

COMMENTARY

Introductory note:

As the JOURNAL embarks on a policy encouraging papers of medical and scientific value from disciplines sometimes unfamiliar to us, such as the social sciences, readers may find that some of the methods used to collect or analyze data are also unfamiliar. In this issue, we present an important report concerning evaluation of techniques for surgical reconstruction (Page 13). A consensus method was used to reach the conclusions in this report, based on responses from several highly experienced individuals. Dr. Hugh Cross has kindly consented to provide the following background information regarding the consensus method used in this study. Ed.

Consensus Methods: A Bridge Between Clinical Reasoning and Clinical Research?

ABSTRACT

Evidence-Based Practice does head the “hierarchy of evidence” upon which developments in clinical practice should be based. There are, however, situations where evidence is either unavailable, unclear, or results between studies are at variance. Consensus is a reliable contingency, and approaches to reaching consensus have acceptable construct validity (Nominal Group Technique, Delphi, and Consensus Development Conference).

Consensus is reached when: (i) the method of investigation tightly controls communication to reduce the obscuring “noise” of divergent discussion; (ii) statistical measures of agreement or dissent screen out the bias that would otherwise be produced by the dictate of vociferous minorities or coalitions that may represent vested interests; (iii) all participants contribute equally to the product of the investigation.

RÉSUMÉ

Le concept de la pratique médicale basée sur des données établies (dénommée « Evidence-Based Practice ») permet réellement d'établir une « hiérarchie des preuves », à partir de laquelle des développements utiles pour la pratique médicale cliniques peuvent être déployés. Il y a cependant des situations où les données cliniques ne permettent pas de clairement étayer une hypothèse médicale ou soutenir une observation; ou bien les résultats observés d'une étude à une autre présentent une variation importante. Le Consensus est alors une méthode robuste dans de tels cas, et la plupart des approches pour atteindre un consensus, telle que par exemple la Technique du Groupe Désigné, la Méthode de Delphi et la Conférence de Développement du Consensus, présente une démarche bien construite et de validité acceptable.

Un consensus est atteint lorsque: (i) la méthode d'investigation contrôle efficacement la communication, afin de réduire le bruit de fond parasite et inutile des discussions divergentes; (ii) des mesures statistiques d'accord ou de désaccord filtrent les biais qui pourraient être produits par le dictat de minorités véhémentes ou bien de coalitions qui pourraient avoir des intérêts cachés; et (iii) tous les participants contribuent de façon équilibrée au produit de l'investigation.

RESUMEN

La Práctica Basada en la Evidencia reconoce la “jerarquía de la evidencia” como la base sobre la cual deben apoyarse los avances en la práctica clínica. Hay, sin embargo, situaciones donde la evidencia no es accesible, es poco clara, o hay variación en los resultados obtenidos. El consenso es una contingencia confiable y los intentos de alcanzar el consenso tienen una aceptable validez constructiva (Nominal Group Technique, Delphi, and Consensus Development Conference).

El consenso se alcanza cuando: (i) el método de investigación controla estrechamente la comunicación para reducir el ruido de la discusión divergente, (ii) las mediciones estadísticas de concordancia o de no concordancia toman en cuenta todas las tendencias o inclinaciones ya que de otra manera las opiniones de minorías o coaliciones vociferantes podrían ser dominantes, (iii) todos los participantes contribuyen igualmente al producto de la investigación.

Consensus is reached when: (i) the method of investigation tightly controls communication to reduce the obscuring “noise” of divergent discussion; (ii) statistical measures of agreement or dissent screen out the bias that would otherwise be produced by the dictate of vociferous minorities or coalitions that may represent vested interests; (iii) all participants contribute equally to the product of the investigation.

With the publication of the Report of the International Leprosy Association Technical Forum ⁽¹⁾ it has become broadly accepted that, wherever possible, further decisions on any proposed developments in clinical practice should be evidence-based. Proponents of evidence-based practice (EBP) suggest that decisions based on the empirical paradigm of science are less likely to be compromised by the unpredictable elements of subjectivity that are probably inextricable from clinical reasoning.

There are, however, circumstances where EBP does not provide answers for those who face problems of decision-making. For situations where there is already a plethora of confusing information, statistical methods such as meta-analysis are now in common use. Where published information is inadequate, non-existent, or contradictory, however, consensus methods provide a means of synthesizing the insights of experts to create a product that decision-makers can use with relative confidence.

Consensus as a valid construct has been supported through exercises in the fields of social science with the result that three objective methods of consensus building and reporting are now in common use: The Delphi Investigation, The Nominal Group Technique (NGT), and Consensus Development Conference. Each shares the common objective of synthesizing judgments when a state of uncertainty exists, but whereas Delphi and NGT are appropriate for smaller scale investigations, the Consensus Development Conference was designed to resolve conflicting opinions and contentious issues that impact on health policy at national or international levels. (Henceforth, this communication will only consider Delphi and Nominal Group Technique as these lie within the experience of the author.)

The central question of consensus reliability was investigated in early studies by Delbecq and Van de Ven ⁽²⁾ who ascertained that

judgmental accuracy may be achieved where the following features are encapsulated in the method of investigation: (i) individuals make independent judgments; (ii) individual judgments are expressed through mathematical rank-ordering and/or rating of items; the mean value of independent judgments are accepted as indicating group decision; re-voting should follow discussion of the mean values.

More recent studies have shown that an indication of the distribution or dispersal of participants’ judgments, not just the measure of central tendency, is more appropriate. In general, the median and the interquartile range are more robust than the mean and standard deviation. Further thematic content analysis of comments and discussion can also enhance the quality of outcomes. ^(2, 3, 4, 5)

The Nominal Group Technique. An essential feature that characterizes a “group” is verbal communication. The reason that the term “Nominal” was adopted is that it denotes group situations in which non-verbal communication is permissible (the group is therefore, by definition only nominally a group). Early researchers applied the term rigidly, and no verbal communication was permitted. However, most contemporary NGT investigations are essentially a development of the approach as both verbal and non-verbal stages are incorporated. Research has shown that by allowing the combination of verbal and non-verbal stages, the optimal benefit from a NGT investigation can be achieved.

Delbecq, *et al.* ⁽⁶⁾ approached the issue of consensus development from psycho-social studies of decision making processes. Their first considerations were the various effects of normative behavior on individuals in groups and on a group as an entity. They also considered studies of alternative processes on the performance of group decision making in terms of the quantity and quality of ideas generated; the affectational (emotional and expressive) overtones of interaction; and the nature of facilitative and inhibitive influences on creative problem solving.

An objective of the NGT is that normative behavior (which basically favors the performance of dominant or aggressive characters) will be controlled by nonconformance tactics so that performance and outcome are maximized, while hidden agendas and negative group dynamics are sup-

pressed. The NGT aims to draw out minority opinions and promote the tolerance of conflicting ideas.

Effective creative, or judgmental, problem solving passes through two essential phases simply defined as the "fact finding phase" and the "evaluation phase." NGT includes processes that encourage deep consideration of problems and the enhancement of idea generation (fact finding), the clarification and synthesis of ideas (evaluation), and extends to all participants, an equal opportunity to contribute to the group product and to influence the direction of the decision outcome.

"The fact finding phase." Van de Ven and Delbecq (7) had reported the human tendency to seek solutions before a question or problem has been adequately grasped (an effect exacerbated by relative degrees of heightened anxiety over the nature of the topic or the perceived situational threat. This tendency leads to poor quality decisions). They also observed that where verbal communication is the method of idea generation, there is the possibility of "focus effect." "Focus effect" denotes a situation where group members are distracted and a single train of thought may be given inappropriate status. As a consequence, time is monopolized without the compensation of enhanced productivity. Van de Ven and Delbecq (7) also found that where group members are denied the opportunity for private reflection on independent thoughts, ideas were expressed as generalizations leaving individuals reluctant to be specific. The NGT method was developed to address these confounding effects as well.

The "fact finding phase" of the NGT is essentially a process where data (in this instance "ideas") are generated in silence and participants are required to write ideas in privacy. Writing forces participants to think through problems and creates a greater sense of task commitment and performance than verbal expression. The "fact finding phase" of the NGT, characterized by the demand for silent reflection and consideration was found to produce a wider range of better quality ideas than interactive verbal methods, e.g., brainstorming.

"The evaluation phase." The process of idea evaluation requires a different approach. NGT was further developed on the basis of investigations that suggested that,

following the generation of ideas, the synthesis of ideas is enhanced by verbal interaction (8). This second phase in the NGT allows verbal interaction where clarification of submitted responses may be required. A limited defense or criticism of ideas may also be permitted, but digressions and prolonged argument are not permitted.

Equality of participation. Selection bias and the definition of expertise are the most commonly cited flaws in consensus investigations generally (2, 9, 10, 11). The choice of participants is a salient consideration. Commitment to the process requires an internal acknowledgment of participant homogeneity and an external recognition of the expertise represented in the group. A tenet of The Nominal Group technique is that idea quality and not presenter status is predominant.

The Delphi investigation. Since its conception in the 1950s when it was used by the Rand Corporation for use in defense related problems (12), Delphi has been applied extensively to clarify issues that have required sharper definition. The method is usually adjusted to suit the requirements of individual applications (it has been widely applied among health disciplines for investigations as diverse as the determination of diagnoses, through policy development to ascertaining criteria for professional competence). Essentially the Delphi method is supported on the same theoretical basis as the NGT but the interaction, controlled by a central facilitator, is conducted by mail. It has the advantage of including participants who are separated geographically. The isolated and wholly anonymous situations in which participants process and respond to information, without the pressure of immediate response, does result in a broader array of high quality ideas. However the positive affects of the group interaction component in the NGT are also lost. It is for this reason that some consider the Delphi method to be inferior to NGT.

Delbecq, *et al.* (6) were instrumental in the early developments of the Delphi Method. They suggested that for a Delphi investigation, sample size should be dictated by the homogeneity of the group and the nature of the investigation. A large sample is necessary if the principal reason for conducting the procedure is to develop awareness within a group, or where diverse reference groups are involved. Where the

desired outcome is to validate opinions based on experience, they suggested that a group of ten to fifteen participants is adequate for a homogenous group. Increasing the size of a homogenous group beyond 30 will not result in more information and only increases administration difficulties.

Methodological details vary according to the requirements of individual Delphi projects. The general approach has been that sequential, structured questionnaires have been used for participants to rank, or rate, responses to indicate his/her priorities related to the topics of interest. On receipt of returned questionnaires, information is collated and analyzed before being redistributed for further refinement and for final comments.

Information, in the form of a statistical analysis, should be dispatched to participants at subsequent rounds of the procedure. Feedback analysis should include the frequency with which participants selected answers, with the mean and/or median and one measure of dispersion. Individuals are asked to reconsider the scores applied previously (in the light of aggregated responses of all members) and, in this manner, consensus is generated.

As with NGT, the selection and definition of experts is cited as being the most potentially confounding effect on a Delphi outcome. Panelists are usually (though not exclusively) recruited by merit of an intimate academic or experiential association with the topic under investigation. Acknowledged expertise or influence may validate a choice of participants; however, Delbecq, *et al.* (6) suggested that such attributes per se are insufficient for the inclusion of participants. They cautioned that commitment to the investigation, motivation to comply with the demands of procedure and the acceptance of the consensus (even though it may be at variance with personal inclination) are fundamental. The "nature of the respondent panel, the obligations of participants, the length of time the Delphi process will take and the information that will be shared among participants" are variables likely to effect the co-operation of invited individuals and should be declared at the initial stages of recruitment.

Criticisms. Some have considered consensus methods from an epistemological perspective and cautioned against "overselling" the methods (2). A principal con-

cern is that there is a risk that observers may place too great a reliance on consensus outcomes than may be warranted. This is a valid concern because consensus methods are used to generate quantitative estimates which could be misconstrued in some cultures as representing a "correct" answer. Consensus, is of course not synonymous with being invariably correct, but a responsibility lies with investigators to present outcomes with due consideration for the target readership. Sackman (9) represents the views of some who argue that Delphi outcomes represent a "forced" consensus that is further compromised because participants are not allowed to discuss issues.

Murphy, *et al.* (3) conducted an extensive review of published research using consensus methods. A result of their endeavor is that a guide has been published that should be considered by those considering the use of either method. One of the objectives of their survey was "To identify the factors that affect the decisions that emerge from consensus development methods." Their study identified recurring methodological issues which they sought to isolate and address. What their study has shown is that the methodological issues that have caused the most controversy can be addressed.

—Hugh Cross

*American Leprosy Missions
Cebu City, The Philippines*

REFERENCES

1. ANONYMOUS. Report of the International Leprosy Association Technical Forum. *Lepr. Rev.* **73(Suppl)** (2002) S3–S61.
2. DELBECQ, A. L., VAN DE VEN, A. H., and GUSTAFSON, D. H. *Group Techniques For Programme Planning: A Guide To Nominal Group and Delphi Processes.* Illinois: Scott Foresman & Company, 1975.
3. DOWNING, A. M., and HUNTER, D. G. Validating clinical reasoning: a question of perspective, but whose perspective? *Man. Ther.* **8(2)** (2003) 117–119.
4. HELMER, O., and RESCHER, N. On the epistemology of the inexact science. *Management Science* **6(1)** (1959) 25–52.
5. MOSTYN, B. The content analysis of qualitative research data: a dynamic approach. In: *The Research Interview*, 1985.
6. MURPHY, M. K., BLACK, N. A., LAMPING, D. L., MCKEE, C. M., SANDERSON, C. F., ASKHAM, J., and MARTEAU, T. Consensus development methods, and their use in clinical guideline development. *Health Technol. Assess.* **2(3)** (1998) 1–88.

7. SACKMAN, H. *Delphi Critique*. Lexington: Lexington Books, 1975.
8. SHARKEY, S. B., and SHARPLES, A. Y. An approach to consensus building using the Delphi technique: developing a learning resource in mental health. *Nurse Educ. Today* **21(5)** (2001) 398–408.
9. TONG, R. The epistemology and ethics of consensus: uses and misuses of ‘ethical’ expertise. *J. Med. Philos.* **16(4)** (1991) 409–426.
10. VAN DE VEN, A. H., and DELBECQ, A. L. The nominal group technique as a research instrument for exploratory health studies. *Exploratory Health Studies*. March (1972) 337–342.
11. VROOM, V. H., GRANT, L. D., and COTTON, T. J. The consequences of social interaction in group problem solving. *J. Appl. Psychol.* **53(4)** (1969) 338–341.
12. WORTMAN, P. M., VINOKUR, A., and SECHREST, L. J. Do consensus conferences work? A process evaluation of the NIH Consensus. *Health Polit. Policy Law.* **13(3)** (1988) 469–498.