

An International Consensus Study of Neuroleptic Malignant Syndrome Diagnostic Criteria Using the Delphi Method

Ronald J. Gurrera, MD; Stanley N. Caroff, MD; Abigail Cohen, PhD;
Brendan T. Carroll, MD; Francis DeRoos, MD, FACMT; Andrew Francis, PhD, MD;
Steven Frucht, MD; Sanjay Gupta, MD; James L. Levenson, MD; Ahsan Mahmood, MD;
Stephan C. Mann, MD; Michael A. Policastro, MD, FACEP; Patricia I. Rosebush, MScN, MD, FRCP(C);
Henry Rosenberg, MD; Perminder S. Sachdev, MD, PhD; Julian N. Trollor, MD;
Varadaraj R. Velamoor, MBBS; Charles B. Watson, MD, FCCM; and Jayne R. Wilkinson, MD

ABSTRACT

Objective: The lack of generally accepted diagnostic criteria for neuroleptic malignant syndrome (NMS) impedes research and clinical management of patients receiving antipsychotic medications. The purpose of this study was to develop NMS diagnostic criteria reflecting a broad consensus among clinical knowledge experts, represented by an international multispecialty physician panel.

Participants: Eleven psychiatrists, 2 neurologists, 2 anesthesiologists, and 2 emergency medicine specialists participated in a formal Delphi consensus procedure.

Evidence: A core bibliography consisting of 12 prominent, current reviews of the NMS literature was identified by an objective, comprehensive electronic search strategy. Each panel member was given a copy of these references and asked to examine them before commencing the survey process.

Consensus Process: After reviewing the core bibliography, panel members were asked to list any clinical signs or symptoms or diagnostic studies that they believed, on the basis of their knowledge and clinical experience, were useful in making a diagnosis of NMS. In subsequent survey rounds, panel members assigned priority points to these items, and items that failed to receive a minimum priority score were eliminated from the next round. Information about individual panel member responses was fed back to the group anonymously in the form of the group median or mean and the number of members who had ranked or scored each survey item. The a priori consensus endpoint was defined operationally as a change of 10% or less in the mean priority score for any individual item, and an average absolute value change of 5% or less across all items, between consecutive rounds. The survey was conducted from January 2009 through September 2009.

Results: Consensus was reached on the fifth round regarding the following criteria: recent dopamine antagonist exposure, or dopamine agonist withdrawal; hyperthermia; rigidity; mental status alteration; creatine kinase elevation; sympathetic nervous system lability; tachycardia plus tachypnea; and a negative work-up for other causes. The panel also reached a consensus on the relative importance of these criteria and on the following critical values for quantitative criteria: hyperthermia, $> 100.4^{\circ}\text{F}$ or $> 38.0^{\circ}\text{C}$ on at least 2 occasions; creatine kinase elevation, at least 4 times the upper limit of normal; blood pressure elevation, $\geq 25\%$ above baseline; blood pressure fluctuation, ≥ 20 mm Hg (diastolic) or ≥ 25 mm Hg (systolic) change within 24 hours; tachycardia, $\geq 25\%$ above baseline; and tachypnea, $\geq 50\%$ above baseline.

Conclusions: These diagnostic criteria significantly advance the field because they represent the consensus of an international multispecialty expert panel, include critical values, provide guidance regarding the relative importance of individual elements, and are less influenced by particular theoretical biases than most previously published criteria. They require validation before being applied in clinical settings.

J Clin Psychiatry 2011;72(9):1222–1228

© Copyright 2011 Physicians Postgraduate Press, Inc.

Submitted: July 23, 2010; accepted October 11, 2010.

Online ahead of print: June 28, 2011 (doi:10.4088/JCP.10m06438).

Corresponding author: Ronald J. Gurrera, MD, Veterans Affairs Boston Healthcare System, 940 Belmont St (116A), Brockton, MA 02301 (ronald.gurrera@va.gov).

Despite the substantial morbidity and mortality associated with neuroleptic malignant syndrome (NMS), agreement on diagnostic criteria remains elusive. Many criteria sets have been proposed,^{1–12} but differences between them are unresolved, and, without an objective test or marker for the disorder, there is no clear basis for evaluating them. One study¹³ formally compared NMS diagnostic schema and found only modest agreement between 3 of the most widely used criteria sets. The absence of universally accepted diagnostic criteria has limited efforts to identify effective interventions and useful markers of vulnerability.

Previous NMS diagnostic criteria were based on the clinical experience of individuals or small groups and/or analyses of published case reports. The absence of known specific risk factors, very low incidence, and unpredictable occurrence have made prospective studies impractical. The Delphi technique was developed to assist decision-making “when accurate information is unavailable or expensive to obtain, or evaluation models require subjective inputs to the point where they become the dominating parameters”^{14(p10)} and is particularly well suited to the challenge of developing a consensus among knowledgeable clinicians regarding diagnostic criteria for NMS.

Dalkey¹⁵ identified several factors that limit the usefulness of information gained from interrogating a group of experts when their expert knowledge is primarily in the form of opinion: (1) group opinion is likely to be highly influenced by the views of the group member who does the most talking, (2) there is no significant correlation between success in influencing the group and competence in the problem being discussed, and (3) the irrelevant or redundant information that attends any discussion obscures the relevant information that is offered. Additional limitations include the tendency of individuals to seek the approval of the group and the reluctance of individuals to change an opinion once it has been publicly expressed.¹⁶ According to Dalkey,¹⁵ the Delphi method minimizes undesirable effects of group interaction by employing 3 key elements: anonymity, controlled feedback, and

- Neuroleptic malignant syndrome (NMS) must be diagnosed promptly to avoid serious injury or death, but there is no consensus on diagnostic criteria for this disorder.
- The Delphi consensus technique is especially useful for developing a consensus among knowledge experts when experimental evidence is lacking or difficult to obtain.
- An international multispecialty panel of clinical experts reached a consensus regarding diagnostic criteria for NMS using the Delphi technique; however, until these consensus criteria are validated by future studies, they should be considered only as an aid to clinical diagnosis and not as the sole basis for excluding a diagnosis of NMS.

statistical “group response.” Participants provide input to 1 or more questions, receive statistically summarized anonymous information about that round of responses, respond again to the question(s), receive statistically summarized anonymous information about the second round of responses, and so on. This iterative process continues until a predetermined criterion is met. Importantly, “there is no particular attempt to arrive at unanimity among the respondents, and a spread of opinions on the final round is the normal outcome. This is a further device to reduce group pressure toward conformity.”^{15(p3)} When the best available information in a topic area is the judgment of knowledgeable individuals, the Delphi method has demonstrated decision-making advantages over traditional conferences, group discussions, and other interactive group activities. The focus in a Delphi study is on the stability of the group opinion rather than on individuals’ opinions, so measuring the group result is superior to measuring individual results.¹⁷

The Delphi method has been applied to many different clinical and health care questions when experimental evidence is lacking or impractical and when a consensus among medical experts is desired, including development of diagnostic criteria.^{18,19} In the present study, an international, multispecialty expert panel of physicians was formed to participate in a Delphi process for the purpose of reaching a consensus regarding clinical criteria for diagnosing NMS.

METHOD

Expert Panel

Recruitment efforts focused on specialists frequently involved in the care of patients with NMS and on broad geographic and medical specialty representation. Psychiatrist panel members were recruited from the Neuroleptic Malignant Syndrome Information Service (NMSIS) Professional Advisory Council, which includes senior psychiatrists who have published clinical studies of NMS, and from the NMSIS staff of volunteer hotline consultants. Nonpsychiatrist members were recruited on the basis of referrals from

national professional societies representing the specialties of neurology, emergency medicine, and anesthesiology. English language fluency was required. Twenty-nine experts were contacted with a detailed explanation of the study and were invited to participate. Of the 20 who responded (69%), 18 agreed to participate; 1 dropped out after the first survey round due to serious illness. The final expert panel comprised 11 psychiatrists, 2 emergency medicine specialists, 2 neurologists, and 2 anesthesiologists. Communications between the study coordinator (R.J.G.) and participants occurred—without disclosing the identities of other participants—via e-mail correspondence with attachments or via a Web-based survey program.²⁰

Core Bibliography

Prior to the survey, each panel member received copies of the most prominent current reviews of NMS, selected as follows. The US National Library of Medicine database was searched with the maximally inclusive parameters “(Neuroleptic Malignant Syndrome [Major_MeSH] OR NMS) AND humans.” The initial search yielded 2,560 NMS-related publications between 1980 and 2008, of which 187 were review-type English-language publications less than 10 years old and 3 or more pages in length. Eleven of these had > 10 citations and appeared in a journal with an impact factor > 2; all were published in 2004 or earlier. One additional study published in 2007 was included despite fewer citations because it presented a comprehensive conceptualization of NMS and appeared in a high-impact journal.

Survey Procedure

The consensus endpoint was defined operationally as (1) a change from one round to the next of 10% or less in the mean priority score for each individual item and (2) a change of 5% or less in the average of absolute-value percent changes in individual mean priority scores across all items. These criteria were established a priori by the study coordinator (R.J.G.) and the statistical consultant (A.C.) to assure that the process would be concluded when both the overall change and the changes in individual items from one round to the next were sufficiently low to indicate that additional rounds were unlikely to produce a better consensus. The acceptance of small amounts of residual change in the terminal round reflects the observation that some oscillation in aggregate scoring is inevitable even when consensus has been reached.²¹ The criteria for carrying any individual item forward to the next round were as follows: In the first 2 rounds, an item was carried forward if it was ranked by at least 2 panel members. In subsequent rounds, in which priority points were assigned, an item was carried forward if it was scored by at least 2 panel members *and* its mean priority score was ≥ 5 (representing at least 5% of the total scoring variance). The rationale for these criteria was that an item endorsed by only 1 panel member, or attracting less than 5% of the total scoring variance, was not clinically important in the view of the expert panel as a whole. These criteria were intended to keep in play any items that the group might

ultimately consider important and to exclude items unlikely to be considered important in any subsequent round.

Panel members were asked initially to review the core references and then to list any clinical signs, symptoms, or clinical diagnostic studies that, on the basis of their knowledge and clinical experience, are useful in making a diagnosis of NMS. They were asked to identify at least 5 distinct features and to list them in descending order of importance or usefulness. For quantitative items, participants were also asked to provide the minimum value at which the finding would be considered unequivocally present. Responses were reviewed by the study coordinator to establish uniform terminology for equivalent features (eg, fever, elevated temperature, and hyperthermia). When phrasing connoted potentially substantive differences (eg, blood pressure oscillation, autonomic instability, and sympathetic nervous system hyperactivity), it was left unchanged.

The survey was conducted from January 2009 through September 2009.

RESULTS

A total of 64 distinct diagnostic features were identified by 1 or more panel members in the preliminary round. Forty-nine of these were ranked by 2 or more panel members and carried forward to the next round.

In the second round, each item was listed along with its median rank and the number of panel members who had ranked it in the previous round. Items were grouped by content (eg, all items related to rigidity were grouped together) in order of decreasing median rank. Participants were asked to select the 7 items they considered most important or valuable in diagnosing NMS and to rank them. Limiting the selection to 7 items balanced the aims of eliminating the least relevant items from further consideration (eg, it is rare for medical disorders to have more than 7 diagnostic criteria) and retaining items that might be important. Sixteen items were ranked by 2 or more experts and carried forward to the next round.

In the third round, each item was presented with its mean rank and the number of participants who had selected it in the previous round. Panel members were asked to assign priority points to 1 or more of these items, indicating importance or value in making a diagnosis of NMS, using a total of 100 points. For quantitative features, participants were also given the mean values of responses from the preliminary round and were asked again to provide their best estimate of the threshold value at which the clinical feature could be considered present. Nine items received a mean priority score ≥ 5 and were carried forward to the next round; 7 items received mean priority scores ≤ 1.7 and were eliminated.

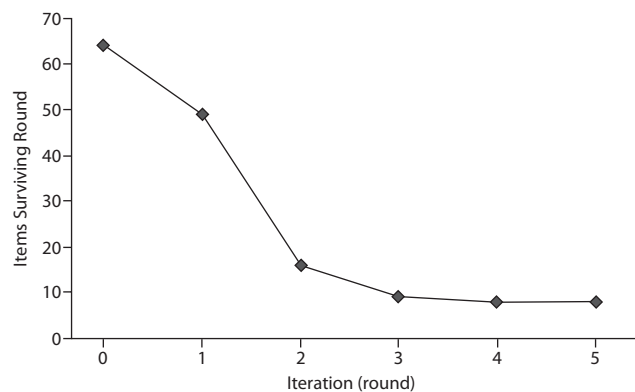
Two items contained duplicate elements that could produce scoring inconsistencies; these were revised prior to round 4 to eliminate redundancies, yielding a list of 9 non-overlapping diagnostic features on which no clinical element appeared more than once. The panel was informed about the mean priority score and the number of individuals scoring

each item on this list and were asked again to assign priority points and critical values. One of the revised items received a mean score of 2.0 and was not carried forward to the next round.

In the fifth and final round, panel members were given mean priority scores and mean critical values from the previous round and the number of respondents who scored each item. Once again, participants were asked to allocate 100 priority points among these items and to provide critical values for quantitative features. Between rounds 4 and 5, 6 items met the individual item consensus criterion (change range, 0.8%–10.2%), and the average absolute value change in mean priority point scores across all 8 items was 6.7%. Two items fell just short of meeting the item-specific change criterion: hyperthermia (13.2%) and creatine kinase elevation (12.5%). On rounds 3 through 5, the creatine kinase elevation item received mean priority scores of 10.5, 9.7 and 10.9, respectively, and the round 5 median score was 10. Given the consistency of these scores over 3 consecutive rounds, the study coordinator and the statistical consultant agreed that consensus had been reached on this item, reflected by a value of 10. Since all priority points sum to 100, only 18 points remained for hyperthermia, and this value equaled the mean of values assigned to hyperthermia across rounds 4 and 5 (17.85). Once again, all 8 diagnostic features received mean priority scores of 5 or greater. Following this analysis, the study coordinator and the statistical consultant agreed that the panel had reached a stable consensus.

Most critical values attained stable consensus by round 4. Changes in mean thresholds for blood pressure elevation above baseline were slightly higher than the a priori consensus criterion (systolic, 11.1%; diastolic 12.0%), but corresponding median values were unchanged. The mean threshold value for heart rate increase changed 12.1% from round 3 to round 4, but, once again, median values were unchanged. The threshold for creatine kinase elevation (defined as multiples of the local laboratory upper limit of normal) satisfied the consensus criterion on round 5, having been assigned mean levels of 4.0 and 3.9 on successive rounds (2.5% change). The only parameter that clearly exceeded a consensus level of change at the conclusion of round 5 was respiratory rate increase. This parameter was given a mean level of 62% above baseline in round 4, and 42% above baseline in round 5, with corresponding median values of 55% and 50%. As this value was the only critical one about which there remained some question as to whether the panel had reached consensus, the study coordinator and the statistical consultant agreed that another round was not necessary. Instead, the median value from the final round was taken as the best estimate of the panel's opinion; this value closely approximated the average of the mean values assigned during rounds 4 and 5.

The mean number of panel members responding to each survey round was 14.6 (85.9%; range, 13–16). The mean round participation rate for each panel member was 4.24 rounds (84.8%; range, 2–5 rounds). Intermittent participation was attributable primarily to vacations and illness.

Figure 1. Candidate Criteria Surviving Successive Iterations of the Survey Process**Table 1. Neuroleptic Malignant Syndrome Diagnostic Criteria: Expert Panel Consensus**

Diagnostic Criterion	Priority Score
Exposure to dopamine antagonist, or dopamine agonist withdrawal, within past 72 hours	20
Hyperthermia (> 100.4°F or > 38.0°C on at least 2 occasions, measured orally)	18
Rigidity	17
Mental status alteration (reduced or fluctuating level of consciousness)	13
Creatine kinase elevation (at least 4 times the upper limit of normal)	10
Sympathetic nervous system lability, defined as at least 2 of the following:	10
Blood pressure elevation (systolic or diastolic $\geq 25\%$ above baseline)	
Blood pressure fluctuation (≥ 20 mm Hg diastolic change or ≥ 25 mm Hg systolic change within 24 hours)	
Diaphoresis	
Urinary incontinence	
Hypermetabolism, defined as heart-rate increase ($\geq 25\%$ above baseline) AND respiratory-rate increase ($\geq 50\%$ above baseline)	5
Negative work-up for infectious, toxic, metabolic, or neurologic causes	7
Total	100

The convergence of panel members' opinions on a set of diagnostic features over the course of 5 iterations is illustrated in Figure 1. The final consensus diagnostic criteria and their relative importance (indexed by mean priority scores) are provided in Table 1.

DISCUSSION

The international multispecialty panel of NMS experts convened for this study reached a consensus regarding the clinical features that are most valuable in making a diagnosis of NMS, the relative importance of these features, and the corresponding critical values.

Previously published diagnostic criteria have been based upon personal impressions or literature reviews, have reflected the opinions of 1 individual or a few individuals, and

have included features considered less relevant by the expert panel (Table 2). The use of formal consensus methodology to formulate diagnostic criteria when a "gold standard" does not exist is an established strategy in health care.²² The diagnostic criteria presented here represent a broad consensus because the panel included knowledgeable clinicians and researchers from different countries, different medical specialties, and different academic centers.

The present study assigned scores to each item corresponding to its importance relative to other diagnostic features. Previous studies have lacked a method for evaluating the relative importance of clinical features, and most made no attempt to do so. For example, one recent study¹² explicitly assigns equal weight to 6 factor-analytically derived domains of NMS phenomenology. Others have employed a major/minor classification scheme,^{2,3} but none have provided individualized estimates of relative importance. According equal salience to diverse clinical features may conflate elements directly related to pathophysiology and secondary or comorbid features, with obvious implications for studies of incidence, prevalence, risk factors, treatment interventions, etc.

In contrast to most prior reports, these criteria provide explicit critical values. These values were also determined by expert consensus and are more conservative than most of those previously published. The application of more conservative and explicitly quantified criteria should improve interrater reliability and diagnostic specificity. Diagnostic criteria reflecting the consensus of clinical knowledge leaders are also more likely to be used consistently in clinical studies and reports, which will make it easier to compare treatment and epidemiologic data.

These diagnostic criteria may also hold less obvious but important advantages for future research on the mechanisms underlying NMS. For example, an ongoing controversy centers on whether NMS is primarily a disorder of central nervous system dopamine regulation or a manifestation of disordered metabolism resulting from abnormal sympathetic nervous system or skeletal muscle function (analogous to malignant hyperthermia). The present consensus criteria do not require a priori acceptance or rejection of any pathophysiologic model because each item is considered independently of other features and was selected solely on the basis of its importance in making a clinical diagnosis of NMS. This theoretical neutrality may promote more systematic and unbiased collection of relevant data in clinical reporting and future research.

Another persistent controversy concerns the relationship of NMS to catatonia spectrum disorders and other extrapyramidal syndromes.²³ The inclusion of catatonic and extrapyramidal elements in NMS diagnostic criteria may confound estimates of co-occurrence and hinder consensus on whether these motor abnormalities are merely associated with NMS or are integral to the disorder. Through the Delphi process, the expert panel eliminated all motor symptoms except rigidity, which has figured prominently in almost every diagnostic scheme that has been published (Table 2). In separating rigidity from other catatonic and extrapyramidal signs, these consensus criteria make it possible to study

Table 2. Comparison of Consensus Neuroleptic Malignant Syndrome Diagnostic Criteria With 12 Previous Criteria Sets^{a,b}

Diagnostic Feature	Current	1	2 ^c	3 ^c	4	5	6	7	8	9	10	11	12
Recent exposure to dopamine antagonist, or dopamine agonist withdrawal	●	◆	○	◇		●		◇	●	●		●	◇
Hyperthermia	●	●	●	●	●		●	●	●	●	●	●	●
Rigidity	●	■	●	●	●		■	■	●	●	■	■	■
Mental status alteration (reduced/fluctuating level of consciousness)	●		●	◆			●	◆	◆	●	◆	◆	◆
Creatine kinase elevation	●		●	●		●	●	●	●	●	●	●	●
Sympathetic nervous system lability (blood pressure elevation/fluctuation, diaphoresis, urinary incontinence)	●		●	●	●	◆	◆	●	◆	●	●	◆	●
Heart-rate increase and respiratory-rate increase	●		◆	◆	■	◆		■	◆	■	■	■	■
Negative work-up for other medical causes	●		○	●		◆	○	●	●	●	●	●	○
Elevated white blood cell count			●	●		●			●	●	●	●	●
Tremor		●		●				■	●	●	■		■
Other severe extrapyramidal symptoms (cogwheeling, oculogyric crisis, dysphagia, dystonia, choreiform movements, etc)		●			●		●	●		◆	◆	◆	◆
Response to bromocriptine, dantrolene, and/or carbidopa/levodopa; or negative autopsy							●						
Metabolic acidosis or hypoxia									●				
Sialorrhea									●				
Hypotension									●				
Mutism										●		■	
Catatonia											●		●
Reduced serum iron or elevated liver enzymes												●	

^aNumbers in the header row correspond to the same-numbered references in this article: (1) Kurlan et al, 1984; (2) Levenson, 1985; (3) Addonizio et al, 1986; (4) Pope et al, 1986; (5) Harsch, 1987; (6) Friedman et al, 1988; (7) Warner et al, 1990; (8) Caroff and Mann, 1993; (9) American Psychiatric Association, 1994; (10) Sachdev et al, 1997; (11) Adityan et al, 1999; (12) Sachdev, 2005.

^bA closed circle (●) denotes that the criterion was present; an open circle (○) indicates that the criterion was present but only implied, not explicit.

A closed diamond (◆) denotes partial correspondence; an open diamond (◇) indicates that the partially corresponding criterion was not implied.

A circle within a square (■) indicates that the criterion was included as 1 element of a larger complex feature. When the authors have proposed criteria for different levels of probability (*definite, probable, etc*), only their *definite* criteria are considered here. Gray shading highlights the criteria identified by consensus in the present study.

^cThese authors designated "major" and "minor" diagnostic criteria.

the co-occurrence of more complex motor syndromes and NMS with fewer diagnostic confounds. However, we note that there are also controversies regarding the definition and phenomenology of these syndromes; for example, rigidity and altered mental status seen in NMS could be considered catatonic signs.

The elimination of other motor signs by the consensus panel should not be construed as the panel's opinion that NMS is not a form of catatonia or extrapyramidal disorder, as others have suggested.²⁴ Rather, the panel's consensus is limited to the clinical features that are most important for operationally diagnosing an NMS episode. Here again, theoretical neutrality may promote more systematic and unbiased collection of relevant data in clinical reporting and future research addressing these nosologic questions. Similarly, laboratory abnormalities associated repeatedly with NMS in previous studies did not survive into the final round, which may signify the need for more research to validate their diagnostic utility. One example is abrupt serum iron reduction, first reported in acute NMS by Rosebush and Stewart.²⁵

It is important to acknowledge the limitations of this study. The Delphi technique is a superior consensus method, but it is based on opinions rather than objective measures. However, there is no biological marker for NMS against which clinical diagnostic criteria might be tested, and the low incidence and unpredictable occurrence of NMS make it difficult to assemble statistically adequate samples for prospective study. The Delphi technique was developed for

situations such as this, when questions must be resolved by expert opinion rather than objective measurement.

In addition to this general limitation, some specific issues merit comment. Perhaps the most salient of these is the potential for biased selection of panel members, which could produce a consensus that does not accurately represent the view of the larger field of experts. We recognize that some bias in panel selection is unavoidable—for example, in who is considered "expert."²⁶ To minimize panel selection bias we actively solicited members from outside the field of psychiatry and from a range of geographic locations and institutional affiliations. We made an effort to invite as many experts as possible, but the recruitment process was constrained by the need to conduct the survey exclusively in English and by the practical limitations of time and other resources needed to administer a survey with multiple iterations. However, many panel members had never met, so participation was not based upon a preexisting concordance of opinion; expert status and willingness to cooperate were the only criteria for panel selection. There is some evidence that individuals who agree to participate in expert panels are similar to those who decline or do not reply.²⁷ We also tried to mitigate potential bias by providing all panel members with a standardized knowledge base selected for salience from the NMS literature on the basis of objective criteria.

The present Delphi process employed a larger panel and more iterations than many of those used to establish diagnostic criteria for other disorders (eg, see Larach et al¹⁸

and Graham et al¹⁹). The reliability of the Delphi process increases with panel size and the number of iterations,²⁸ but there are no studies to suggest what the optimal values for these parameters might be. Using a bootstrap data expansion technique, Akins et al demonstrated that a small panel of “similarly trained experts (who possess a general understanding in the field of interest)”^{17(p37)} provides a stable and reliable estimate of the larger expert group’s opinion in a concentrated field of study. Others have observed that expert panel sizes of 12 to 13 members are sufficient to reach a stable, representative consensus, with larger panels yielding diminishing returns.^{22,29} The strategies employed here for panel creation, survey construction, and iterative feedback were similar to those in other medical applications of the Delphi technique (eg, see Graham et al¹⁹). Thus, the criteria generated by the present study should approximate the latent consensus of all NMS experts, which could not be obtained directly.

It was determined that consensus had been reached even though some change scores between rounds 4 and 5 slightly exceeded a priori study termination criteria. Previous work suggests that “internal system noise,” or an inherent error function, of the Delphi process will inevitably produce an oscillatory change of up to 13% for individual items, no matter how many iterations are conducted.²¹ Notably, this magnitude of fluctuating change occurs even at the mode, which is relatively insensitive to outlier effects.²¹ In the present study, terminal changes in mean individual priority scores were all 13% or less. Thus, in retrospect, our original termination criteria were unrealistically stringent.

Any increase in diagnostic specificity that might be achieved by application of these consensus criteria may entail a reduction in sensitivity, which could have negative clinical implications if associated with delayed recognition. Therefore, until sufficiently validated by future studies, these criteria can be considered only an aid to clinical diagnosis and should not be used as the sole basis for excluding a diagnosis of NMS. This recommendation is consistent with current practice in the application of NMS diagnostic criteria. In subsequent work, we intend to examine whether these criteria, and the priority point metric, can be used to create a valid and reliable diagnostic tool.

Author affiliations: Department of Psychiatry, Harvard Medical School, and Veterans Affairs (VA) Boston Healthcare System, Boston, Massachusetts (Dr Gurrera); Department of Psychiatry (Dr Caroff), Center for Clinical Epidemiology and Biostatistics (Dr Cohen), Department of Emergency Medicine (Dr DeRoos), and Department of Clinical Neurology (Dr Wilkinson), University of Pennsylvania School of Medicine, and Philadelphia VA Medical Center (Drs Caroff, DeRoos, and Wilkinson), Philadelphia; College of Osteopathic Medicine, Ohio University, Athens, and Chillicothe VA Medical Center, Chillicothe, Ohio (Dr Carroll); Department of Psychiatry and Behavioral Sciences, Stony Brook University, The State University of New York, Stony Brook (Dr Francis); Department of Neurology, Mount Sinai Medical School, New York, New York (Dr Frucht); Department of Psychiatry, Upstate Medical University, The State University of New York, Syracuse (Dr Gupta), and Department of Psychiatry, University at Buffalo, The State University of New York, Buffalo (Drs Gupta and Mahmood); Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond (Dr Levenson); Buffalo Psychiatric Center, Buffalo, New York (Dr Mahmood); Department of Psychiatry and Behavioral Sciences,

University of Louisville School of Medicine, Louisville, Kentucky, and Central Montgomery Mental Health & Mental Retardation Center, Norristown, Pennsylvania (Dr Mann); Department of Emergency Medicine, Wright State University Boonshoft School of Medicine, Dayton, Ohio, and Good Samaritan Hospital, Cincinnati, Ohio (Dr Policastro); Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada (Dr Rosebush); Departments of Medical Education and Clinical Research, Saint Barnabas Medical Center, Livingston, New Jersey, and the Malignant Hyperthermia Association of the United States, Sherburne, New York (Dr Rosenberg); School of Psychiatry (Drs Sachdev and Trollor) and Department of Developmental Disability Neuropsychiatry (Dr Trollor), University of New South Wales, Sydney, Australia, and Neuropsychiatric Institute, The Prince of Wales Hospital, Randwick, Australia (Dr Sachdev); Department of Psychiatry, University of Western Ontario Schulich School of Medicine and Dentistry, London, Ontario, Canada (Dr Velamoor); Department of Anesthesia, Bridgeport Hospital, Yale New Haven Health System, Bridgeport, Connecticut (Dr Watson). Dr Mahmood’s current affiliation is the Department of Psychiatry, Indiana University School of Medicine, Indianapolis.

Potential conflicts of interest: Dr Gurrera has served as a volunteer consultant for the Neuroleptic Malignant Syndrome Information Service (NMSIS) and is a member of the NMSIS Professional Advisory Council. Dr Caroff has served as a consultant for Eli Lilly and is a member of the NMSIS Professional Advisory Council. Dr Carroll has received speakers bureau honoraria from Eli Lilly and Pfizer and is an NMSIS hotline volunteer consultant. Dr Francis is a member of the NMSIS Professional Advisory Council. Dr Frucht has served as a consultant for UCB Pharma, Lundbeck, GE Healthcare, Gerson Lehrman Group, and Guidepoint Global and is a member of the NMSIS Professional Advisory Council. Dr Gupta has been a consultant for Eli Lilly; has received research support from Eli Lilly, GlaxoSmithKline, AstraZeneca, Pfizer, Takeda, Forest, Myriad, and Ono; has served on advisory panels for Eli Lilly and Pamlab; has received speakers bureau honoraria from Eli Lilly, GlaxoSmithKline, Novartis, Forest, Merck, and AstraZeneca; and is a member of the NMSIS Professional Advisory Council. Dr Levenson is a member of the NMSIS Professional Advisory Council. Dr Mann is an NMSIS hotline volunteer consultant. Dr Policastro has been an employee of Qualified Emergency Specialists, Cincinnati, Ohio; has been a consultant to the US Drug Enforcement Administration and Dyax Corporation; has received honoraria from the Malignant Hyperthermia Association of the United States and NMSIS; and has served as an advisory panel member for Dyax Corporation and the Malignant Hyperthermia Association of the United States. Drs Rosebush and Rosenberg are members of the NMSIS Professional Advisory Council. Drs Sachdev and Trollor are NMSIS International Partners. Dr Velamoor is a member of the NMSIS Professional Advisory Council. Dr Watson is an NMSIS hotline volunteer consultant. Drs Cohen, DeRoos, Mahmood, and Wilkinson report no financial relationships with any commercial interest. All expert panel members received a small honorarium from NMSIS for their participation.

Funding/support: This study received administrative support from NMSIS (www.nmsis.org) and received financial support from the Shah Educational Awareness Fund of NMSIS and from an independent educational grant (#48804) from AstraZeneca Pharmaceuticals LP to NMSIS. Dr. Cohen was supported by the Outcomes Measurement Methods Core of the Center for Translational Science Award, National Center for Research Resources grant number UL1-RR024134. This work was also supported by the resources of and the use of the facilities at the VA Boston Healthcare System.

Role of sponsor: The Neuroleptic Malignant Syndrome Information Service is a private, nonprofit organization created for the purpose of educating health professionals and the public about neuroleptic malignant syndrome; NMSIS has no financial interest in these diagnostic criteria and exercised no censorship or editorial influence on any portion of the project.

Previous presentations: Presented as a poster at the 18th European Congress of Psychiatry, February 27–March 2, 2010, Munich, Germany; the 65th Annual Scientific Convention of the Society of Biological Psychiatry, May 20–22, 2010, New Orleans, Louisiana; and the 163rd Annual Meeting of the American Psychiatric Association, May 22–26, 2010, New Orleans, Louisiana.

Acknowledgments: The authors thank Jeffrey R. Strawn, MD, of the Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, Ohio, for his assistance in assembling the core bibliography for the expert panel and Marilyn Green

Larach, MD, FAAP, who is affiliated with the North American Malignant Hyperthermia Registry of the Malignant Hyperthermia Association of the United States, Pittsburgh, Pennsylvania, for her helpful comments about the Delphi consensus process. Dr Strawn has received research support from the American Academy of Child & Adolescent Psychiatry and Eli Lilly and has received honoraria from Qualcom Medical Media. Dr Larach has received a consulting fee from the Malignant Hyperthermia Association of the United States to develop an ambulatory surgery center transfer-of-care document for malignant hyperthermia patients and has been reimbursed by the Association for travel expenses to attend scientific conferences.

REFERENCES

1. Kurlan R, Hamill R, Shoulson I. Neuroleptic malignant syndrome. *Clin Neuropharmacol.* 1984;7(2):109–120.
2. Levenson JL. Neuroleptic malignant syndrome. *Am J Psychiatry.* 1985; 142(10):1137–1145.
3. Addonizio G, Susman VL, Roth SD. Symptoms of neuroleptic malignant syndrome in 82 consecutive inpatients. *Am J Psychiatry.* 1986;143(12): 1587–1590.
4. Pope HG Jr, Keck PE Jr, McElroy SL. Frequency and presentation of neuroleptic malignant syndrome in a large psychiatric hospital. *Am J Psychiatry.* 1986;143(10):1227–1233.
5. Harsch HH. Neuroleptic malignant syndrome: physiological and laboratory findings in a series of nine cases. *J Clin Psychiatry.* 1987; 48(8):328–333.
6. Friedman JH, Davis R, Wagner RL. Neuroleptic malignant syndrome: the results of a 6-month prospective study of incidence in a state psychiatric hospital. *Clin Neuropharmacol.* 1988;11(4):373–377.
7. Warner MD, Peabody CA, Boutros NN. A survey of neuroleptic malignant syndrome. *J Nerv Ment Dis.* 1990;178(10):664–665.
8. Caroff SN, Mann SC. Neuroleptic malignant syndrome. *Med Clin North Am.* 1993;77(1):185–202.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Washington, DC: American Psychiatric Association; 1994.
10. Sachdev P, Mason C, Hadzi-Pavlovic D. Case-control study of neuroleptic malignant syndrome. *Am J Psychiatry.* 1997;154(8):1156–1158.
11. Adityanjee, Mathews T, Aderibigbe YA. Proposed research diagnostic criteria for neuroleptic malignant syndrome. *Int J Neuropsychopharmacol.* 1999;2(2):129–144.
12. Sachdev PS. A rating scale for neuroleptic malignant syndrome. *Psychiatry Res.* 2005;135(3):249–256.
13. Gurrera RJ, Chang SS, Romero JA. A comparison of diagnostic criteria for neuroleptic malignant syndrome. *J Clin Psychiatry.* 1992;53(2):56–62.
14. Linstone HA, Turoff M. *The Delphi Method: Techniques and Applications*. Reading, MA: Addison-Wesley; 1975. Reproduced in pdf format in 2002 and available free of charge at <http://is.njit.edu/pubs/delphibook/>.
15. Dalkey NC. *Delphi*. P-3704. Santa Monica, CA: RAND Corporation; 1967.
16. Brown BB. *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts*. P-3925. Santa Monica, CA: RAND Corporation; 1968.
17. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Med Res Methodol.* 2005;5(1):37–48.
18. Larach MG, Localio AR, Allen GC, et al. A clinical grading scale to predict malignant hyperthermia susceptibility. *Anesthesiology.* 1994;80(4): 771–779.
19. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol.* 2003;56(12):1150–1156.
20. SurveyMonkey. User Manual: Customer Guide for Account Navigation, Survey Creation, Distribution & Analysis. <http://s3.amazonaws.com/SurveyMonkeyFiles/UserManual.pdf>. Verified April 6, 2011.
21. Scheibe M, Skutsch M, Schofer J. Experiments in Delphi methodology. In: Linstone HA, Turoff M, eds. *The Delphi Method: Techniques and Applications*, Chapter IV.C. Reading, MA: Addison-Wesley; 1975: 257–281.
22. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess.* 1998;2(3):i–iv, 1–88.
23. Levinson DF, Simpson GM. Neuroleptic-induced extrapyramidal symptoms with fever: heterogeneity of the 'neuroleptic malignant syndrome.' *Arch Gen Psychiatry.* 1986;43(9):839–848.
24. Fink M, Taylor MA. The catatonia syndrome: forgotten but not gone. *Arch Gen Psychiatry.* 2009;66(11):1173–1177.
25. Rosebush P, Stewart T. A prospective analysis of 24 episodes of neuroleptic malignant syndrome. *Am J Psychiatry.* 1989;146(6):717–725.
26. Fink A, Kosecoff J, Chassin M, et al. Consensus methods: characteristics and guidelines for use. *Am J Public Health.* 1984;74(9):979–983.
27. McKee M, Priest P, Ginzler M, et al. How representative are members of expert panels? *Qual Assur Health Care.* 1991;3(2):89–94.
28. Dalkey NC. *The Delphi Method: An Experimental Study of Group Opinion*. RM-5888-PR. Santa Monica, CA: RAND Corporation; 1969.
29. Milholland AV, Wheeler SG, Heieck JJ. Medical assessment by a Delphi group opinion technic. *N Engl J Med.* 1973;288(24):1272–1275.