ELSEVIER

Journal of Clinical Epidemiology 56 (2003) 1150–1156

Journal of Clinical Epidemiology

Delphi as a method to establish consensus for diagnostic criteria

Brent Graham*, Glenn Regehr, James G. Wright

Toronto Western Hospital, 399 Bathurst Street, FP-174 Toronto, ON M5T 2S8, Canada

Accepted 29 April 2003

Abstract

Background and Objectives: To achieve a consensus, among a panel of experts, on the best clinical criteria for the clinical diagnosis of carpal tunnel syndrome (CTS).

Method: Experts rated the diagnostic importance of items from the clinical history and physical examination for CTS. The ratings were expressed on a 10-cm visual analog scale. The average and standard deviation of the scores for each item were returned to the panelists. The panel members evaluated the items a second time with knowledge of the group responses from the first round. The scores were standardized to minimize scaling variations and, after the second round, the items were ranked in order of importance assigned by the group. Cronbach's α was used as a measure of homogeneity for the rankings. Increasing homogeneity was considered to be an indication of consensus among the panelists.

Results: Cronbach's α increased from 0.86 after the first round to 0.91 after the second iteration. Panelists who were relative outliers on the first round demonstrated a much higher correlation with the entire group after the second round.

Conclusions: Delphi is an effective method of establishing consensus for certain clinical questions. Cronbach's α was a useful statistic for measuring the extent of consensus among the panel members. Delphi was chosen from the possible methods of group process because of its inherent feasibility. The absence of a need by the panelists to meet in person removed any constraint on the geographic location of the panel members. In addition, the anonymous nature of Delphi was thought to be a key factor in avoiding a result that might be skewed by one or more persuasive panelists. Both of these characteristics were felt to be particularly important to the topic on which consensus was sought, the clinical diagnostic criteria for CTS. This movement in the opinions of some of the panelists appeared to result from the feedback of information describing the group opinion.

Keywords: Group process; Delphi; Consensus; Carpal tunnel syndrome; Cronbach's α; Diagnostic criteria

1. Introduction

Delphi is a well-recognized group process in the social sciences [1–8]. Although prior studies have used this method to establish appropriateness criteria for treatment [9–16], it has received less attention as a tool to establish consensus among health care professionals on diagnosis.

Delphi is a completely anonymous process in which the participants never meet. Delphi otherwise resembles the nominal group in structure. Ideas are expressed to the participants in the context of a mailed questionnaire. Responses to the items in the questionnaire are collated and analyzed. Items may be dropped or added in a second round in which the group responses to the first round are reported to the participants. New responses to the items are recorded and repeated iterations of the process carried out until a consensus appears to have been reached. The determination that a consensus has been achieved requires an operational definition that is appropriate to the issue under consideration.

Delphi is particularly attractive for the task of achieving consensus, especially among health care professionals. First, the absence of an obligation to meet in person greatly improves the feasibility of Delphi and lowers the cost significantly. Second, and perhaps more importantly, there are less likely to be constraints on either the size or composition of the group. Participants may be recruited from diverse geographic locations and clinical backgrounds. Third, the reliability of group consensus for the issue being examined improves as the number of panelists is increased [2]. A panel size appropriate to the issue under consideration is easier to achieve because of the inherent feasibility of the Delphi process. Finally, the anonymous nature of the exercise ensures that a single influential participant will not have a disproportionate impact on the outcome of the group as can occur with other group processes.

The Delphi process has been criticized as being subject to bias because the investigator limits the scope of the issue

* Corresponding author. Tel.: 416-603-5647; fax: 416-603-5716.
E-mail address: brent.graham@utoronto.ca (B. Graham).

0895-4356/03/$ – see front matter
doi: 10.1016/S0895-4356(03)00211-7

evaluated by the panelists. Thus, as the breadth of the issue under consideration is at least partially controlled by the investigator, any consensus that may emerge may be somewhat distorted [5]. The Delphi method has also been criticized for the fact that the panelists never meet together. Other group processes depend on the interaction between the participants as a source of novel insight into an issue. Due to the nature of Delphi, no discussion takes place, and any consensus that the group appears to have developed can only derive from information provided to it by the investigator. Where there is discussion among panelists, like in other types of group process, the consensus reached may be significantly different from that expected prior conducting the group process [4]. Finally, criteria for determining that group consensus has been achieved have not been established.

Carpal tunnel syndrome (CTS) is a diagnosis commonly made in industrialized societies. The prevalence of CTS reportedly varies with the geographic location of the population under study [17–19]. There also are variations in the reported prevalence of the condition between different industries [20–27]. Potential explanations for these variations may include intrinsic differences among the populations, different exposures, or variations in the diagnostic criteria for identifying CTS. One of the more likely explanations for these variations is the different case definition for this condition among reports. These vary in their nature, emphasis, and stringency.

The clinical evaluation remains an important aspect of the diagnostic process for CTS. Electrodiagnostic studies are often considered to be a gold standard diagnostic test for CTS [28–30]. In other words, electrodiagnostic tests are frequently taken to represent a demonstration of the essential lesion for the clinical condition called CTS. There is not consensus on this issue, and the assumption that these tests represent the essential lesion in CTS may be flawed for the following reasons. First, electrodiagnostic testing does not have perfect sensitivity or specificity, and the tests may be normal despite clinically significant nerve compression [31,32]. Second, the standard interpretation of electrodiagnostic data assumes a normal distribution of nerve conduction velocities, and arbitrarily designates velocities more than two standard deviations from the mean as abnormal. This results in a misclassification of asymptomatic individuals as being affected by CTS. In addition, the literature reports that assumption of a normal distribution of nerve conduction velocity may not be reasonable [33]. The third point is that cut point for defining an abnormality of nerve conduction velocity varies in the literature [30,34,35]. There is no established consensus on the electrical evidence for CTS. The specificity and sensitivity of sensory nerve conduction measurements is affected by the threshold defining CTS. This would not be expected of a gold standard test in the usual connotation for this term.

The use of electrodiagnostic tests is not universal among experts who treat CTS [36]. Like any diagnostic test, electrodiagnostic evaluations should be interpreted from within a clinical context. Ideally, they should be used in a bayesian manner to modify the pretest probability of CTS established clinically. Thus, there is a need to standardize the clinical criteria for the diagnosis of CTS.

Clinical experience at our center has indicated that CTS is a condition that is diagnosed and treated by a broad spectrum of specialist and primary care clinicians. These diagnosticians bring varying experiences to the task of diagnosing CTS, experiences that have been obtained within an intellectual framework or paradigm specific to a particular clinical specialty. These unique clinical experiences could form the basis for diagnostic criteria that are not uniform among diagnosticians from different training backgrounds. The absence of uniform diagnostic criteria makes the study of potentially important factors, like industrial exposures, difficult or even impossible. Thus, CTS is a common clinical condition diagnosed using widely varying clinical criteria.

Diagnostic criteria for any condition must be both valid and clinically sensible. Establishing a consensus among clinical experts on what the criteria should comprise does not ensure validity. As clinical experience evolves, the opinions of experts may also change, together with their diagnostic practices. The development of methods for seeking consensus must consider this and be flexible so that the criteria can be re-examined and revised at intervals. Obtaining agreement among clinical opinion makers should be seen as a starting point for establishing criteria that are likely to have significant clinical sensibility and that can be tested to evaluate validity. The key issue is the development of a consensus so that a gold standard diagnostic criterion can be established.

The objective of this study was to determine whether the Delphi method could be used to achieve consensus among influential expert clinicians representing all of the involved clinical disciplines for diagnostic criteria for CTS. An additional goal was to measure agreement within the panel using Cronbach's α as a measure of the internal consistency of the group.

2. Method

2.1. Selection of the panelists

Panelists were recruited from clinical disciplines involved in the diagnosis and treatment of CTS including neurology, neurosurgery, rheumatology, occupational health, plastic surgery, and orthopedic surgery. We attempted to identify experts defined in at least one of two ways. First, some panelists were leaders in their clinical fields as evidenced by their roles as opinion makers within national organizations such as the American Society for Surgery of the Hand and the American Society for Peripheral Nerve. Second, a MEDLINE search of the literature on CTS was carried out to identify individuals who had authored one or more papers directly concerned with CTS in a peer-reviewed medical journal during the preceding 3 years.

Although primary care physicians frequently evaluate CTS, none were included in the Delphi panel. The reason for this was that the overall objective was to eventually develop a diagnostic instrument for CTS that modeled the practices of expert specialist clinicians. This phase in the development of the diagnostic instrument was concerned with establishing the presence of consensus among the clinicians upon whose practices the scale would be based. Other aspects of the development of the diagnostic scale will be described in subsequent publications. The focus of this article was to evaluate a method measuring consensus with a particular group of clinicians, in this case specialist physicians and surgeons, participating in a group process, Delphi.

Prospective panelists were sent an information package containing a synopsis of the study goals and a description of Delphi. One individual who had authored a peer-reviewed publication refused participation because he was an epidemiologist and not a clinician. Fourteen panelists participated in the initial round and 12 completed the second round. The membership of the panel was not disclosed to the participants.

2.2. Delphi items

The Delphi panelists considered 57 items that represented findings on the history and physical examination of patients in whom CTS is diagnosed (Table 1). These items had been compiled from four sources: (1) focus groups involving specialist clinicians. These focus groups were conducted at the meetings of the academic organizations to which these individuals belonged; (2) key informant interviews of local faculty members with experience in assessing and treating CTS; (3) textbooks from the clinical specialties; and (4) the published literature on the topic of CTS. The investigators extracted the items from the literature and from the transcripts of the focus groups and key informant interviews. They did not participate in any capacity in either the key informant interviews or in the focus group discussions.

In the first Delphi round each member of the panel evaluated each of the 57 items for its importance to the clinical diagnosis of CTS. For each item, each panelist was asked to answer the question: "How important to you is [item] in diagnosing carpal tunnel syndrome." A 10-cm visual analog scale with the anchors "completely unimportant" and "extremely important" was used to record the responses. The mean score and standard deviation for the group as a whole was established for each individual item.

In the second round the panelists considered the same items, and this time were also informed what the group mean score and standard deviation was for each item in the first round. The panelists were asked to evaluate the items again in light of this information. The process was terminated after the second round because a consensus had been reached using the criteria described in the next section.

2.3. Analysis of Delphi data

For this study, the concept of consensus within a group was defined as a condition of homogeneity or consistency of

Table 1
Delphi panel items

Patient characteristics
Female gender
Middle age
Obesity
History
Relief with shaking hand
Relief by hanging hand over bedside
Relief with splinting
Relief following a steroid injection into carpal canal area
Symptoms improve during periods of inactivity
Onset with recent change in work activity
Symptoms
Nocturnal sensory symptoms
Numbness or tingling in median nerve distribution
Symptoms only on radial side of ring finger
Burning pain on radial side of palm
Hypersensitivity in digits
Sense of coldness in digits
Swelling of digits
Diaphoresis in hand
Pain in hand
Subjective weakness in hand
➡ dropping objects
➡ weak pinch
➡ clumsiness
Precipitation by driving
Precipitation by reading newspaper
Precipitation by typing at computer keypad
Precipitation by grasping telephone
Pain in wrist
Pain in fingers
Radiating pain in forearm but not proximal to elbow
Radiating pain into proximal limb, shoulder or neck
Precipitation by manual work activity
➡ specific occupations e.g. meat cutting
➡ vibratory activity
Bilateral symptoms
Coexisting diagnoses
Fibromyalgia
Inflammatory arthropathy
Diabetes
Previous fracture of distal radius
Osteoarthritis
Cervical disc disease
Hypothyroidism
Pregnancy
Physical signs
Tinel sign
Phalen sign
Local swelling
Thenar atrophy
Weakness of abductor pollicis brevis muscle
Thickening of flexor tendons
Negative findings
Absence of sensory symptoms
Failure to respond to splinting
Failure to respond to steroid injection
Sensory loss outside median nerve distribution
Pain with passive flexion/extension of wrist
Loss of two point discrimination
Hypersensitivity in digits
Decreased vibratory sensation
Tenderness to palpation over median nerve
Weakness with intrinsic positive posture

opinion among the panelists. Cronbach's α is one statistical index, among others, that has been used to quantify the reliability of a summation of entities, in this case, panelists [37]. Where the responses of the panelists are highly correlated, they are considered to be internally consistent or homogeneous.

Assuming that each clinical item was characterized by a constant, but unknown, importance as a diagnostic criterion, the opinions of the panelists, represented by their scores recorded on the visual analog scale, could be considered multiple measures of this characteristic. In this sense, each item could be thought of as being evaluated by the different criteria in a measurement scale, each of these criteria being represented by a panelist. Thus, the internal consistency of the panel for each item would be expected to reflect the extent of consensus within the group for the importance of that item.

Were the reliabilities of each rater for item importance known to be equal, then the Spearman-Brown formula could be used to estimate the reliability of the sum of the raters. However, Bravo and Potvin [38] indicate that the Spearman-Brown formula is not applicable where the reliability of each panelist may differ. The reliability of the sum of the panelists' responses, is more meaningful than the sum of reliabilities. Cronbach's α estimates the reliability of the sum of panelists responses:

$$\alpha_x = k/k-1\ (1-\Sigma\sigma^2_{yi}/\sigma^2_x)$$

where k is the number of panelists, $\sigma^2 yi$ are the variances of each individual panelist responses, and $\sigma^2 x$ is the variance of the sum of responses for each individual panelist.

The more the individual panelists covary, that is, the smaller the variance between panelists, compared with the variance within each panelist, the closer Cronbach's α will be to 1.0.

The interpretation of Cronbach's α is that the panelists chosen for the study are a sample from a population of content experts on the topic of CTS. Evaluation of the items by a new panel, chosen from this population, would have an expected correlation with the index sample equal to Cronbach's α [37]. This line of reasoning justifies the use of Cronbach's α as the summarizing measure of internal consistency for the assessment of all 57 items by the panel because it suggests that the responses reported by the index panel would be correlated with other samples from the same population. Where Cronbach's α is close to 1.0, it can be argued that there is consistency in the responses of the index panel, suggesting consensus, and that this agreement is likely to observed in other samples selected in the same way from the same population. The meaning of the value of Cronbach's α depends on the context. Bland and Altman [37] and others [39] suggest that a reliability coefficient like Cronbach's α should be above 0.90 for a diagnostic scale to be useful in clinical practice.

3. Results

Cronbach's α for the first round of the Delphi process was 0.86. The individual panelist-group correlation ranged between 0.23 and 0.73 (Table 2). This suggested that there were some panelists who were relative outliers. Two of the three panelists with the lowest correlation with the entire group were from the same clinical specialty, rheumatology. The three highest values were orthopedic hand surgeons.

The items were ranked in descending order according to the average score assigned by the entire panel. Items that were clearly closely related to each other had similar rankings. For example, the top-ranked item was "thenar atrophy," and the third most highly ranked item was "weakness of the abductor pollicis brevis muscle," both of which are physical examination findings suggesting motor denervation of the median nerve-innervated muscles in the hand. The items ranked second, fourth, and eighth all related to the distribution of the sensory complaint of numbness. These findings appeared to corroborate the concept that there was consistency in rating the items on the visual analog scale.

In the second round panelists were sent a slightly revised list of items. In a few instances items were reworded because they were noted by the panelists to be ambiguous in their meaning on the first iteration. Two items were added at the suggestion of several of the panelists.

Only 12 of the panelists responded to the second round. Two of the three neurologists from the first round did not respond for unknown reasons. The same analysis was used to evaluate the data from the 12 panelists. Cronbach's α increased from 0.86 in the first round to 0.91 in the second. The panelist-group correlation substantially increased in 11 of 12 instances (see Table 2). These values ranged from 0.58 to 0.75, corresponding to the higher Cronbach's α observed

Table 2
Individual panelist-group correlation over two Delphi iterations

Panelist	Iteration 1	Iteration 2
Neurol 1	0.43	[a]
Neurol 2	0.68	0.74
Neurol 3	0.57	[a]
Neurosurg 1	0.62	0.74
Neurosurg 2	0.51	0.61
Occ Med 1	0.58	0.76
Occ Med 2	0.38	0.59
Ortho 1	0.62	0.79
Ortho 2	0.73	0.81
Ortho 3	0.73	0.66
Plast 1	0.23	0.63
Plast 2	0.62	0.74
Rheum 1	0.32	0.75
Rheum 2	0.36	0.64
Cronbach's α	0.86	0.91

[a]Did not respond to second iteration.

Abbreviations: Neurol, Neurologist; Neurosurg, Neurosurgeon; Occ Med, Occupational Medicine Specialist; Ortho, Orthopedic surgeon; Plast, Plastic surgeon; Rheum, Rheumatologist.

for the entire group. The disparity in correlation with the group among members of particular specialties also was markedly attenuated.

The ranking of items in the second iteration was not qualitatively different from those obtained in the first round. Fourteen of the 15 most highly ranked items in the first round appeared among the top 15 items in the second iteration, although the order of the rankings was slightly different. Among the 20 top items identified in the first iteration, 17 were in the top 20 ranked items after the second round (Table 3).

4. Discussion

The concept of consensus within a group is easily understood, but the best way to measure this phenomenon is unclear. Furthermore, criteria that indicate a consensus has been achieved will vary with the setting in which agreement is sought and the method being utilized.

Consensus within the group in general should be reflected in decreases in the variance of the responses. Consensus among groups has often been quanitated using group means and standard deviations [16]. In our study, the standard deviation for the group decreased in the second round, but the difference from the first round was not statistically significant.

Where the true answer to a question is known, experiments in group opinion have suggested that the variance and standard deviation of panelists' answers are not as accurate as some measure of group confidence [3] in gageing whether a consensus has been reached by the panel. The extent of group confidence can be evaluated by having each panelist express how confident they are in the opinion they have stated. Decreases in the variance of this response are interpreted as showing increasing group confidence in the opinion established by the process. This approach may be most useful where panelists are not necessarily experts on the topic under discussion [3]. In contrast, studies involving experts on the issue under discussion have not supported the group confidence concept as an indication of consensus. For example, in a study of the management of blunt hepatic trauma, involving 14 expert general surgeons, Milholland et al. [16] found that ratings of confidence in the answers given by the panelists were almost uniformly high. The authors concluded that this approach could not be used with groups comprising experts because, by their nature, experts will usually have a high degree of confidence in their opinion. As a result, the variance of any measure of confidence for the group will likely be small, even where no consensus on the issue has been achieved.

In the present study, consensus was equated to a homogeneity of opinion expressed by a group of individuals rating items on a visual analog scale. Agreement defined in this way can be evaluated using Cronbach's α. Cronbach's α can be directly interpreted as the correlation expected between the present panel and a second panel selected from the same population of content experts [37]. The question of the *extent* of homogeneity of opinion required to state with confidence that a consensus exists was set high because these criteria will ultimately be used at the individual patient level. The value of 0.91 observed in our study after the second Delphi round should be considered to be very substantial, and is consistent with reliability values obtained for validated scales in clinical use [37,39].

Even during the first Delphi round Cronbach's α was 0.86, indicating substantial homogeneity among the panelists from the beginning. This might be somewhat surprising given the widely divergent criteria in the literature for the diagnosis of CTS. Variations in the peer-reviewed literature may be due to poor study methodology, misleading or naïve statistical analyses, failure to publish negative results, and many other sources of systematic error leading to inconsistency. In this study, the reasonably high consistency among experts from diverse clinical specialties, even in the first round, is somewhat reassuring, and suggests that clinical judgment and consensus, provided that these are based on experts, is a reasonable approach to establishing diagnostic criteria.

Table 3
Clinical features of CTS, ranked by importance on Delphi iteration 1 and 2

Rank	Iteration 1	Iteration2
1	Thenar atrophy	Sensory symptoms restricted to median nerve
2	Absence of sensory symptoms	Thenar atrophy
3	Weakness of APB	Sensory symptoms only outside median nerve
4	Sensory symptoms restricted to median nerve	Weakness of APB
5	Nocturnal sensory symptoms	Absence of sensory symptoms
6	Coexisting condition: pregnancy	Nocturnal sensory symptoms
7	Coexisting condition: diabetes mellitus	Sensory symptoms described as numbness or tingling
8	Sensory symptoms only outside median nerve	Coexisting condition: pregnancy
9	Sensory symptoms described as numbness or tingling	Phalen sign
10	Phalen sign	Loss of two point discrimination
11	Precipitation by: driving	Tinel sign
12	Loss of two point discrimination	Coexisting condition: previous distal radius fracture
13	Coexisting condition: Hypothyroidism	Precipitation by: Driving
14	Decreased vibratory sensation	Decreased vibratory sensation
15	Response to: splinting	Response to: splinting
16	Tinel sign	Response to: steroid injection
17	Coexisting condition: previous distal radius fracture	Precipiptation by: grasping the newspaper
18	Exposure to vibration	Exposure to vibration
19	Coexisting condition: inflammatory arthropathy	Coexisting condition: Hypothyroidism
20	Dropping objects	Precipitation by: grasping the phone

Delphi was useful not only for evaluating consistency among experts but also for improving agreement. Cronbach's α improved from 0.86 in the first iteration to 0.91 after the second round. There was particular movement among the members of the panel who had the lowest correlation with the rest of the group after the first round. In fact, there was general improvement in the correlation with the whole group observed in almost all of the panelists. This finding suggests that feedback to the group of the results in the first round influenced the responses of the panelists in the second round, resulting in a greater degree of agreement. The fact that the participants, who are acknowledged to be opinion leaders in this area, appeared to have been influenced by the group as a whole, indicates that Delphi can be effective even in the context of a well-informed, and possibly even opinionated, group of panelists.

The loss of two panelists for the second iteration could have influenced the outcome of the study. However, Cronbach's α is sensitive to the number of items, or in this case, panelists. The loss of two panelists would be most likely to decrease the value of Cronbach's α. In fact, Cronbach's α increased in the second round, suggesting that the increase in consistency within the remaining group more than offset the effect of losing two participants.

The two panelists who did not participate in the second round were both neurologists. Neurologists play a prominent role in the diagnosis of CTS, and the effect of the absence of these two individuals on the achievement of consensus is not clear. It's possible that the qualitative nature of the consensus that was established might have been different with the full participation of these individuals; however, this does not seem likely. Although it is impossible to predict what the response of the two neurologists who did not complete the study might have been in the second round, it's unlikely that they would have differed significantly from the rest of the group. Eleven of the 12 remaining panelists showed an increase in the magnitude of their individual correlation with the group. In addition, the responses of the remaining neurologist who participated in both rounds of the study were highly correlated with the rest of the panel, suggesting that this individual contributed substantively to the consensus.

The objective of the study was not to compare Delphi to any other method of establishing clinical agreement but rather to determine whether this method could be used for this purpose. Owing to its inherent feasibility, certain questions requiring agreement might be particularly well suited to the advantages of the Delphi method. For example, Delphi could be used to establish diagnostic criteria for other conditions on which consensus is lacking. Treatment protocols and other "best practice" issues where agreement is desirable might also benefit from a group process approach using either Delphi or some other technique. The study of many of these issues using traditional clinical research methods like randomized trials may simply be infeasible. There seems to be clear value in pooling the experiences of experts in a field to establish a consensus on a controversial topic at least as a starting point toward collecting actual evidence on a clinical question. Group processes such as Delphi may be a cost-effective approach to some of these problems. Questions of preferences for resource allocation have already been studied using Delphi [40].

This study has several limitations. Although Cronbach's α allowed quantification of group consensus and comparison of the level of consensus to generally accepted standards, no direct comparison was made between Cronbach's α and any other measures of consensus. It is unknown how useful Cronbach's α would be in other situations.

Our study attempted to minimize the risk of bias associated with the Delphi methods in several ways. First, the items for inclusion in the study were not chosen by the investigators but resulted from a formal process of item generation that included literature review, focus group discussions, and key informant interviews. The investigators did not participate in the identification of the items for study, beyond extracting them from the source material. The participants in the focus groups and key informant interviews were from all of the clinical specialties commonly involved in the care of CTS. These individuals were also from a wide range of geographic locations, mostly in North America, but in a few instances from Europe as well. Both of these factors limited the risk of generating a pool of items for consideration by the Delphi panel that was based on a limited clinical point of view.

The absence of face-to-face encounters by the panelists that characterize Delphi was felt to be an advantage for the goal of this project—the establishment of consensus on the controversial issue of diagnostic criteria for CTS. Although the potential for productive interaction was lost, the benefits gained in the lack of constraint on the location and clinical background of the participants outweighed this disadvantage. The broadly based process of item generation also seems likely to have limited the risk of neglecting important issues.

No attempt has been made to compare Delphi with any other method of obtaining consensus on this issue. For the reasons listed above Delphi was considered best suited, among a number of potentially useful techniques, to meet the objectives of this particular question. It is unknown whether or not it would be as informative in another context. Nonetheless, our results indicate that under the appropriate circumstances, Delphi is a useful consensus method.

In conclusion, Delphi was useful for the goal of establishing consensus on diagnostic criteria in CTS. The concept of consensus, reframed as one of consistency witnin the group, allowed the use of Cronbach's α. This statistic detected movement of the group opinion where other measures of consensus, like the group standard deviation, were uninformative. The next step in this project will be to establish weights for the most highly ranked clinical items and to create a diagnostic scale for clinical use. Once weightings have been determined for the items included in the final

instrument, validity testing will be carried out. The long-term objective is to create a scale that is a model of expert clinician practice. Among the potential uses of such an instrument would be the possibility of screening patients for further assessment with electrodiagnostic tests. Cases in which the scale indicates that the likelihood of CTS is either very high or very low may not require electrodiagnostic evaluation. This has implications for sparing patients the expense, inconvenience, and discomfort of testing that might not materially contribute to their care.

Acknowledgments

Dr. Wright is supported as a the R.B. Salter Chair of Surgical Research and as an Investigator of the Canadian Institute for Health Research. This work was supported by Physicians' Services Incorporated Foundation Grant 97-52.

References

[1] Dalkey NC, Helmer O. An experimental application of the Delphi method to the use of experts. Manage Sci 1963;9:458–67.
[2] Dalkey NC. The Delphi method: an experimental study of group opinion. Santa Monica, CA: RAND Corporation; 1969.
[3] Dalkey N, Brown B, Cochran S. The Delphi Method III. Use of self-ratings to improve group estimates. Santa Monica, CA: RAND Corporation; 1969.
[4] Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984;74(9): 979–83.
[5] Pill J. The Delphi method: substance, context, a critique and an annotated bibliography. Socio-Econ Plan Sci 1971;5:57–71.
[6] Romm FJ, Hulka BS. Developing criteria for quality of assessment: effect of the Delphi technique. Health Serv Res 1979;14(4):309–12.
[7] Weinberg AD, Ullian L. The Delphi technique as a method for determining the continuing education needs of physicians regarding coronary artery disease. J Assoc Hosp Med Educ 1977;Second Quarter:32–6.
[8] Zinn J, Zalokowski A. The use of the Delphi panel for consensus development on indicators of laboratory performance. Clin Lab Manage Rev 1999;13:386–408.
[9] Couper MR. The Delphi technique: characteristics and sequence model. ANS Adv Nurs Sci 1984;7(1):72–7.
[10] Gale RP, Park RE, Dubois R, et al. Delphi-panel analysis of appropriateness of high-dose chemotherapy and blood cell or bone marrow autotransplants in diffuse large-cell lymphoma [see comments] [published erratum appears in Leuk Lymphoma 1999 Aug;34(5–6):following 639]. Leuk Lymphoma 1998;32(1–2):139–49.
[11] Hillman BJ, Hessel SJ, Swensson RG, Herman PG. Improving diagnostic accuracy: a comparison of interactive and Delphi consultations. Invest Radiol 1977;12(2):112–5.
[12] Jairath N, Weinstein J. The Delphi methodology (part two): a useful administrative approach. Can J Nurs Admin 1994;7(4):7–20.
[13] Koplan JP, Farer LS. Choice of preventive treatment for isoniazid-resistant tuberculous infection. Use of decision analysis and the Delphi technique. JAMA 1980;244(24):2736–40.
[14] Matthews ME, Mahaffey MJ, Lerner RN, Bunch WL. Profiles of the future for administrative dietitians via the Delphi technique. J Am Diet Assoc 1975;66(5):494–9.
[15] McKenna HP. The Delphi technique: a worthwhile research approach for nursing? J Adv Nurs 1994;19(6):1221–5.
[16] Milholland AV, Wheeler SG, Heieck JJ. Medical assessment by a Delphi group opinion technic. N Engl J Med 1973;288(24):1272–5.
[17] Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosen I. Prevalence of carpal tunnel syndrome in a general population [see comments]. JAMA 1999;282(2):153–8.
[18] de Krom MCTF, Knipschild PG, Thijs CT, Boekkooi PF, Spaans F. Carpal tunnel syndrome: prevalence in the general population. J Clin Epidemiol 1992;45(4):373–6.
[19] Dieck G, Kelsey JL. An epidemiologic study of the carpal tunnel syndrome in an adult female population. Prev Med 1985;14:63–9.
[20] Barnhart S, Rosenstock L. Carpal tunnel syndrome in grocery checkers: a cluster of a work related illness. West J Med 1987;147(1):37–40.
[21] Bleecker ML. Medical surveillance for carpal tunnel syndrome in workers. J Hand Surg [Am] 1987;12(5 Pt 2):845–8.
[22] Cannon LJ, Bernacki EJ. Personal and occupational factors associated with carpal tunnel syndrome. J Occup Med 1981;23(4):255–8.
[23] Cummings K, Maizlish M, Rudolph L, Dervin K, Ervin A. Occupational disease surveillance: carpal tunnel syndrome. Morbidity and Mortality Weekly Report 1989;38(28):485–9.
[24] Franklin GM, Haug J, Heyer N, Checkoway H, Peck N. Occupational carpal tunnel syndrome in Washington state, 1984–1988. Am J Public Health 1991;81:741–6.
[25] Margolis W, Kraus JF. The prevalence of carpal tunnel syndrome symptoms in female supermarket checkers. J Occup Med 1987;29(12): 953–6.
[26] Osorio AM, Ames RG, Jones J, Castorina J, Rempel D, Estrin W, Thompson D. Carpal tunnel syndrome among grocery store workers. Am J Ind Med 1994;25:229–45.
[27] Silverstein BA, Fine LJ, Armstrong TJ. Occupational factors and carpal tunnel syndrome. Am J Ind Med 1987;11:343–58.
[28] Boniface SJ, Morris I, MacLeod A. How does neurophysiologic assessment influence the management and outcome of patients with carpal tunnel syndrome? Br J Rheum 1994;33:1169–70.
[29] Golding DN, Rose DM, Selvarajah K. Clinical tests for carpal tunnel syndrome: an evaluation. Br J Rheum 1986;25:388–90.
[30] Nathan PA, Keniston RC, Meadows KD, Lockwood RS. Predictive value of nerve conduction measurements at the carpal tunnel. Muscle Nerve 1993;16:1377–82.
[31] Braun RM, Jackson WJ. Electrical studies as a prognostic factor in the surgical treatment of carpal tunnel syndrome [see comments]. J Hand Surg [Am] 1994;19(6):893–900.
[32] Grundberg A. Carpal tunnel decompression in spite of normal electromyography. J Hand Surg 1983;8(3):348–9.
[33] Robinson LR. A statistical problem in diagnosis of carpal tunnel syndrome. Muscle Nerve 1994;17:1490–1.
[34] Jackson DA, Clifford JC. Electrodiagnosis of mild carpal tunnel syndrome. Arch Phys Med Rehabil 1989;70:199–204.
[35] Stevens J. AAEE minimonograph #26: the electrodiagnosis of carpal tunnel syndrome. Muscle Nerve 1987;10:99–113.
[36] Duncan KH, Lewis RC, Foreman KA, Nordyke ME. Treatment of carpal tunnel syndrome by members of the American Society for Surgery of the Hand: results of a questionnaire. J Hand Surg 1987;12A(3): 384–91.
[37] Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ 1997; 314:572–3.
[38] Bravo G, Potvin L. Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: toward the integration of two traditions. J Clin Epidemiol 1991;44(4–5):381–90.
[39] Shrout P, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420–8.
[40] Moscovice I, Armstrong P, Shortell S, Bennett R. Health services research for decision-makers: the use of the Delphi technique to determine health priorities. J Health Polit Policy Law 1977;2(3):388–410.