Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies

Ivan R. Diamond[a,b,c], Robert C. Grant[a], Brian M. Feldman[a,c,d,e], Paul B. Pencharz[d,f], Simon C. Ling[d], Aideen M. Moore[a,c,d], Paul W. Wales[a,c,g,*]

[a]*Child Health Evaluative Sciences Program, The Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada*
[b]*Department of Medical Imaging, The University of Toronto, 263 McCaul Street, 4th Floor, Toronto, ON M5T 1W7, Canada*
[c]*Institute of Health Policy Management and Evaluation, The University of Toronto, 155 College Street, 4th floor, Toronto, ON M5T 3M7, Canada*
[d]*Department of Paediatrics, The Hospital for Sick Children and The University of Toronto, 555 University Avenue, Toronto, ON M5G 1X8, Canada*
[e]*Dalla Lana School of Public Health, The University of Toronto, 155 College Street, 6th floor, Toronto, ON M5T 3M7, Canada*
[f]*Department of Nutritional Sciences, The University of Toronto, 150 College Street, Room 316, Toronto, ON M5S 3E2, Canada*
[g]*Department of Surgery, The University of Toronto, 149 College Street, 5th Floor, Toronto, ON M5T 1P5, Canada*

Accepted 8 December 2013

Abstract

Objective: To investigate how consensus is operationalized in Delphi studies and to explore the role of consensus in determining the results of these studies.

Study Design and Settings: Systematic review of a random sample of 100 English language Delphi studies, from two large multidisciplinary databases [ISI Web of Science (Thompson Reuters, New York, NY) and Scopus (Elsevier, Amsterdam, NL)], published between 2000 and 2009.

Results: About 98 of the Delphi studies purported to assess consensus, although a definition for consensus was only provided in 72 of the studies (64 a priori). The most common definition for consensus was percent agreement (25 studies), with 75% being the median threshold to define consensus. Although the authors concluded in 86 of the studies that consensus was achieved, consensus was only specified a priori (with a threshold value) in 42 of these studies. Achievement of consensus was related to the decision to stop the Delphi study in only 23 studies, with 70 studies terminating after a specified number of rounds.

Conclusion: Although consensus generally is felt to be of primary importance to the Delphi process, definitions of consensus vary widely and are poorly reported. Improved criteria for reporting of methods of Delphi studies are required. © 2014 Elsevier Inc. All rights reserved.

Keywords: Delphi study; Consensus; Systematic review; Publication quality; Reporting guidelines; Multi-disciplinary

1. Introduction

A Delphi study is a widely used method to obtain input from a group of experts [1–3]. Although initially developed to predict cold war enemy attack probabilities, the technique has been widely used across numerous disciplines as a method to seek expert opinion in an iterative structured manner. The key features of the method are anonymity between participants with controlled feedback provided in a structured manner. Participants then may adjust their initial ratings based on feedback from the group in a number of subsequent iterations [1,4]. Although classically the method seeks to elicit consensus on the topic under study, there are variations of the Delphi such as the dissensus or Policy Delphi that aim to seek a broad range of opinions without achieving consensus [5].

Although the notion of consensus is fundamental to many Delphi studies, the definition of what constitutes consensus is less clear [3,4,6]. Investigators have adopted different approaches to defining consensus including formal measures of agreement, degree of uncertainty around a point estimate, decreases in variance of group responses, or the proportion of participants agreeing to a particular viewpoint [1,2,6,7].

The objective of this systematic review was to document the range of definitions of consensus in a representative, randomly selected, cross-disciplinary selection of Delphi studies. The rationale for the study was based on the fact

---

What is new?

- Definitions of consensus in Delphi studies vary widely and are poorly reported.

- Attainment of consensus is infrequently used as a criterion for ending a Delphi study.

- Methodologic criteria are proposed for the reporting of Delphi studies.

---

that although the Delphi technique is widely regarded as a consensus development technique, our impression from reading Delphi publications was that achievement of consensus is often assumed to occur by virtue of performing a Delphi study. We were concerned that many studies do not adequately define criteria for achievement of consensus and that even when consensus has been defined, it is not always clear whether the prespecified criteria for consensus have been a factor in deciding when to stop the Delphi process. We believe that failure to adequately define and use criteria for consensus challenges the notion that the results of a Delphi study reflect the consensus of the group of experts.

2. Method

2.1. Search strategy

The ISI Web of Science (Thompson Reuters, New York, NY) and Scopus (Elsevier, Amsterdam, NL) databases were used to conduct the search. The objective of using these databases was to obtain a cross-disciplinary selection of Delphi studies. The search was performed using ''Delphi'' as a text word term. The results were limited to articles published in calendar years 2000–2009 in English. The results were exported into an EndNote X4.0 database (Thompson Reuters) and duplicates removed via both the duplicate search function in EndNote and by manually reviewing the list of articles. All of the studies were assigned a unique identifier using consecutive numbers from one to the total number of articles. These articles formed the ''study population.''

2.2. Selection of publications for detailed review

Using a random number generator, the study population was randomly sorted and a second sequential number (review number) was assigned to each article. Beginning with the article with the lowest review number, we reviewed the abstract of each article sequentially to ensure that the article described the conduct of a Delphi study. To be selected, the article was required to describe a study whereby greater than one participant engaged in an iterative anonymous process that involved ranking or rating items by qualitative

or quantitative means. Where the abstract was not available, the full manuscript was obtained for review. We excluded articles describing the synthesis of results from Delphi studies or subsequent use of instruments and/or measures that had previously been developed by a Delphi process, but not the Delphi process itself. For feasibility reasons, only the first 100 articles to describe the conduct of a Delphi study were selected as the ''study sample'' and subjected to detailed review.

2.3. Review of Delphi studies

All manuscripts in the ''study sample'' were reviewed using a standardized datasheet by a single reviewer (I.R.D.). Using a random number table, a random selection of 23 articles was reviewed by a second reviewer (R.C.G.). In the development of the datasheet used for the review, the two reviewers jointly reviewed a number of Delphi studies from the study population, which had a review number that was greater than 2,500 and therefore were unlikely to be included in the study sample. This was done to ensure usability of the datasheet and consistency of the review. None of the articles comprising this ''validation data set'' were discussed among the primary and secondary reviewer before review and quantification of the measures of agreement.

Agreement between the reviewers was assessed using three key fields and the Kappa statistic. The fields assessed were: (1) whether the Delphi study had consensus development as a goal, (2) whether consensus was defined, and (3) the reason for ending the Delphi. It was decided a priori that provided the kappa statistic was greater than 0.8, the review of the primary reviewer would be considered to be valid and that double review of all articles would not be performed. In the event that the Kappa statistic fell below this threshold, it was planned for the second reviewer to review all articles, with a third reviewer adjudicating discrepancies.

The primary focus of the review was to document how consensus was defined. We recorded the description of consensus from each manuscript and then classified the various definitions into a number of mutually exclusive categories. We also specified whether the definition of consensus was stated a priori in the article and whether a specific threshold criterion for consensus was provided at that time. We also examined whether the decision to end the Delphi was based on achievement of consensus. If the decision was not based on consensus, we recorded the primary reason for ending the Delphi process. The results and discussion section of the manuscripts were also reviewed to determine whether the investigators commented on a consensus being reached. This was done irrespective of whether a definition for what constituted consensus was provided in the methods section.

Study characteristics were also collected including: discipline, region of origin, year of publication, number of participants, and the number of rounds that the Delphi study was performed for. Although the intention was not

to report the overall quality of the Delphi studies reviewed, we collected four variables that we view to be key indicators of a good quality Delphi study (Table 1). These measures were summed to create an overall quality score.

2.4. Analysis and interpretation

Initially we had planned, as our primary analysis to only focus on studies identified as having consensus as their goal, with a secondary analysis to include all studies. However, only two studies were selected that did not aim to address consensus. Consequently, only a single analysis was performed. However, the two studies that did not purport to assess consensus were excluded from the analyses of consensus-related issues. Analyses were done with PASW Statistics Version 17 (IBM SPSS, Chicago, IL).

Our analyses were primarily descriptive, with frequency counts provided for most variables. We also performed a limited number of analyses for trends within categorical variables (chi-square or Fisher's exact test). These analyses examined the relationship between measures of consensus, reasons for termination of the Delphi, and the quality score. We also explored the relationship between these indicators and whether the Delphi study was from the health care field (medicine, nursing, or allied health) or a non-health care field.

3. Results

A total of 3,056 unique manuscripts formed the "study population." About 247 of these articles were reviewed to obtain the 100 manuscripts describing a Delphi process (Appendix).

Characteristics of the "study sample" are provided in Table 2. The articles selected were evenly balanced among the years covered by the review. Almost half of the studies originated in North America, with studies from Europe making up the second largest group. Almost three-quarters of the Delphi studies were within health care. About 48 of the studies were completed in two rounds with 42 having three rounds. All of the studies with a single round had been planned for multiple rounds, although the predefined criterion for consensus was met on the first round.

3.1. Accuracy of a single reviewer

The level of agreement was excellent on the primary measures assessed with perfect agreement on whether the Delphi addressed consensus and whether consensus was

Table 1. Data fields indicating quality of a Delphi study

1. Stopping criteria specified	(Yes)
2. Planned number of rounds specified	(Yes)
3. Reproducible criteria for selection of participants	(Yes)
4. Criteria for dropping items at each round	(Yes or not applicable)

Table 2. Characteristics of the 100 studies in study sample

Parameters	Number
Year published	
2000–2001	17
2002–2003	27
2004–2005	12
2006–2007	21
2008–2009	24
Region	
North America	44
Europe	32
Asia	7
South Pacific	6
Middle East	2
South America	1
Multiple regions	8
Field of study	
Medicine	40
Nursing	12
Other health care	22
Other science	5
Education	6
Business/economics	15
Number of rounds	
1	2
2	48
3	42
4	4
≥ 5	4
Number of participants in final round	
≤ 10	14
11–25	40
25–50	24
51–100	12
≥ 100	5
Not reported	5

defined. The kappa was 0.82 for why the Delphi was stopped. We also assessed agreement on a number of secondary fields that we deemed to be important for the review (a priori definition of consensus, was consensus reached, was the decision to stop based on consensus, and our quality score measures). Agreement on these measures exceeded the prespecified criterion of 0.8.

3.2. Consensus and relationship to Delphi process

Table 3 examines the relationship between Consensus and the Delphi process. A total of 98 of the Delphi studies had development of consensus as their aim. In the two studies that did not purport to address consensus, there was a statement in the methods that the objective of these studies was to obtain divergent opinions.

Of the 98 studies that had consensus as an objective, a definition of consensus was given in 72 (73.5%) studies. This definition was clearly specified a priori in 64 of 72 (89%) studies. However, the criterion for the achievement of consensus was not specified in 21 of the 72 (33%) studies. The authors commented on consensus being reached in 86 (87.8%) of the 98 studies. In 15 of these studies, aside from a comment that consensus was reached,

Table 3. Consensus and relationship to Delphi process

Parameters	n (%)
Did the Delphi aim to address consensus?[a]	
Yes	98 (98)
No	2 (2)
Was consensus defined?[b]	
No	26 (26.5)
Yes	72 (73.5)
Was definition of consensus a priori?[c]	
No	8 (11.1)
Yes, but not the scale	21 (29.2)
Yes, including the scale	43 (59.7)
Was consensus reached?[b]	
No	12 (12.2)
Yes	86 (87.8)
What was the reason to stop Delphi?[b]	
Number of rounds	70 (71.4)[b]
Consensus reached	23 (23.5)
Stability of responses	3 (3.1)
Unclear	2 (2.0)

Denominators for calculating the percentages are as follows.
[a] All 100 Studies.
[b] The 98 studies that had consensus as an aim.
[c] The 72 studies that provided a definition for consensus.

no definition of consensus was available. Of the 64 studies where consensus was defined a priori, the authors stated that consensus was not achieved in only 1 study (1.5%).

3.3. Definitions of consensus

Table 4 lists the various definitions of consensus for the 72 studies where a definition was provided. The table also specifies whether the definition was a priori and whether a specific threshold for consensus was provided. The most common definition for consensus was percent agreement (25 studies), although a threshold was provided in only half of the studies. The proportion of ratings within a range was the next most common definition for consensus (16 studies). Although most studies provided this definition a priori, consensus was restricted to a limited portion of the range in half of the studies. Of those definitions based on a percentage or proportion, the median threshold, when specified, for determination of consensus was 75% (range: 50−97%).

3.4. Consensus and termination of the Delphi process

Despite consensus being commented on as being achieved in 88% of the studies, it was only deemed to be the primary reason for termination of the Delphi process in 23 (23.5%) studies. In most of these studies (19/23, 82.6%), an a priori definition of consensus was available. However, in only 14 (60.8%) studies was the fact that the Delphi would be terminated with the achievement of consensus stated in the methods. Most Delphi studies were run for a prespecified number of rounds (71.4%).

The Delphi studies from health care tended to be more likely to be stopped owing to the consensus being reached (20/74 vs. 3/26, Fishers exact *P*-value = 0.08). In fact, of

the 23 studies that were stopped for consensus, 20 of these were within health care. However, the proportion of studies commenting that the consensus was reached (20/26 vs. 66/74, Fishers exact *P*-value = 0.113) did not differ significantly within and outside of health care.

3.5. Proposed quality indicators

Table 5 lists the proposed quality indicators for the 98 consensus Delphi studies. Reproducible criteria for participants were available in two thirds of the studies. In most (82.7%), the number of rounds that the Delphi was to be performed was noted in the methods section of the article. Criteria for stopping the Delphi process, other than a statement of number of rounds, were present in 15 (15.3%). These criteria were related to consensus in 14 and stability of response in 1.

Three-quarters of the studies had a quality score of either two or three. Only 4 of the 98 studies had a quality score of four and 3 studies had a score of zero. The quality score did not appear to be related to whether consensus was defined or not [$\chi^2 = 2.67$, degrees of freedom (df) = 4, $P = 0.615$] or whether the study came from the health care field ($\chi^2 = 3.53$, df = 4, $P = 0.473$). However, a quality score of four was associated with a decision to stop based on consensus. All four studies with this score were stopped for this reason ($\chi^2 = 14.81$, df = 4, $P = 0.005$).

4. Discussion

Most Delphi studies were conducted for a specific number of rounds, without a formal criterion for consensus. Therefore, when authors conclude, as most did, that the results of the study reflect the consensus opinion, it would seem that the achievement of consensus within a Delphi study is assumed to be an integral part of the technique, as has been suggested in the literature [1−3,6,7]. The fact that consensus was not reached in only 1 of the 64 studies where an a priori definition of consensus was specified, provides some empirical support for this notion. Despite the fact that consensus may be the expected outcome of the Delphi process, we believe that there is a need to better define criteria for consensus and to document the degree of agreement together with the results of the Delphi process.

Despite the fact that the most Delphi studies in our cohort had consensus as their aim, in only a minority of the Delphi studies reviewed was consensus defined with a specific criterion. Furthermore, this criterion was the reason for termination of the Delphi process, usually on the basis of an a priori definition. However, we believe that an optimal approach would be to formally define criteria a priori, for what constitutes consensus rather than assume it to be an automatic outcome at the conclusion of a Delphi study. In addition to this, investigators should also specify alternate stopping criteria, including possibly a maximum

Table 4. Definitions of consensus

Definition	Example	Overall number	A priori with threshold	A priori with no threshold	Post hoc
Formal measure of agreement	Kappa statistic, Crohnbach's alpha, intraclass correlation coefficient, Kendal's *W*	4		4	
Rand criteria	No more than two ratings outside of a three-point range including the median. Valid if rated as 7+ without disagreement	4	4		
Measure of central tendency	Median ranking used to indicate groups ranking	8	2	4	2
Percent agreement	>80% With same rating	25	11	10	4
Central tendency within a specific range (restricted)	Mean greater than seven on a nine-point scale	2	1	1	
Central tendency within a specific range (unrestricted)	Median between seven and nine for appropriate, one to three for inappropriate, or four to six for equivocal with a range less than three	5	5		
Proportion within a range (restricted)	90% Scoring 7+ on a nine-point scale	8	8		
Proportion within a range (unrestricted)	75% of the Participants rated 7, 8, 9 or 1, 2, 3	8	7		1
Decrease in variance	Interquartile range less than three on nine-point scale	6	4	1	1
Stability	<15% Change in distribution of responses	1	1		
Rank	Rank order	1		1	

number of rounds that the Delphi would be performed. If studies are to be performed for a specific number of rounds, then the authors should specify how the degree of agreement of the experts at the conclusion of the study will be quantified. Clear criteria for dropping or combining items should also be specified based on the level of agreement or disagreement with individual items. One of the limitations of a priori specification is that certain items may fall just below the threshold for what is fundamentally an arbitrary cut off. In the event that items, believed to be important fell just below the threshold for inclusion in the study, the authors could consider including these items as posteriori considerations provided that sufficient justification was provided.

Table 5. Proposed quality criteria for the 98 Delphi studies that aimed to assess consensus

Criteria	Number
Quality score components	
Were criteria for participants reproducible?	
Yes	65
No	33
Was the number of rounds to be performed stated?	
Yes	81
No	17
Were criteria for dropping items clear?	
Yes or not applicable	59
No	39
Stopping criteria other than rounds specified?	
Yes	15
No	83
Quality score (items)	
0	3
1	16
2	37
3	38
4	4

The authors used a wide range of definitions of consensus as has been previously documented [1,2,6,7]. The most common definition of consensus in this study was based on percent agreement, followed by the proportion of participants agreeing in a specific rating range. When consensus was defined by either a measure of central tendency or proportion for ratings within a range, in 10 of the 23 studies making use of such a criterion, consensus was restricted to a specific part of the range. Although this may be appropriate for particular questions, the definition fails to account for consensus within other regions of the scale. Owing to the diversity of topics covered by the Delphi method, there is no way to ascertain the validity of any specific definition of consensus [2].

To our knowledge, there are no validated quality indicators for Delphi studies. We therefore elaborated a set of four criteria that we propose as quality indicators. These criteria were selected on the basis of those that we believe would allow the Delphi study to be replicated by other investigators. The quality of the studies reviewed, as judged by these indicators, was moderate. It is, however, important to recognize that this score is based on what was reported in the article and may reflect issues with reporting, rather than with the quality of the study itself. We therefore propose that these criteria form a set of suggested elements that should be included in all Delphi publications.

It is important to acknowledge the limitations of this review. First, our intention in selecting the ISI Web of Science and Scopus databases, rather than a discipline specific database, was to obtain as much of a cross-disciplinary selection of publications as possible. Most publications, however, came from the health care field. It is possible that the reason for this relates to the fact that Delphi studies outside of health care field may be less likely to be published in academic journals. Also, we limited our search strategy to English,

Table 6. Proposed key methodologic criteria to report in publications of Delphi studies

Study objective
- Does the Delphi study aim to address consensus?
  - Is the objective of the Delphi study to present results (eg, a list or statement) reflecting the consensus of the group, or does the study aim to merely quantify the level of agreement?

Participants
- How will participants be selected or excluded?

Consensus definition
- How will the consensus be defined?
- If applicable, what threshold value will be required for the Delphi to be stopped based on the achievement of consensus?
  - What criteria will be used to determine when to stop the Delphi in the absence of consensus?

Delphi process
- Were items dropped?
  - What criteria will be used to determine which items to drop?
- What criteria will be used to determine to stop the Delphi process or will the Delphi be run for a specific number of rounds only?

which likely restricted studies from certain geographic regions. It is possible that had a more diverse set of articles been selected, alternate definitions and uses of consensus may have been found. Also, although we created a quality score, no attempt was made to validate this score—which would be difficult to do given the absence of a "gold standard," and the diverse subjects covered by the Delphi studies. However, we believe that at a minimum this score represents a reasonable initial approach for assessing the quality of reports of Delphi studies in the absence of other metrics. An additional limitation relates to the fact that all of the measures captured in this review are based on what was reported in the manuscript, reflecting that the issues identified may be one of poor reporting rather than of conduct of the studies themselves.

5. Conclusion

We believe that there is a need to improve the reporting of Delphi studies, along the lines of a CONSORT-like guideline, as is used for randomized controlled trials [8]. We propose that the investigators who undertake Delphi studies should report a standard set of quality indicators. Proposed indicators are listed in Table 6. If the aim of the Delphi study is to elicit consensus, then a clear definition for what constitutes consensus should be provided a priori together with threshold values that specify when consensus is reached. If the investigators plan to only quantify the degree of consensus, but not have consensus as a criterion to stop the Delphi study, this should also be explicitly stated.

Acknowledgments

References

[1] Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12:1—7.

[2] Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. J Health Serv Res Policy 1999;4:236—48.

[3] Duffield C. The Delphi technique: a comparison of results obtained using two expert panels. Int J Nurs Stud 1993;30:227—37.

[4] Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J Adv Nurs 2006;53:205—12.

[5] Steinert M. A dissensus based online Delphi approach: an explorative research tool. Technol Forecast Soc Change 2009;76:291—300.

[6] Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 2003;56:1150—6.

[7] Lindstone HA, Turoff M. The Delphi method: techniques and applications. NJ: Murray Turoff and Harold A. Newark, NJ: Linstone; 2002.

[8] Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;63:834—40.

Appendix. Articles included in the review

[1] Barrett S, Kristjanson LJ, Sinclair T, Hyde S. Priorities for adult cancer nursing research: a West Australian replication. Cancer Nurs 2001;24(2):88—98.

[2] Basson R, Berman J, Burnett A, Derogatis L, Ferguson D, Fourcroy J, et al. Report of the international consensus development conference on female sexual dysfunction: definitions and classifications. J Urol 2000;163:888—93.

[3] Beattie E, Mackway-Jones K. A Delphi study to identify performance indicators for emergency medicine. Emerg Med J 2004;21(1):47—50.

[4] Bennett RM, Broom DM, Henson SJ, Blaney RJP, Harper G. Assessment of the impact of government animal welfare policy on farm animal welfare in the UK. Anim Welf 2004;13(1):1—11.

[5] Bonnemaizon A, Cova B, Louyot MC. Relationship marketing in 2015: a delphi approach. Eur Manag J 2007;25(1):50—9.

[6] Broder MS, Landow WJ, Goodwin SC, Brook RH, Sherbourne CD, Harris K. An agenda for research into uterine artery embolization: results of an expert panel conference. J Vasc Interv Radiol 2000;11(4):509—15.

[7] Broomfield D, Humphris GM. Using the Delphi technique to identify the cancer education requirements of general practitioners. Med Educ 2001;35(10):928—37.

[8] Butler MM, Meehan TC, Kemple M, Drennan J, Treacy M, Johnson M. Identifying research priorities for midwifery in Ireland. Midwifery 2009;25(5):576—87.

[9] Campbell SM, Shield T, Rogers A, Gask L. How do stakeholder groups vary in a Delphi technique about primary mental health care

and what factors influence their ratings? Qual Saf Health Care 2004; 13:428–34.

[10] Cardoso MJ, Cardoso J, Santos AC, Barros H, Cardoso de Oliveira M. Interobserver agreement and consensus over the esthetic evaluation of conservative treatment for breast cancer. Breast 2006; 15(1):52–7.

[11] Catlin A, Carter B. Creation of a neonatal end-of-life palliative care protocol. J Perinatol 2002;22(3):184–95.

[12] Chang PC, Wang CP, Yuan BJC, Chuang KT. Forecast of development trends in Taiwan's machinery industry. Technol Forecast Soc Change 2002;69(8):781–802.

[13] Chang YC, Hsu CJ, Williams G, Pan ML. Low cost carriers' destination selection using a Delphi method. Tourism Manag 2008;29(5): 898–908.

[14] Clark AC, Scales AY. A delphi study of assessment practices in engineering education. Eng Des Graphics J 2003;67(2):21–32.

[15] Coben JH. Measuring the quality of hospital-based domestic violence programs. Acad Emerg Med 2002;9(11):1176–83.

[16] Cook SP, Birrell M. Defining an occupational therapy intervention for people with psychosis. Br J Occup Ther 2007;70(3):96–106.

[17] Costa CA. The status and future of sport management: a Delphi study. J Sport Manag 2005;19(2):117–42.

[18] Cramer CK, Epstein JB, Sheps SB, Schechter MT, Busser JR. Modified Delphi survey for decision analysis for prophylaxis of post-radiation osteonecrosis. Oral Oncol 2002;38:574–83.

[19] Dart RC, Goldfrank LR, Chyka PA, Lotzer D, Woolf AD, McNally J, et al. Combined evidence-based literature analysis and consensus guidelines for stocking of emergency antidotes in the United States. Ann Emerg Med 2000;36(2):126–32.

[20] De Lima L, Krakauer EL, Lorenz K, Praill D, Macdonald N, Doyle D. Ensuring palliative medicine availability: the development of the IAHPC list of essential medicines for palliative care. J Pain Symptom Manage 2007;33:521–6.

[21] Dean B, Barber N, Schachter M. What is a prescribing error? Qual Health Care 2000;9(4):232–7.

[22] Deane KHO, Ellis-Hill C, Dekker K, Davies P, Clarke CE. A Delphi survey of best practice occupational therapy for Parkinson's disease in the United Kingdom. Br J Occup Ther 2003;66(6):247–55.

[23] Dinnebeil L, McInerney W, Hale L. Understanding the roles and responsibilities of itinerant ECSE teachers through Delphi research. Topics Early Child Spec Educ 2006;26(3):153–66.

[24] Dole RL, Arvidson K, Byrne E, Robbins J, Schasberger B. Consensus among experts in pediatric occupational and physical therapy on elements of individualized education programs. Pediatr Phys Ther 2003;15(3):159–66.

[25] Edwards LH. Research priorities in school nursing: a Delphi process. J Sch Nurs 2002;18:157–62.

[26] Evans D, Volery T. Online business development services for entrepreneurs: an exploratory study. Entrepren Reg Dev 2001;13:333–50.

[27] Falkson CB, Bezjak A, Darling G, Gregg R, Malthaner R, Maziak DE, et al. The management of thymoma: a systematic review and practice guideline. J Thorac Oncol 2009;4:911–9.

[28] Faull K, Hills MD, Cochrane G, Gray J, Hunt M, McKenzie C, et al. Investigation of health perspectives of those with physical disabilities: the role of spirituality as a determinant of health. Disabil Rehabil 2004;26(3):129–44.

[29] Fick DM, Cooper JW, Wade WE, Waller JL, Maclean JR, Beers MH. Updating the Beers criteria for potentially inappropriate medication use in older adults: results of a US consensus panel of experts. Arch Intern Med 2003;163:2716–24.

[30] Fongwa MN, Sayre MM, Ruth Anderson NL. Quality indicator themes among African Americans, Latinos, and Whites. J Nurs Care Qual 2008;23(1):50–7.

[31] Gallagher P, Ryan C, Byrne S, Kennedy J, O'Mahony D. STOPP (Screening Tool of Older Person's Prescriptions) and START (Screening Tool to Alert doctors to Right treatment). Consensus validation. Int J Clin Pharmacol Ther 2008;46(2):72–83.

[32] Gebbie K, Merrill J, Hwang I, Gupta M, Btoush R, Wagner M. Identifying individual competency in emerging areas of practice: an applied approach. Qual Health Res 2002;12(7):990–9.

[33] Gorney B, Ness RG. Evaluation dimensions for full-time head coaches at NCAA Division II institutions. J Person Eval Educ 2000;14(1):47–65.

[34] Griffith LE, Hogg-Johnson S, Cole DC, Krause N, Hayden J, Burdorf A, et al. Low-back pain definitions in occupational studies were categorized for a meta-analysis using Delphi consensus methods. J Clin Epidemiol 2007;60:625–33.

[35] Guenthner JF, Michael KC, Nolte P. The economic impact of potato late blight on US growers. Potato Res 2001;44(2):121–5.

[36] Guimaraes TA, Borges-Andrade JE, Dos Santos Machado M, Vargas MRM. Forecasting core competencies in an R&D environment. R&D Manag 2001;31(3):249–55.

[37] Hallowell MR, Gambatese JA. Activity-based safety risk quantification for concrete formwork construction. J Construct Eng Manag 2009;135(10):990–8.

[38] Hatcher T, Colton S. Using the internet to improve HRD research: the case of the web-based Delphi research technique to achieve content validity of an HRD-oriented measurement. J Eur Ind Train 2007; 31(7):570–87.

[39] Hawk C, Schneider M, Ferrance RJ, Hewitt E, Van Loon M, Tanis L. Best practices recommendations for chiropractic care for infants, children, and adolescents: results of a consensus process. J Manipulative Physiol Ther 2009;32(8):639–47.

[40] Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM. Expert consensus on the desirable characteristics of review criteria for improvement of health care quality. Qual Health Care 2001; 10(3):173–8.

[41] Hejblum G, Ioos V, Vibert JF, Boelle PY, Chalumeau-Lemoine L, Chouaid C, et al. A web-based Delphi study on the indications of chest radiographs for patients in ICUs. Chest 2008;133:1107–12.

[42] Hess GR, King TJ. Planning open spaces for wildlife. Landsc Urban Plann 2002;58(1):25–40.

[43] Ireland RS. A clinical minimum data set for primary dental care. Br Dental J 2001;190(12):663–7.

[44] Jauhiainen A, Saranto K, Tossavainen K. Consumer-centered nursing with ICT: a futuristic viewpoint. Stud Health Technol Inform 2006; 122:425–9.

[45] Junttila K, Lauri S, Salanterã S, Hupli M. Initial validation of the perioperative nursing data set in Finland. Nurs Diagn 2002;13(2): 41–52.

[46] Kaliner M, Amin A, Gehling R, Crivera C, Chiao E, Vishalpura T, et al. Impact of inhaled corticosteroid-induced oropharyngeal adverse effects on treatment patterns and costs in asthmatic patients: results from a Delphi panel. Pharm Ther Dig 2005;30(10):573–601.

[47] Kasturiratne A, Pathmeswaran A, Fonseka MM, Lalloo DG, Brooker S, de Silva HJ. Estimates of disease burden due to land-snake bite in Sri Lankan hospitals. Southeast Asian J Trop Med Public Health 2005;36(3):733–40.

[48] Keil M, Tiwana A, Bush A. Reconciling user and project manager perceptions of IT project risk: a Delphi study. Inf Syst J 2002; 12(2):103–19.

[49] Kelly JS, Zimmerman LA, Reed K, Enger KS. Immunization information systems national research and evaluation agenda. J Public Health Manag Pract 2007;13(1):35–8.

[50] Kirchberger I, Glaessel A, Stucki G, Cieza A. Validation of the comprehensive international classification of functioning, disability and health core set for rheumatoid arthritis: the perspective of physical therapists. Phys Ther 2007;87:368–84.

[51] Kokko S, Kannas L, Villberg J. The health promoting sports club in Finland - A challenge for the settings-based approach. Health Promotion Int 2006;21(3):219–29.

[52] Kuo NW, Hsiao TY, Yu YH. A Delphi-matrix approach to SEA and its application within the tourism sector in Taiwan. Environ Impact Assess Rev 2005;25(3):259–80.

[53] Lachance LL, Hawthorne V, Brien S, Hyland ME, Lewith GT, Verhoef MJ, et al. Delphi-derived development of a common core for measuring complementary and alternative medicine prevalence. J Altern Complement Med 2009;15(5):489–94.

[54] Lanza ML, Kazis L, Lee A, Ericsson A. A community meeting protocol for assault prevention. Int J Group Psychotherapy 2003;53(3): 285–302.

[55] Lee CF, King BE. Using the Delphi method to assess the potential of Taiwan's hot springs tourism sector. Int J Tourism Res 2008;10(4): 341–52.

[56] Lee DS, Tran C, Flintoft V, Grant FC, Liu PP, Tu JV, et al. CCORT/CCS quality indicators for congestive heart failure care. Can J Cardiol 2003;19(4):357–64.

[57] Lightfoot WS, Hefti A, Mariotti A. Using a Delphi panel to survey criteria for successful periodontal therapy in anterior teeth. J Periodontol 2005;76:1508–12.

[58] Lin C, Tan B, Chang S. An exploratory model of knowledge flow barriers within healthcare organizations. Inf Manage 2008;45(5): 331–9.

[59] Lundvall AM, Olson DK. Agricultural health nurses. Job analysis of functions and competencies. AAOHN J 2001;49(7):336–46.

[60] Marshall AP, Currey J, Aitken LM, Elliott D. Key stakeholders' expectations of educational outcomes from Australian critical care nursing courses: a Delphi study. Aust Crit Care 2007;20(3): 89–99.

[61] Massoudi MS, Barker L, Schwartz B. Effectiveness of postexposure vaccination for the prevention of smallpox: results of a delphi analysis. J Infect Dis 2003;188:973–6.

[62] Mayo PH, Beaulieu Y, Doelken P, Feller-Kopman D, Harrod C, Kaplan A, et al. American College of Chest Physicians/La Societe de Reanimation de Langue Francaise statement on competence in critical care ultrasonography. Chest 2009;135:1050–60.

[63] McGory ML, Shekelle PG, Ko CY. Development of quality indicators for patients undergoing colorectal cancer surgery. J Natl Cancer Inst 2006;98:1623–33.

[64] McKenna H, Keeney S, Bradley M. Generic and specialist nursing roles in the community: an investigation of professional and lay views. Health Soc Care Community 2003;11(6):537–45.

[65] Meadows AB, Finstuen K, Hudak RP. Pharmacy executives: leadership issues and associated skills, knowledge, and abilities in the U.S. Department of Defense. J Am Pharm Assoc 2003;43(3): 412–8.

[66] Medeiros LC, Kendall P, Hillers V, Chen G, DiMascola S. Identification and classification of consumer food-handling behaviors for food safety education. J Am Diet Assoc 2001;101:1326–39.

[67] Meijer R, Ihnenfeldt D, Vermeulen M, De Haan R, Van Limbeek J. The use of a modified Delphi procedure for the determination of 26 prognostic factors in the sub-acute stage of stroke. Int J Rehabil Res 2003;26(4):265–70.

[68] Miltner RS. Identifying labor support actions of intrapartum nurses. J Obstet Gynecol Neonatal Nurs 2000;29(5):491–9.

[69] Mitchell RJ, Williamson AM, O'Connor R. The development of an evaluation framework for injury surveillance systems. BMC Public Health 2009;9:260.

[70] Moffat SO, Cubbage FW, Holmes TP, O'Sullivan E. Characterizing the sustainable forestry issue network in the United States. Forest Pol Econ 2001;2(3–4):307–18.

[71] Moldrup C, Traulsen JM, Almarsdottir AB. Medically-enhanced normality: an alternative perspective on the use of medicines for non-medical purposes. Int J Pharm Pract 2003;11(4):243–9.

[72] Monterosso L, Dadd G, Ranson K, Toye C. Priorities for paediatric cancer nursing research in Western Australia: a Delphi study. Contemp Nurse 2001;11(2):142–52.

[73] Moore SA, Wallington TJ, Hobbs RJ, Ehrlich PR, Holling CS, Levin S, et al. Diversity in current ecological thinking: implications for environmental management. Environ Manag 2009;43(1): 17–27.

[74] Morris CJ, Cantrill JA, Hepler CD, Noyce PR. Preventing drug-related morbidity—determining valid indicators. Int J Qual Health Care 2002;14(3):183–98.

[75] Norcross JC, Hedges M, Prochaska JO. The Face of 2010: a delphi Poll on the future of Psychotherapy. Prof Psychol Res Pract 2002; 33(3):316–22.

[76] O'Boyle C, Jackson M, Henly SJ. Staffing requirements for infection control programs in US health care facilities: Delphi project. Am J Infect Control 2002;30(6):321–33.

[77] Oliver Schwarz J. Assessing the future of futures studies in management. Futures 2008;40(3):237–46.

[78] Playfor S, Jenkins I, Boyles C, Choonara I, Davies G, Haywood T, et al. Consensus guidelines on sedation and analgesia in critically ill children. Intensive Care Med 2006;32(8):1125–36.

[79] Preheim GJ, Armstrong GE, Barton AJ. The new fundamentals in nursing: introducing beginning quality and safety education for nurses' competencies. J Nurs Educ 2009;48(12):694–7.

[80] Quintana JM, Arostegui I, Azkarate J, Goenaga JI, Tobio R, Aranburu JM, et al. Use of explicit criteria for total hip joint replacement fixation techniques. Health Policy 2002;60(1):1–16.

[81] Qureshi AI, Feldmann E, Gomez CR, Johnston SC, Kasner SE, Quick DC, et al. Consensus conference on intracranial atherosclerotic disease: rationale, methodology, and results. J Neuroimag 2009; 19(Suppl 1):1S–10S.

[82] Rice K. Priorities in k-12 distance education: a delphi study examining multiple perspectives on policy, practice, and research. Educ Technol Soc 2009;12(3):163–77.

[83] Riddle MS, Tribble DR. Reaching a consensus on management practices and vaccine development targets for mitigation of infectious diarrhoea among deployed US military forces. J Eval Clin Pract 2008;14:266–74.

[84] Robertson HA, MacKinnon NJ. Development of a list of consensus-approved clinical indicators of preventable drug-related morbidity in older adults. Clin Ther 2002;24:1595–613.

[85] Shriberg D, Bonner M, Sarr BJ, Walker AM, Hyland M, Chester C. Social justice through a school psychology lens: definition and applications. School Psych Rev 2008;37(4):453–68.

[86] Smith DW, Kelley P, Maushak NJ, Griffin-Shirley N, Lan WY. Assistive technology competencies for teachers of students with visual impairments. J Vis Impair Blind 2009;103(8):457–69.

[87] Suwaratchai P, Sithisarankul P, Sriratanban J, Chenvidhya D, Phonburee W. Utilize the modified Delphi technique to develop trauma care indicators. J Med Assoc Thai 2008;91(1):99–103.

[88] Taleai M, Sharifi A, Sliuzas R, Mesgari M. Evaluating the compatibility of multi-functional and intensive urban land uses. Int J Appl Earth Obs Geoinf 2007;9(4):375–91.

[89] Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101(8): 1900–20.

[90] van der Vorm A, Vernooij-Dassen MJ, Kehoe PG, Olde Rikkert MG, van Leeuwen E, Dekkers WJ. Ethical aspects of research into Alzheimer disease. A European Delphi Study focused on genetic and non-genetic research. J Med Ethics 2009;35:140–4.

[91] van Tubergen A, van der Heijde D, Anderson J, Landewe R, Dougados M, Braun J, et al. Comparison of statistically derived ASAS improvement criteria for ankylosing spondylitis with clinically relevant improvement according to an expert panel. Ann Rheum Dis 2003;62(3):215–21.

[92] Vargus-Adams JN, Martin LK. Measuring what matters in cerebral palsy: a breadth of important domains and outcome measures. Arch Phys Med Rehabil 2009;90:2089–95.

[93] Weir CR, Hicken BL, Rappaport HS, Nebeker JR. Crossing the quality chasm: the role of information technology departments. Am J Med Qual 2006;21(6):382–93.

[94] Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep

quality in patients with rheumatoid arthritis. J Rheumatol 2007;34: 280–9.

[95] Whitehead T, Preston J, Holvad T. The whole-life impacts of transport-charging interventions on business performance: a time-marching framework. Environ Plan A 2005;37(5):877–94.

[96] Wilde VE, Ford JJ, McMeeken JM. Indicators of lumbar zygapophy-seal joint pain: survey of an expert panel with the Delphi technique. Phys Ther 2007;87:1348–61.

[97] Woollard M. Emergency calls not requiring an urgent ambulance response: expert consensus. Prehosp Emerg Care 2003;7(3): 384–91.

[98] Wyrwich KW, Tierney WM, Babu AN, Kroenke K, Wolinsky FD. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. Health Serv Res 2005;40:577–91.

[99] Zhang W, Doherty M, Leeb BF, Alekseeva L, Arden NK, Bijlsma JW, et al. EULAR evidence based recommendations for the management of hand osteoarthritis: report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007;66(3):377–88.

[100] Zins C. Conceptual approaches for defining data, information, and knowledge. J Am Soc Inf Sci Technol 2007;58(4):479–93.