinical Grading Scale is a great improvement over personal opinion as to what constitutes an MH episode. The scale has certain limitations—some D6 cases may have been excluded because of insufficient data reporting, most often laboratory results (arterial blood gases, serum potassium, urine myoglobin). The scale is somewhat subjective because it relies on the reporting clinician's suspicion that a given sign is abnormal and inappropriate. However, when the scale was

created, its developers believed that the reporting clinician must have some discretion about the appropriateness of the signs. Only the clinician observing the event can consider a clinical sign in real-time and in light of the patient's premorbid condition, the surgical procedure, and any medications.

Our results may be limited by our exclusion of other subjects. For example, we excluded biopsy centers that did not report at least one case subject and ten control subjects, biopsy reports that were incomplete, and tests that did not adhere to the published CHCT standards. However, this was the only way to ensure a uniform study population and reduce possible bias.

Control subjects represent a nonrandom surgical pop-

ulation chosen because their surgical procedure allows an incidental muscle biopsy to be performed, such as total hip arthroplasty. Because of this, control and case subjects differ in age (mean age, 53 vs. 25 yr; table 4), but all control subjects were ambulatory and had negative personal and family histories. Malignant hyperthermia appears to be a disorder that more frequently affects young males,¹⁰ so it is not surprising that the case subjects were predominantly men. However, the results of the two-component CHCT were not confounded by subject age or sex, supporting the validity of comparing the case and control subjects as we have done.

Other studies support the high sensitivity of the CHCT performed according to the North American standards. In swine studies, in which each animal's MH status is known, the CHCT has been accurate.¹¹ The major advantage of animal studies is the ability to use anesthetic challenge with triggering agents as a final test of MH susceptibility. Another swine study questioned the role of the RYR1 mutation in MH susceptibility based on the animals' lack of response to halothane and succinylcholine.¹²

Recently, the European Malignant Hyperthermia Group published a report of the performance of the CHCT according to their protocol.¹³ Using the Clinical Grading Scale,⁴ 20 centers provided data on 105 patients thought to be MH susceptible who were rated as D6 and 202 low-risk subjects who underwent muscle biopsy and contracture testing. The sensitivity of the European protocol was determined to be 99% (95% CI, 94.8–100%); the specificity was 93.6% (95% CI, 89.2–96.5%). It was not clear how the 202 low-risk subjects were selected because the European Malignant Hyperthermia Group does not have a central registry, but all met criteria similar to those that we used in this study. Only 13 of 20 centers included D6 patients and low-risk subjects; in the present study, only centers that provided both types of subjects were included.

The European and North American CHCT protocols are similar but not identical. The European protocol uses slightly different caffeine concentrations for the caffeine contracture test.¹³ More importantly, it uses an incremental dosing technique for the halothane contracture test (0.5%, 1%, 2%, 3%) instead of the single 3% halothane exposure used in the North American protocol.

The European protocol uses a uniform diagnostic threshold of ≥ 0.2 g, with an abnormal response occurring at a caffeine concentration ≤ 2.0 mM or at a

halothane concentration $\leq 2\%$ (0.44 mM). An abnormal contracture response to caffeine and to halothane must occur for an "MHS" diagnosis to be given to the test results. If an abnormal response occurs only to caffeine or to halothane, the test result is labeled as "MHE" or equivocal. Patients are counseled as if they are thought to be MH susceptible, pending further refinements of various diagnostic tests. The North American protocol considers the CHCT results abnormal if there is an abnormal response either to caffeine or to halothane. The difference in the specificity between the two protocols may be due to the higher likelihood of a false-positive response with the single exposure to 3% halothane used in the North American protocol.¹⁴

Isaacs and Badenhorst reported four cases of possible false-negative CHCT results using the European protocol.¹⁵ All four patients had suspicious episodes under anesthesia, but subsequent CHCT results were negative for MH. One patient underwent muscle biopsy on two different occasions, and both CHCT results were negative. This case report illustrates that false-negative CHCT results may occur. Like nearly all tests used in clinical medicine, a degree of uncertainty exists in CHCT results.¹⁶

False-positive results are also possible with many tests, including the CHCT. Serfas *et al.*¹⁷ studied a large Mennonite family and compared CHCT results for the presence of the Arg614Cys MH mutation on chromosome 19. One subject had a positive CHCT (3.9 g response to 3% halothane; 0.8 g response to 2 mM caffeine) but lacked the chromosome 19 mutation. The authors suggested that the CHCT result was a false-positive one; otherwise there was no linkage of MH susceptibility to the chromosome 19 mutation in this family. Most MH biopsy center directors would argue that such CHCT responses are abnormal and unlikely to be a false-positive result. Perhaps the subject represents a new mutation. Because there is no ethical way to determine the correct diagnosis, clinically the patient would be considered MH susceptible. Unlike swine, there appear to be several mutations that lead to MH susceptibility in humans.¹⁸ The presence of heterogeneity is supported by another study from the United Kingdom of a large kindred in which linkage to the RYR1 region of chromosome 19 was demonstrated, but the Arg614Cys mutation was not present.¹⁹

An important reason for reporting this study is to provide molecular geneticists with the sensitivity and specificity of the CHCT at various thresholds. Genetic studies require subjects who are likely to be MH suscep-

SENSITIVITY AND SPECIFICITY OF CHCT

tible when their CHCT results are positive; *i.e.*, they must have high specificity or a low false-positive rate. The diagnostic thresholds used by geneticists may be different from those used by clinicians.²⁰ We suggest that genetics investigators may wish to use thresholds of ≥ 0.7 g for the halothane contracture test, ≥ 0.3 g for the caffeine contracture test, or both, which have a sensitivity of 88% (95% CI, 71–97%) and a specificity of 81% (95% CI, 73–88%). However, we caution that the confidence intervals for sensitivity and specificity between these thresholds and those recommended for clinical diagnosis (*i.e.*, halothane contracture ≥ 0.5 g, 2 mm caffeine contracture ≥ 0.3 g) are similar. We also believe that, for uniformity, the Clinical Grading Scale should be used as a clinical case definition to identify potentially affected persons for genetics studies.

In September 1994, members of the North American Malignant Hyperthermia Group met to discuss the results of this study at the Fifth Malignant Hyperthermia Biopsy Standards Conference. The group decided that an equivocal range of contracture responses should be adopted for both parts of the two-component test. A positive response to 3% halothane was identified as a contracture ≥ 0.7 g, an equivocal response to be ≥ 0.5 to < 0.7 g, and a negative response as < 0.5 g. A positive response at 2 mm caffeine was set as a contracture ≥ 0.3 g, an equivocal response to be ≥ 0.2 to < 0.3 g, and a negative response as < 0.2 g. This yields a test sensitivity of 88% and a specificity of 81% for an unequivocally positive response. The equivocal category was suggested to give individual biopsy centers more latitude in diagnosing persons as MH susceptible. However, it was agreed that the Registry may use the thresholds of ≥ 0.5 g for 3% halothane and ≥ 0.3 g for 2 mm caffeine for investigations that require maximum test sensitivity.

Because MH is potentially fatal, the thresholds for diagnostic testing require a high degree of sensitivity and an acceptable degree of specificity. That is, physicians are willing to accept false-positive responses to avoid false-negative ones because the consequences of a false-negative diagnosis might be disastrous. Sensitivity and specificity are stable properties of a test; *i.e.*, they usually do not vary with disease prevalence. However, as one deals with less severe forms of a disease and with different patients, test sensitivity and specificity may decrease (spectrum bias) and reduce the utility of the test.²¹ This problem usually occurs when tests are used to screen asymptomatic populations. In the case of the CHCT, because it requires an open muscle biopsy and is performed at a limited number of centers, patients

are carefully selected after their clinical records are reviewed and their pretest probability is already high. There may be differences in the onset and severity of MH among affected individuals, but such a theory cannot be tested in humans. Other factors, both intrinsic and extrinsic to the affected patient, may be involved. However, if MH is inherited in an autosomal dominant pattern, then susceptibility should be all-or-none, and no spectrum of susceptibility should exist. Therefore the estimates of sensitivity and specificity of the North American CHCT presented here are the best estimates available.

The CHCT is not a screening test and is poorly predictive of MH susceptibility when applied to the general population because of the low prevalence of the disorder. However, it has excellent predictive value at intermediate levels of probability (25–50%). In families, for example, CHCT results can be used to accurately predict risk of MH susceptibility once the proband has been correctly identified. Similarly, patients suspected of having an MH episode (intermediate pretest probability) should undergo muscle biopsy and contracture testing.²² Without correctly identifying the proband, many patients and families will be mislabeled as MH susceptible.²³ Such mislabeling may have adverse effects on their ability to obtain optimal medical or dental care, or in career choices such as military service.

In conclusion, we have determined the sensitivity and specificity of the CHCT performed according to the North American Malignant Hyperthermia Group standards. We believe that these results can be used to improve risk assessment in patients and families, and to accurately identify subjects for future studies, especially for molecular genetics. The hope for the development of a less invasive test to screen for MH susceptibility lies in correctly identifying affected persons. This requires muscle biopsy and contracture testing under standardized conditions.

The authors thank Richard Landis, Ph.D., and Tom Hen Have, Ph.D., from the Center for Biostatistics and Epidemiology, for their statistical expertise during preparation of this manuscript; Julian Loke, M.D., and David MacLennan, Ph.D., of the University of Toronto, for their critical reading of the manuscript; and Cindy Brubaker for assistance in preparing the manuscript. The authors also appreciate the insightful discussions that occurred at the 1994 North American Malignant Hyperthermia Group meeting in Chicago that helped to shape this manuscript.

References

1. The European Malignant Hyperpyrexia Group: A protocol for the investigation of malignant hyperpyrexia (MH) susceptibility. *Br J Anaesth* 1984; 56:1267–9

2. Larach MG, for the North American Malignant Hyperthermia Group: Standardization of the caffeine halothane muscle contracture test. *Anesth Analg* 1989; 69:511-5
3. Larach MG, Landis JR, Bunn JS, Diaz M, The North American MH Registry: Prediction of malignant hyperthermia susceptibility in low-risk subjects. *ANESTHESIOLOGY* 1992; 76:16-27
4. Larach MG, Localio AR, Allen GC, Denborough MA, Ellis FR, Gronert GA, Kaplan RF, Muldoon SM, Nelson TE, Ørding H, Rosenberg H, Waud BE, Wedel DJ: A clinical grading scale to predict malignant hyperthermia susceptibility. *ANESTHESIOLOGY* 1994; 80:771-9
5. Kuritz SJ, Landis JR, Koch GG: A general overview of Mantel-Haenszel methods: Applications and recent developments. *Ann Rev Public Health* 1988; 9:123-60
6. Hosmer DW, Lemeshow S: *Applied Logistic Regression*. New York, John Wiley & Sons, 1989
7. Hanley JA, McNeil BJ: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148:839-43
8. Youden WJ: Index for rating diagnostic tests. *Cancer* 1950; 3:32-5
9. Leisenring W, Sullivan Pepe M, Longton G: A marginal regression modelling framework for evaluating medical diagnostic tests. *Stat Med* 1997; 16:1263-81
10. Larach MG, Fuhrmann LJ, Allen GC: The epidemiology of malignant hyperthermia events in North America, Malignant Hyperthermia—Proceedings of the 3rd International Symposium on Malignant Hyperthermia, 1994. Edited by M Morio, H Kikuchi, O Yuge. New York, Springer, 1996, pp 39-42
11. Allen GC, Fletcher JE, Huggins FJ, Conti PA, Rosenberg H: Caffeine and halothane contracture testing in swine using the recommendations of the North American Malignant Hyperthermia Group. *ANESTHESIOLOGY* 1990; 72:71-6
12. Fletcher JE, Calvo PA, Rosenberg H: Phenotypes associated with malignant hyperthermia susceptibility in swine genotyped as homozygous or heterozygous for the ryanodine receptor mutation. *Br J Anaesth* 1993; 71:410-7
13. Ørding H for the European Malignant Hyperthermia Group: In vitro contracture test for diagnosis of malignant hyperthermia following the protocol of the European MH Group: Results of testing patients surviving fulminant MH and unrelated low-risk subjects. *Acta Anaesthesiol Scand* 1997; 41:955-66
14. Ørding H, Bendixen D: Sources of variability in halothane and caffeine contracture tests for susceptibility to malignant hyperthermia. *Eur J Anaesth* 1992; 9:367-76
15. Isaacs H, Badenhorst M: False-negative results with muscle caffeine halothane contracture testing for malignant hyperthermia. *ANESTHESIOLOGY* 1993; 79:5-9
16. Johnson HA: Diminishing returns on the road to diagnostic certainty. *JAMA* 1991; 265:2229-31
17. Serfas KD, Bose D, Patel L, Wrogemann K, Phillips MS, MacLennan DH, Greenberg CR: Comparison of the segregation of the RYR C1840T mutation with segregation of the caffeine/halothane contracture test results for malignant hyperthermia susceptibility in a large Manitoba Mennonite family. *ANESTHESIOLOGY* 1996; 84:322-9
18. Allen GC: Malignant hyperthermia: Recognition and management of susceptible patients. *Curr Opin Anesthesiol* 1996; 9:271-5
19. Wallace AJ, Wooldridge W, Kingston HM, Harrison MJ, Ellis FR, Ford PM: Malignant hyperthermia—A large kindred linked to the RYR1 gene. *Anaesthesia* 1996; 51:16-23
20. MacKenzie AE, Allen G, Lahey D, Crossan M-L, Nolan K, Mettler G, Worton RG, MacLennan DH, Korneluk R: A comparison of the caffeine halothane muscle contracture test with the molecular genetic diagnosis of malignant hyperthermia. *ANESTHESIOLOGY* 1991; 75:4-8
21. Fletcher RH, Fletcher SW, Wagner EH: *Clinical Epidemiology: The Essentials*, 3rd ed. Baltimore, Williams and Wilkins, 1996, pp 54-6
22. Larach MG: Should we use muscle biopsy to diagnose malignant hyperthermia susceptibility? *ANESTHESIOLOGY* 1993; 79:1-4
23. Bachand M, Vachon N, Boisvert M, Mayer FM, Chartrand D: Clinical reassessment of malignant hyperthermia in Abitibi—Temiscamingue. *Can J Anaesth* 1997; 44:696-701