Consensus Development for Producing Diagnostic Procedure Guidelines: SPECT Brain Perfusion Imaging with Exametazime

James W. Fletcher, Steven H. Woolf and Henry D. Royal

Division of Nuclear Medicine, Department of Internal Medicine, Saint Louis University Health Sciences Center, and Mallinckrodt Institute of Radiology, Division of Nuclear Medicine, St. Louis, Missouri; and Department of Health and Human Services, Public Health Service, Washington, D.C.

A method for developing diagnostic practice guidelines is presented in which a team of experts used a semiquantitative scoring system to reach consensus on a standard procedure for SPECT cerebral perfusion imaging. **Methods:** An expert panel generated a list of elements that they thought were important for the optimal performance of cerebral perfusion imaging as the first phase of a modified Delphi panel technique. Panel members then scored each statement to indicate the importance of that statement for the performance of cerebral perfusion imaging. The scores were recorded for each statement and the average score, s.d. and variance for each statement were determined for each successive panel round. A total of three panel rounds were conducted. The change in average s.d. between scoring rounds was analyzed for significance using both parametric and nonparametric tests. **Results:** The average s.d. decreased by 35% from 2.1 to 1.32 between the first and final panel round. This change in average s.d., which indicated enhanced consensus, was significant at $p < 0.0001$. Following consensus, all statements were grouped into four categories based on average score: *critical elements, important elements, less important elements* and *elements of uncertain importance.* This grouping formed the basis for a guideline summary narrative. Results were generated in 3 mo, at low cost and with clear documentation of rationale. **Conclusion:** Through simple adaptations of this methodology, expert panels that develop practice guidelines can replace informal discussion with systematic scoring methods to rate the quality of evidence, generalizability to practice conditions, appropriate indications and strength of recommendations.

Key Words: consensus development; practice guidelines; Delphi panel; cerebral perfusion; SPECT

J Nucl Med 1994; 35:2003–2010

Physicians have witnessed a growing interest in clinical practice guidelines in recent years. Practice guidelines are official statements issued by medical groups or government agencies that specify the proper care of specific conditions or performance of procedures (*1*). Their development has escalated recently because of growing concerns that inappropriate or unnecessary procedures may be contributing to the rising costs of health care (*2*). The development of guidelines has been promoted by federal and state governments, specialty societies, academic medical centers, payers, employers and others concerned with quality (*3*).

An important step in developing diagnostic practice guidelines is to define the procedure to be examined. Nuclear medicine has faced difficulties with this step because methods for performing and interpreting tests vary widely between medical centers and practitioners. The variability with which nuclear medicine procedures are performed has several disadvantages. Unnecessary variation can hamper quality control, increase costs and interfere with the efficient transfer of research findings to clinical practice. Inconsistent protocols also make it difficult for meta-analysts to combine the data from small studies that lack statistical power into a larger pool of more meaningful data. This inability to synthesize data weakens scientific arguments for performing procedures. Inconsistency also affects physicians, who must decide individually how best to perform procedures and who cannot apply study results to their practices unless they follow the same protocol.

This problem is best remedied by reaching some agreement on a standard protocol for performing the procedure. The authors have called such agreements procedure guidelines. These guidelines can be produced by evidence-based or opinion-based methods (*4*). Evidence-based methods rely on a careful analysis of scientific evidence to determine which techniques produce the best clinical outcomes. This process, which typically involves expert panel meetings and background research by staff, is labor intensive, can require 1 to 3 yr to reach completion and depends on the availability of reliable evidence. Opinion-based methods, in which a group of experts simply reach consensus on

Received Feb. 14, 1994; revision accepted July 12, 1994.

For correspondence or reprints contact: James W. Fletcher, MD, Saint Louis University Health Science Center, Division of Nuclear Medicine, 3635 Vista Avenue at Grand Blvd., P.O. Box 0250, St. Louis, MO 63110-0250.

a protocol, are faster, require fewer resources and have been used for many years by specialty societies and hospital committees. The problems with this approach are that the recommendations may not be evidence based and the process by which consensus was reached is often poorly documented. This makes it difficult for persons outside the process to evaluate the analytic rationale and invites speculation about conflicts of interest and partiality.

Although evidence-based methods offer the best strategy to circumvent these problems, the lack of scientific evidence and the time and resources required to produce evidence-based procedure guidelines suggest the need for an interim process that can produce temporary guidelines based on expert consensus with clear documentation of how that consensus was derived. In this article, the authors present a method for developing recommendations for a standard procedure for SPECT cerebral perfusion imaging in which a team of experts used a semiquantitative scoring system to reach consensus. The results were generated in 3 mo, at low cost and with clear documentation of rationale. The authors believe that similar methods could be used to generate interim procedure guidelines quickly for other imaging tests.

METHODS

Members of an expert panel were chosen because of their contributions to the SPECT cerebral perfusion imaging literature and their past participation in related conferences (*5*) and seminars. Three of the 11 panel members had PhD degrees, one had an MD/PhD degree and the remaining had MD degrees. All of these physician members of the panel were nuclear medicine specialists with additional training in either internal medicine or radiology. After agreeing to serve on the expert panel, the modified Delphi panel technique (*6,7*) was described to each panelist.

The first task of the panel was to generate a list of elements that they thought were important to the performance of cerebral perfusion imaging. Each panelist was sent a fax stating, "*Indicate those elements that are important for the optimal performance of a high quality SPECT Cerebral Perfusion Imaging study with exametazime.*" Instructions were otherwise nondirective. Panelists were allowed to submit as many statements as they wished. No guidance was provided as to potential subtopics (e.g., patient preparation, instrumentation setup or radiopharmaceutical preparation).

First Scoring Round

The submitted written statements were collated by one of the authors (J.W.F.) into specific subtopic areas: (1) radiopharmaceutical, (2) instrumentation and setup, (3) patient preparation, (4) computer setup and acquisition parameters and (5) analysis (processing and quantitation). The statements were not edited, although duplicate or similar statements were eliminated so that only one representative statement appeared in the final listing. The statements were numbered sequentially on a standardized form and faxed to the panelists for scoring. This portion of the process required 1 hr of the panelist's time and was completed in approximately 3 wk.

Panel members were instructed to write a value from 1 to 9 for each statement to indicate the importance of that statement to the performance of the cerebral perfusion imaging. A score of "1" indicated that the statement was definitely not important. A score of "5" indicated that the importance of the statement was "uncertain." A score of "9" indicated that the statement was extremely important. Panel members were also instructed to add any additional statements to the end of the statement list if an item they thought was important was not reflected in any of the statements. Panel members scored each statement for importance using the standard form and returned the forms by fax. The scores from all members of the panel were recorded for each statement, and the average score, s.d. and variance for each statement were determined. Collating the scores required approximately 4 hr and was completed within approximately 3 wk of receiving the scores from the panelists.

Conference Call

The statement list, with added statements and with the scores of all panelists, was returned to the panel members by fax. After an opportunity to review the statements and scores, the panel was convened for a 60-min telephone conference. The purpose of the teleconference was to allow panel members to exchange views on the importance of particular items. The rules of the conference were established to allow each member of the panel who wished to comment on a particular statement to be able to do so.

Prior to the conference, the statements that had the highest s.d. (2.5 or more), reflecting the weakest consensus, were selected for discussion in the conference. Each of these statements was presented for comment to all members of the panel. Panel members were not required to revise or make scoring decisions during the teleconference. The teleconference resembled the approach used by the RAND Corporation to develop appropriateness and necessity scores for selected medical interventions (e.g., coronary artery bypass surgery and angioplasty). RAND panels actually met for at least 1 or more days to accomplish the same tasks as this teleconference.

Second Scoring Round

Following the teleconference, the panel members were again asked to review the statement list and to rescore each of the statements using the same instructions as for the First Panel Round. As in the teleconference, panel members could now see the scores that other members assigned each statement, including the average score, s.d. and variance. They returned the forms via fax, and their individual scores, average score, s.d. and variance for each statement were assembled and calculated.

Agreement and Disagreement

Statements from the second scoring round were also evaluated for agreement using the method described by the RAND Corporation (*8*). The method does not depend on parametric statistics but instead uses an empirical grouping of scores (Table 1). For example, under the most strict definition for agreement (A11S), all 11 panel members assign a score that falls within a specific three-point range (1–3, 4–6, or 7–9). The next level of agreement is more relaxed (A11R) and requires that all panel members assign a score within any three-point range, even if it straddles more than one of the above ranges (e.g., all ratings between 3 and 5). The first level of disagreement (D11S) applies if at least one panel member assigns a score of 1 and at least one assigns a score of 9. The scores provided by panel members for SPECT brain perfusion imaging were stratified according to these eight levels of agreement and disagreement.

TABLE 1
Rand Agreement/Disagreement Definitions

Agreement	A11S:	All 11 of the ratings fell within a single three-point region (1–3, 4–6 or 7–9).
	A11R:	All 11 of the ratings fell within any three-point range.
	A9S:	After discarding one extreme high and one extreme low rating, the remaining nine ratings all fell within a single three-point region (1–3, 4–6 or 7–9).
	A9R:	After discarding one extreme high and one extreme low rating, the remaining nine ratings fell within any three-point range.
Disagreement	D11S:	Considering all 11 ratings, at least one was a 1 and at least one was a 9.
	D11R:	Considering all 11 ratings, at least one fell in the lowest three-point region and at least one fell in the highest (7–9).
	D9S:	After discarding one extreme high and one extreme low rating, at least one of the remaining nine ratings was a 1 and at least one was a 9.
	D9R:	After discarding one extreme high and one extreme low rating, at least one of the remaining nine ratings fell in the lowest three-point region and at least one fell in the highest (7–9).

Statistical Analysis

The change in average s.d. was analyzed for significance using both parametric (Student's unpaired t-test) and nonparametric tests (Mann-Whitney U and Kolmogorov-Smirnov).

RESULTS

The first solicitation generated 73 statements relating to radiopharmaceuticals (9 statements), instrumentation and setup (18 statements), patient preparation (7 statements), computer setup/acquisition (15 statements) and analysis (24 statements). No statements were submitted relating to interpretation of SPECT. Scores for 16 statements suggested weak consensus (s.d. ≥ 2.5), 26 statements received moderate consensus (s.d. = 2.0–2.5) and 31 statements received strong consensus (s.d. < 2.0). The 16 statements with weak consensus (s.d. ≥ 2.5) were the principal focus of the teleconference.

Following the first scoring round, the average score (±1 s.d.) for all statements was 6.62 ± 2.10. The average score (±1 s.d.) following the final round was 6.72 ± 1.32. These results indicate enhanced consensus because the average s.d. decreased by 35% from 2.10 to 1.32. This change in average s.d. was significant at $p < 0.0001$. The nonparametric analysis was more valid because Z-score histograms of the s.d. from both rounds showed a nongaussian distribution. Range values (±1 s.d.) from the first round and final round were 6.3 ± 1.7 and 4.2 ± 2.0. This change in range was also significant at $p < 0.0001$ using the same parametric and nonparametric tests described earlier. The mean score and median value score did not change significantly between the first and final round (mean 1 = 6.63, mean 2 = 6.72; median 1 = 7.0, median 2 = 6.9).

Figure 1 shows the s.d. distribution for all statements following the first and second scoring rounds. The s.d. distribution shifted to the left with successive rounds. Because there was a statistically significant decrease in s.d. between the first and second rounds, it was decided that additional scoring rounds would not be conducted and that a reasonable consensus had been achieved after two scoring rounds.

The final results for each statement are presented in Table 2. *Critical elements* are statements that received an average score of 8.0 or greater, *important elements* received an average score of 7.0 to 8.0, *less important elements* received an average score of 6.0 to 7.0 and *elements of uncertain importance* received a score below 6.0. Most of the critical and important elements related to instrumentation and setup (12 statements), analysis (9 statements), computer setup and acquisition (7 statements), radiophar-

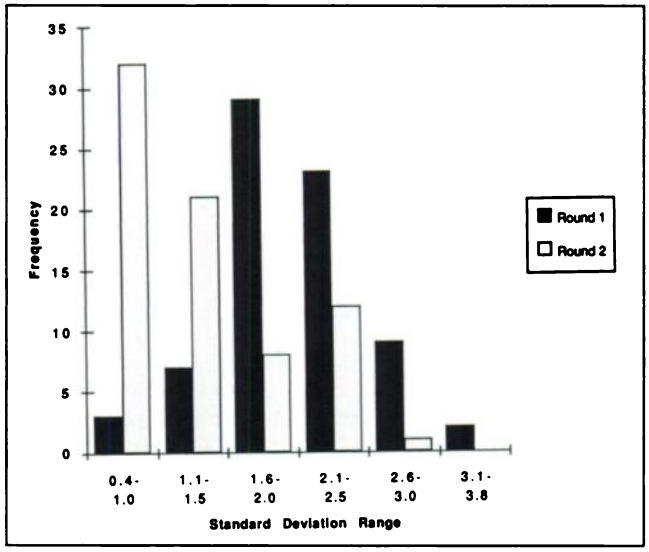

FIGURE 1. The frequency distribution of the average s.d. from all statements from Round 1 of the Delphi panel is compared with Round 2. The distribution has been aggregated into six ranges for ease of comparison. Note that there is a shift to the left and much higher frequency of lower s.d. after successive panel rounds.

TABLE 2
Expert Panel Statement List

Elements of Diagnostic Procedure	Average Score	s.d.	RAND A/D
Critical Elements			
Radiopharmaceutical			
Inject exametazime as soon as possible after preparation and QC.	8.5	0.8	A11S
Use only freshly eluted (not older than 2.0 hr) Tc04.	8.1	0.5	A11S
Employ dose range of 20–30 mCi of Ceretec.	8.0	1.2	A9S
Use short form of QC.	8.0	1.0	A9S
Instrumentation and setup			
Strive for minimum radius of rotation for your camera system (typically 12–14 cm).	8.6	0.5	A11S
Check to make sure complete 360° rotation is unobstructed.	8.5	0.9	A9S
With multihead camera, make sure energy windows on all heads are identically peaked.	8.2	0.9	A9S
Use high-resolution or special-purpose collimation (fan-beam or cone-beam).	8.1	0.7	A11S
Patient preparation			
Employ patient education regarding scan procedure.	8.2	0.4	A11S
Use comfortable patient head restraint device.	8.0	0.6	A11S
Analysis (processing and quantification)			
Tomographic images in transverse, sagittal and coronal planes should be available for interpretation.	8.7	0.5	A11S
After acquisition, review cine of all tomographic views for movement artifacts or other errors; repeat acquisition if necessary.	8.5	0.5	A11S
Have CT or MRI study available to evaluate anatomy.	8.2	0.6	A11S
When using image restoration filters (e.g., Metz), perform SPECT acquisitions of a uniform phantom to ensure that artifacts are not being introduced.	8.2	0.6	A11S
Important Elements			
Radiopharmaceutical			
Inject radiopharmaceutical within 30 min of kit preparation.	7.7	2.3	D11S
Prepare Ceretec immediately before injection and inject within 10 min of preparation using short form of chromatography.	7.5	1.8	D11R
Instrumentation and setup			
Preferably use multihead SPECT system.	7.9	0.9	A9S
For high-quality SPECT cerebral perfusion studies, perform studies on dedicated SPECT systems or multiple head systems.	7.8	1.3	A9S
Give rigorous attention to QC and setup if single-head camera system is used.	7.7	0.8	A9S
Employ conventional QC.	7.6	1.1	A9S
Use easily adjustable table and head rest to allow minimum radius of rotation.	7.6	1.0	A*
Check planar lateral views to make sure temporal lobes and as much cerebellum as possible are in the field of view.	7.5	1.6	D11R
Rotate camera within 1–2 cm of the patient's nose; warn patient.	7.5	0.9	A9R
Patient preparation			
Comfortable positioning of the patient is necessary to prevent patient movement and may require use of arm board extensions, Velcro security band around abdomen and knee and leg support.	7.6	1.4	A*
Administer radiopharmaceutical in a low noise level, dimly lit room under conditions that keep visual, auditory and cognitive stimulation to a minimum for at least 10 min before and after radiopharmaceutical injection.	7.5	1.0	A9R
Computer setup and acquisition			
Monitor patient to ensure no patient movement during imaging.	7.9	0.8	A11S
Choose matrix size so that pixel size is about two to four times smaller than expected camera resolution.	7.8	0.6	A11S
If 128 × 128 matrix, use at least 120 views for angular sampling.	7.7	0.9	A9S
Use manufacturer recommendations as starting point.	7.5	0.7	A11S
Collect projections at 6 degree or finer angular increments using a linear pixel sampling that provides for a 4- to 6-mm sampling in the reconstructed slice.	7.3	1.0	A9R
If 64 × 64 matrix, use at least 60 views for angular sampling.	7.2	1.2	A9R
Use sequential short (4–5 min) acquisitions in agitated patients.	7.0	1.5	A*
Analysis (processing and quantification)			
Use video displays and hard copy for interpretation.	7.8	1.0	A9S
Employ oblique reorientation to obtain transaxial slices at 10° to OM line.	7.5	1.0	A*
Attenuation correction should be used to maintain qualitative relationship between cortical and deep structures.	7.4	1.2	A9S
Ensure that transverse slices are parallel to OM line.	7.0	0.9	A9R

TABLE 2 (continued)

	Average Score	s.d.	RAND A/D
Less Important Elements			
Radiopharmaceutical			
Elute generators on Sunday for work on Monday.	6.0	1.9	D11R
Instrumentation and setup			
Use written manual with meticulous standardized quality control steps.	6.6	1.4	A*
Record radius of rotation.	6.6	1.4	A*
Employ fan-beam collimator rather than parallel collimator.	6.5		D11R
With multihead camera, make sure patient does not feel claustrophobic.	6.5	1.5	D11R
With single-head camera, start from a lateral position with rotation behind the patient's head for the first 180°.	6.5	1.1	A*
Patient preparation			
Insert indwelling venous catheter 10 min before HMPAO injection.	6.8	1.2	A9R
Computer setup and acquisition			
Shorten acquisition to 15–20 min if patient is agitated.	6.7	0.9	A9R
Duration of acquisition should be approximately 30 min.	6.7	1.5	A*
Need to use 3 degrees or less for angular increment as 6 degrees is too coarse.	6.6	1.4	A*
Employ imaging time range between 30–45 min.	6.2	2.0	D11R
Employ minimum of 60 views with minimum of 20 sec/view.	6.0	1.3	A*
Analysis (processing and quantification)			
Limited motion correction for longitudinal movement should be employed if necessary.	6.6	1.0	A9R
Use image-restoration filters for multihead camera because they can yield better image appearance, but correct specification is dependent on system performance.	6.5	1.2	A*
Filter employed should be empirical choice based on actual patient data.	6.4	2.1	D11S
It is useful to have two standardized methods of filtering, one for standard length acquisition and one for shortened acquisitions.	6.4	1.5	D11R
Employ semiquantitation of cortical counts using whole brain/cerebellum for normalization.	6.0	0.8	A11R
Use comparison with age-matched normal database.	6.0	2.0	D11R

OM = orbitomeatal; IV = intravenously; QC = quality control; HMPAO = hexamethylpropyleneamineoxime, and A* = statements that did not fit any of the eight A/D definitions.

continued

maceuticals (6 statements) and patient preparation (4 statements).

Rand Agreement/Disagreement (A/D) Categorization

Table 3 shows the A/D scores. The number of statements and the average panel score and s.d. for the statements in each A/D category are shown. The 37 statements that were in the A11S, A11R, A9S or A9R agreement categories had average s.d. between 0.6 and 1.1. In general, the strict definitions of A (A11S and A9S) had higher average scores than the more relaxed A categories (A11R and A9R). This would imply that the panel members agreed most about the statements that they assigned a higher score to or that they believed were most important. There were 13 statements that did not fit any of the eight A/D definitions, and they are labeled A* in Tables 2 and 3. This group of statements had an average s.d. value of 1.3.

Fifteen statements with an average s.d. of 1.8 were assigned to the D11R category. The average s.d. for the nine D11S and D9R statements (2.4 and 2.2, respectively) were the highest of all A/D categories. The average panel scores for these statements were also the lowest, implying that the panel disagreed most about the statements to which they assigned the lowest scores (i.e., that they thought were less important). The A/D scores for all statements are shown in the last column of Table 2.

Evaluation of the statement list from the second scoring round using the RAND methodology to determine agreement shows an almost identical result in the grouping of statements.

A narrative procedure guideline summary derived from the statements in Table 2 is presented in the Appendix. Statements were selected for this narrative if they were in the *critical* and *important elements* group in Table 2 and showed agreement according to the RAND A/D definitions in Table 1.

DISCUSSION

Formal methods to obtain consensus and collate expert opinion have been used by business leaders and social scientists for decades. The Delphi technique, on which this method is based, was introduced in the 1960s (*9*). Its significance in medicine began in the mid-1980s, when RAND Corporation used a modified Delphi technique to judge the appropriateness of clinical procedures in specific clinical circumstances (*8*). Outside of the work of the RAND Corporation, however, the use of formal consensus-development methods to set medical policy has been limited. This

TABLE 2 (continued)

	Average Score	s.d.	RAND A/D
Elements of Uncertain Importance			
Radiopharmaceutical			
Use package insert QC instructions to test binding of lipophilic compound.	5.0	2.3	D9R
Use minimum of 30-mCi dose.	4.2	2.1	D11S
Instrumentation and setup			
Record angle of pitch with regard to the OM line.	3.5	1.1	A*
Single-head camera system is not optimal for brain imaging and should not be used.	2.8	2.7	D11S
Patient preparation			
Position head by means of light beam over OM line to produce transaxial cuts.	4.8	1.9	D11R
If patient is restless, consider diazepam 2–5 mg IV or morphine 2–5 mg IV.	4.6	2.1	D11R
Computer setup and acquisition			
Begin imaging at least 60 min after exametazime injection.	5.8	2.6	D11S
Initiate imaging at 2 hr following injection.	5.5	2.3	D11R
Imaging should begin 30 min after Ceretec administration.	4.5	1.6	A*
Analysis (processing and quantification)			
Use slice thickness of 0.8–1.0 cm in transverse plane to facilitate comparison to CT or MRI.	5.9	1.1	A9R
Employ quantification procedures.	5.9	2.1	D9R
Three-dimensional smoothing of the entire volume of the reconstructed data yields best appearance.	5.7	1.6	D11R
Display semiquantitative color scale with definition of abnormality threshold.	5.5	2.3	D9R
For single-head camera, use Butterworth filter with frequency cutoff of 0.44–0.5 cycles/cm.	5.1	2.0	D11R
Image-restoration filters should be rarely used because of their resultant noisy images.	5.0	1.2	D11R
Use of contrast enhancement should not be applied to single slices or planes.	4.5	2.3	D9R
For multihead camera, use Butterworth filter with frequency cutoff of 0.6–0.7 cycles/cm.	4.0	1.3	A9R

OM = orbitomeatal; IV = intravenously; QC = quality control; HMPAO = hexamethylpropyleneamineoxime, and A* = statements that did not fit any of the eight A/D definitions.

is largely because physicians are unfamiliar with these methods, those who favor science-based medical policy are uncomfortable with relying on opinion and those who are comfortable with opinion often prefer simple discussion over tedious group voting procedures.

The common tendency of physicians to defend clinical practices based on global subjective judgment (*10*) rather than on explicit scientific arguments is quickly becoming outdated. The health care crisis has motivated policy makers, payers and others within the health care system to scrutinize the rationale for medical procedures. Clinical practices based on informal consensus and vague, poorly documented rationale are increasingly difficult to defend. In response to this trend, practice guidelines are being developed in increasing numbers to lay out clearly the scientific rationale for clinical procedures. In 1989, the federal government established the Agency for Health Care Policy and Research, which has a congressional mandate to develop practice guidelines. More than 40 specialty societies have formed practice guideline committees or task forces. About 1500 practice guidelines have been published or are under review (*3*).

TABLE 3
Rand Agreement/Disagreement Categorization

No. of statements	RAND agreement score	Average panel score	s.d.
13	A11S	8.2	0.6
2	A11R	6.7	0.8
11	A9S	7.9	1.0
11	A9R	6.7	1.1
5	D11S	5.4	2.4
15	D11R	5.8	1.8
4	D9R	5.2	2.2
13	A*	6.4	1.3

A* Did not fit RAND scoring methodology.

In this environment, wide variations in the performance of nuclear medicine procedures are increasingly problematic. Early efforts to develop standardized protocols for diagnostic imaging included the establishment of Workgroups on Development of Standards and Guidelines at the 1992 Radiology Summit Meeting (*11*), the Commission on Standards and Accreditation of the American College of Radiology and the office of Health Care Policy (now the Commission on Health Care Policy) of the Society of Nuclear Medicine. The Society is exploring collaborative relationships between nuclear medicine and other groups that are developing relevant practice guidelines. Ultimately, nuclear medicine will need to adopt evidence-based methods to link diagnostic imaging techniques to the quality of the evidence so that their performance will improve clinical outcomes.

In the meantime, expert opinion must be used to reach consensus on protocols and to reduce unnecessary varia-

tions in clinical practice. The optimal consensus-based approach is (1) *transparent*, providing explicit documentation of how the recommendations were derived; (2) *expedient*, producing recommendations without extensive delays; and (3) *efficient*, generating product at minimal cost. The authors believe the approach described in this article achieves these objectives. The approach is transparent; the importance of procedural elements are defined explicitly on the basis of defined scoring criteria. The process is expedient, requiring about 5 hr of expert time and about 3 to 6 mo to complete the project. It is also efficient; principal costs include the use of a fax machine, a telephone conference call and the time required for data analysis.

The process provides physicians with information of direct relevance to clinical practice. In this case, the process clarified that the most critical components of brain SPECT imaging with ^{99m}Tc-HMPAO are radiopharmaceutical preparation, dose, quality control and timing of injection; using dedicated camera imaging systems fitted with special purpose collimation; educating the patient about the procedure; and preventing patient motion during acquisition. A summary narrative of these components is presented in the Appendix.

This approach is not without its limitations. The conclusions of the panel represent expert opinion and are therefore subject to the biases and other limitations associated with global subjective judgment (*10*). There is no explicit linkage between the conclusions and supporting evidence or outcomes data, making it unclear whether there is a scientific basis for judgments about the importance of procedural elements. The cost effectiveness of the recommendations was not considered. The conclusions are based on the views of 11 experts, and their opinions may not be representative of other authorities. The views of experts may not be applicable to common practice conditions in which equipment and staff limitations may not be optimal or experience with certain imaging procedures may be limited. The conclusions of the panel define which procedural elements are important but do not provide practice guidelines on how to perform the procedure.

Although this process does not produce practice guidelines, the authors believe this technique is applicable to expert panels engaged in guideline development, who must also reach consensus and make group judgments based on opinion. Expert opinion is incorporated into all practice guidelines, and even evidence-based groups are forced to make subjective judgments about the quality of evidence or its generalizability to clinical practice (*12*). Ironically, groups that use some of the most explicit criteria for judging individual studies often rely on informal, poorly documented discussion to reach conclusions about the overall evidence and the wording of recommendations. Through simple adaptations of this methodology, expert panels that develop practice guidelines can replace this informal discussion with systematic scoring methods for rating the quality of evidence, generalizability to practice conditions, appropriate indications and strength of recommendations.

This method and other explicit approaches for capturing expert opinion help to ensure that the independent views of individual panel members are captured accurately and are not overshadowed by their more outspoken colleagues. Explicit criteria also provide persons outside the process with a clear explanation of how the consensus was derived. Future research is needed to refine this methodology and similar approaches, test internal and external validity, compare the reliability of conclusions between different panels, introduce methods for linking expert opinions to scientific evidence and collect outcomes data on whether formal consensus-development methods achieve better results than informal discussion. Such efforts are likely to improve the quality of diagnostic imaging and other clinical practices to which they are applied.

APPENDIX

Exametazime should be prepared with only freshly eluted TcO_4^- and injected as soon as possible after preparation and completion of a quality control step that uses the short form of chromatography. A dose of 20 to 30 mCi should be injected in a low noise level, dimly lit room under conditions that keep visual, auditory and cognitive stimuli to a minimum for at least 10 min before and after radiopharmaceutical injection.

Multihead SPECT systems fitted with special-purpose collimation with identically peaked energy windows on all heads are preferred. Single-head camera systems should be used only if there is rigorous attention to quality control practices. Always position the patient and scanner to achieve a minimum radius of rotation. Prepare the patient for the procedure by indicating what to expect during the examination. Prevent any possible patient motion by ensuring that the patient is as comfortable as possible during the procedure and by using Velcro security bands on the torso and extremities and a comfortable head motion restraint device. Use sequential short (4–5 min) acquisitions in agitated patients. Choose a matrix resolution so that pixel size is about two to four times smaller than the expected camera resolution. Employ at least 120 views if the matrix resolution is 128 × 128 or 60 views if matrix is 64 × 64. Alternatively, collect projections at 6 degree or finer angular increments using a linear pixel sampling that provides for a 4- to 6-mm sampling in the reconstructed slice. Have neuroanatomic studies (CT or MRI) available for comparison and anatomic detail.

ACKNOWLEDGMENTS

The authors acknowledge the assistance of Sheryl Stern, MS, Associate Director, Commission on Health Care Policy, Society of Nuclear Medicine, for data collection and collation and the members of the expert panel, Robert Caretta, MD, Michael Devous, PhD, Michael Graham, MD, PhD, Robert Hellman, MD, Tom Hill, MD, Jack Juni, MD, Ishmael Mena, MD, Henry Wellman, MD, Alan Maurer, MD, Ron Ticofsky, PhD, Dave Weber, PhD, and Ronald VanHeertum, MD, who contributed their time and expertise to the development of this information.

REFERENCES

1. Field MJ, Lohr KN, eds. *Guidelines for clinical practice: from development to use. Institute of Medicine Committee on Clinical Practice Guidelines.* Washington, DC: National Academy Press; 1992.
2. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain

geographic variations in the use of health care services? A study of three procedures. *JAMA* 1987;258:2533–2537.
3. Woolf SH. Practice guidelines: a new reality in medicine. I. Recent developments. *Arch Intern Med* 1990;150:1811–1818.
4. Woolf SH. Practice guidelines: a new reality in medicine. II. Methods of developing guidelines. *Arch Intern Med* 1992;152:946–952.
5. Brain SPECT perfusion imaging: image acquisition, processing, display and interpretation. In: Springfield, VA: National Technical Information Service; 1991. Weber DA, Devous MD, Tikofsky RS, *Proceedings of workshop held at Brookhaven National Laboratory*, October 8–9, 1991.
6. Bramson R, Parlette N. Methods of data collection for decision making. *Personnel J* 1978;7:243–245.
7. Fink A, Kosecoff J, Chassin M, Brook R. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74:979–983.
8. Brook R, Chassin M, Fink A, Solomon D. *A method for the detailed assessment of the appropriateness of medical technologies. A Rand Note N-3376-HHS.* Santa Monica, CA: Rand Corp.; 1991.
9. Gordon TJ, Helmer O. *Report on a long-range forecasting study. RAND Paper P-2982.* Santa Monica, CA: RAND Corp.; 1964.
10. Eddy DM. Practice: where do they come from. *JAMA* 1990;263:1265–1275.
11. McClennan B. Radiology summit meeting 1992: the role of radiology in health care reform. *AJR Am J Roentgenol* 1993;160:649–657.
12. Woolf SH. Practice guidelines: a new reality in medicine. III. Impact on patient care. *Arch Intern Med* 1993;153:2646–2655.