

TECHNICAL REPORT

Department of Computer Science

The DELPHI Method as a Consensus and Knowledge
Acquisition Tool for the Evaluation of the DIABETES
System for Insulin Administration.

Barbara-Vivian Ambrosiadou
Dimitrios G Goulis

June 1999

Report Number: 332

The DELPHI Method as a Consensus and Knowledge Acquisition Tool for the Evaluation of the DIABETES System for Insulin Administration

Barbara-Vivian Ambrosiadou ¹

Dimitrios G.Goulis ²

1. Computer Science Department, Faculty of Engineering and Information Sciences, University of Hertfordshire, Hatfield Campus, College Lane, Hatfield, Herts, AL10 9AB, UK
2. Section of Endocrinology and Metabolic Medicine, Imperial College School of Medicine, St. Mary's Hospital, Praed St., London W2 1NY, UK

Address for correspondence :

Dr. Barbara Vivian Ambrosiadou

Senior Lecturer

Department of Computer Sciences,

Faculty of Engineering and Information Sciences,

University of Hertfordshire,

Hatfield Campus, College Lane, Hatfield, Herts

AL10 9AB, UK

tel. 01707-284347

FAX 01707-284303

e-mail b.v.ambrosiadou@herts.ac.uk

Running title :

The DELPHI Method for the DIABETES expert system.

Abstract

DIABETES is a decision-support system in the field of insulin administration. System performance evaluation is particularly difficult because of the absence of a uniform decision-making model followed by the specialists. The DELPHI method has been selected since it is appropriate for those domains where there is divergence among experts' opinions. The DELPHI approach helps a number of diabetologists arrive at a consensus and in thus it facilitates performance evaluation and further knowledge acquisition.

Insulin administration regimes, for 100 diabetic subjects, were proposed by DIABETES and five diabetologists (round 1). These suggestions were compiled and forwarded back to the specialists who proposed a second management approach (round 2). In each case, the experts were asked to justify their decision and comment on the suggestions of their colleagues and DIABETES. A novel scoring system for quantification of agreement was adopted.

The DELPHI procedure significantly increased the agreement among the diabetologists from 67% to 84% (X^2 , $p=0.0001$). The agreement between experts' and DIABETES recommendations was to a level of 54%. A total of 3500 comments were acquired by the experts.

Keywords :

DELPHI method, diabetes mellitus, performance evaluation, knowledge acquisition.

1. Introduction

Diabetes Mellitus is a chronic disease that may result in serious complications if it is not managed effectively from its earliest stages [1,2]. Due to its major medical, social and economic significance, medical informatics has tried for many years to provide the medical profession with tools for the diagnosis and processing of clinical and laboratory information, in order to achieve the best possible management [3-6].

In recent years, information technology applications have been developed for clinical diabetes care. Decision support systems in this domain require extensive evaluation concerning system functionality, reliability and user-friendliness, as well as correctness, usefulness and doctor acceptance. For these reasons, very few have been applied in everyday clinical practice.

One of the most common problems arising in the development and evaluation of a decision support system in the field of diabetes mellitus and in particular, in the domain of insulin administration, is the **absence of a uniform decision making model** which is followed by the specialists. Some diabetologists fully discard a number of insulin regimes and prefer to adjust insulin dosages, in order to control the glucose level, whereas others use a wider range of insulin regimes and insulin types, even combinations of insulin injections and hypoglycaemic agents. In other words, there is no 'absolute' or 'objective' method for treatment, as the same patient could be effectively controlled by two specialists following a different insulin regime - dosage combination.

An additional problem that arises, as in many other cases of decision support system development, is **knowledge acquisition**. In domains where there is divergence of specialists' opinions, such as the management of diabetes mellitus, the standard procedure of interviewing the experts may be insufficient, as the system developer may not know how to handle the many different and, usually, conflicting items of knowledge that have been acquired. Automated approaches which promise to resolve the so called problem of 'knowledge acquisition', such as machine learning algorithms, can not face the challenge of extracting knowledge from multiple sources of expertise, which often disagree with one another. The example cases presented to any such algorithmic approach may find it difficult to converge even to a locally optimal solution.

This paper addresses the issues of the DIABETES decision support system performance evaluation and knowledge acquisition based on the knowledge acquired from multiple sources of expertise by applying the DELPHI procedure for consensus of opinions.

2. Objectives

DIABETES is a rule-based decision support system [7]. It facilitates insulin administration and dose adjustment for optimal blood glucose control, in patients with insulin-treated diabetes mellitus. The DELPHI method is a technique that can be used for knowledge acquisition [8] and convergence of specialist opinion in a specific area [9-14].

In this study, a modified DELPHI approach was followed in order to arrive at a consensus of opinion within a group of diabetologists. This consensus was subsequently considered as “standard” domain knowledge. Areas of relative disagreement were highlighted. Simultaneously, additional knowledge was acquired from the specialists, in the form of suggested regimes and clinical comments. DIABETES’ performance was evaluated through comparison with the human experts. Finally, in addition to performance evaluation, the results from the DELPHI method indicated the “weak” points in DIABETES’ knowledge base and inference engine. The indication of these “weak” points and the new knowledge acquired can contribute, at a later time, to DIABETES’ improvement.

In summary, the objectives of the study were :

1. convergence of specialist opinion (consensus) in the field of insulin therapy of diabetes mellitus
2. indication of areas of relative disagreement among experts
3. knowledge acquisition from experts, in the domain of insulin therapy
4. DIABETES performance evaluation, through comparison with human experts
5. indication of “weak” areas in DIABETES’ knowledge base and inference engine

3. DIABETES system and DELPHI method

3.1 *The DIABETES System*

The DIABETES rule-based decision support system can adjust the insulin regime and dose of type 1 and 2 insulin-dependent diabetic patients [7]. As it can explain its advice and reasoning, it can be used as an aid in objective, consistent and well-reasoned patient management. As **input** the system uses (1) patient blood glucose profile, (2) clinical symptoms at specific time ranges, in those cases where blood glucose values are unavailable, and (3) insulin regime (number of injections, insulin types, number of insulin units per injection and time of application).

DIABETES, using the input data and its inference engine, produces a set of **intermediate values** : blood glucose estimates, M-values, adjusted insulin dosages and an assessment of the patient glycaemic control. As **output**, it gives (1) adjustments to insulin regime, (2) explanations for these adjustments, (3) comments on patient glycaemic control and (4) suggestions for follow-up.

The **production rule method** has been adopted for knowledge representation purposes. A typical DIABETES rule is :

IF (patient is on regime 1)
THEN (pre-lunch blood glucose should not be low)
AND (should not have symptoms of hypoglycaemia before lunch)

The **knowledge base** is divided into two parts, the static and the dynamic knowledge bases. The first consists of knowledge rules, which do not change as the program is modified. The second holds data values, specific to the patient case being processed by the program at the time. These dynamic data are created at run time and are removed at the end of a program execution session. The dynamic data are organized using the “object/attribute/value” paradigm, which is instantiated to different values from case to case. The conditions of a production rule are usually expressed in terms of such instantiations.

The **inference engine** of the DIABETES system functions by top-down refinement. DIABETES creates the main tasks (adjustments to insulin regime) which are decomposed into subtasks, for example adjusting the breakfast insulin or the lunch insulin. DIABETES performs these subtasks by forward chaining on

production rules. The conflict resolution strategy that is applied in firing the rules is that the special-case rule is preferred to the general-case rule.

The **user interface** accepts data input in a way that mimics a human expert's insulin management system. Explanation and advice are given in a paragraph of natural language.

3.2 The DELPHI Method

The DELPHI method can be used as a technique for increasing agreement among experts on a specific subject. This method is characterised by anonymity, feedback and iteration [15-21]. Every specialist in the group is expected to give his/her opinion without reference to the others (anonymity). In this way, there is interdependence and objectivity in the decision making of the specialists. A 'coordinator' selects the results of the first round, which are re-submitted to the experts (feedback), without knowing who provided which opinion. With this additional information, the specialists give their opinion once again (iteration). This process can be repeated for several rounds until the judgments converge.

Although the DELPHI method has been criticized for its possible lack of accuracy, it has been widely used in the medical field and accepted as a viable knowledge acquisition process. The method is most valuable in situations where the "gold standard" is unknown. DELPHI studies have been performed in the fields of internal medicine [22], infectious diseases [23-24], gastroenterology [25], cardiology [26-27], neurology [28-29], paediatrics [30-32], psychiatry [33], obstetrics [34], anesthesiology [35], AIDS [36], oncology [37], lung [38] and nuclear medicine [39]. In the wider field of health, there are also applications in nursing care [40], dentistry [41], medical informatics [42], and medical education [43]. Very few studies have been performed for diabetes mellitus [44].

4. Subjects and Methods

4.1 Subjects

For this retrospective study, all insulin-treated diabetics that attended the outpatient clinic of a large district hospital, during a period of four months, were recruited (n=384). Of them, the patients receiving one of the four insulin regimes supported by DIABETES were suitable for the study (n=295). The subjects were subsequently divided into four groups, according to the insulin regime that they were using. At the end of the classification process, each group contained approximately 75 patient records. A random number generator selected 100 cases, 25 for each of the four groups. The characteristics of the 100 diabetic patients that were included are shown in table 1.

In this DELPHI study five experienced diabetologists participated. All of them had more than eight-year experience in the management of diabetes mellitus at inpatient and outpatient basis. None of them was familiar with the cases included in the study, which, in addition, were distributed anonymously.

4.2 Methods

General. In all patient case sheets the format of an insulin regime was according to the pattern :

regime / 1st injection units, 2nd injection units, 3rd injection units, 4th injection units

For the uniformity of the results, it has been assumed that all regimes consisted of four injections. Of course, if the regime included fewer injections, then zero unit “injections” were noted.

Before initiation of the DELPHI procedure, all 100 cases were processed by the DIABETES system, and a similar number of suggestions and comments were provided by the system.

Round 1 : input. Each expert received a copy of each one of the 100 cases. These case-sheets contained the same data used for DIABETES input, namely insulin regime, units of each injection, blood glucose values during the previous 24 hours and patient’s symptoms during the same period. In addition, the DIABETES output was included. A typical round 1 input sheet is illustrated in table 2.

Round 1 : output. The specialist, in his answer sheet, had to suggest the insulin regime and units which constituted the best therapeutic approach, according to his opinion. In addition, he had to specify his level of agreement with DIABETES and, finally, to add a short agreement or disagreement comment in order to

justify his suggestion. Thus, during the first round, the five specialists provided $100 \times 5 = 500$ answers and comments. Table 3 is part of the answer sheet of one of the specialists.

Round 1 : processing. After round 1, the study coordinator collected and processed the answers. The main task was to summarise the experts' answers in each individual case and make the comments shorter. This was in order to produce a case-sheet that had to be re-submitted to the specialists for round 2.

Round 2 : input. Each specialist received the 100 cases again, and a similar number of the summarised answer-sheets, such as the one shown in table 4. He was also informed of the six answers from round 1 (his own answer, DIABETES' and four from his colleagues), without knowing who had provided each answer. He was also aware of the six comments attached to each answer.

Round 2 : output. The expert had to suggest the patient treatment once again and comment on each of the "answer-comment" pairs from round 1 (as shown in table 4). Thus, during round 2, $100 \times 5 = 500$ answers and $100 \times 6 \times 5 = 3000$ comments were produced.

Round 2 : processing. In each round, there was a need to make a quantitative estimation of the degree of agreement (i) among specialists and (ii) between specialists and DIABETES. For each case five answers were given in each round, one from every specialist. Consequently, there were $(5 \times 4) / 2 = 10$ "pairs of answers". Thus, the degree "10" meant absolute agreement among specialists while the degree "0" absolute disagreement. The degree of agreement between specialists is given by the point scoring system of table 5. In the same way, the basis for point scoring of agreement between specialists and DIABETES is given in table 6.

Ethics. The study was approved by the "Hippocraton" General Hospital Ethics Committee, Greece, from where all the cases were recruited. All patients gave informed written consent in order that data concerning their care be included in the study.

Statistics. Subject characteristics were of normal distribution and results were expressed as mean \pm standard error of the mean (SEM). The Chi-square test was used to indicate differences between groups, after each round of the DELPHI process. Statistical analysis was performed using Arcus Quickstat software, Biomedical version 1.0, Longman Logotron.

5. Results

Consensus among experts. The maximum score of 1000 points refers to complete agreement (10 points per case) among specialists, for each one of the 100 case reports. During round 1, a total of 674 points were collected (agreement between specialists of 67.4%), which increased to 842 points after round 2 feedback (percentage agreement 84.2%). The difference was statistically significant (X^2 , $p=0.001$). The results are presented in table 7. The difference of 168 points between round 1 and 2 occurred because some of the specialists changed their opinion during round 2. As is obvious from table 5, a change in specialist opinion could result in different score change: (1) if, as a consequence of the change, total agreement was achieved (all five specialists had the same opinion), then four points were added (from 6 to 10), (2) if near total agreement was achieved (four specialists had the same opinion and only one remained to disagree), then two points were added (from 4 to 6), and (3) in any other occasion, only one point was added. Table 8 illustrates the absolute and per cent points added during round 2, according to degree of agreement. There was no change in opinion that decreased the degree of agreement among specialists.

Knowledge acquisition. During the DELPHI process, 3500 comments (500 from round 1 and 3000 from round 2), were collected from the 100 cases presented to the experts. From these, new knowledge was acquired. The agreement comments gave a better understanding of the rationale behind specific decisions, and the disagreement comments were valuable in indicating areas of contradiction among experts or between experts and DIABETES.

Disagreement among experts. The disagreement comments that existed among the specialists were classified into main areas. The commonest reasons of contradiction were related to : (1) existence of approximately 20 different insulin treatment regimes, a fact that was an obstacle for greater opinion convergence, (2) consistent non-use of specific regimes by some experts, (3) combination of insulin therapy with anti-diabetic tablets by some experts, (4) lack of data in the input sheet, that may have increased opinion convergence (that is, patient's age and latest value of glycosylated haemoglobin (HbA_{1c})), and (5) slightly different approaches in the adjustment of insulin units (that is, some experts change insulin by only an even number of units).

DIABETES performance evaluation. A maximum score of 1000 points referred to complete agreement (10 points per case) between specialists and DIABETES for each one of the 100 case reports. During

round 1, a total of 522 points were collected (52.2%), which increased to 544 (54.4%) after round 2 feedback. The difference was not statistically significant (X^2 , $p=NS$). Results are presented in table 7.

Indication of “weak” areas in DIABETES knowledge base and inference engine. The disagreement comments between experts and DIABETES can be classified into three main areas. The comments indicated that, in order for the level of agreement to be increased, the new system must include : (1) new insulin treatment regimes, (2) combinations of insulin with anti-diabetic tablets, and (3) new input parameters as decision-making factors : age of patient, existence of diabetic complications, existence of concomitant diseases, glycosylated haemoglobin (HbA_{1c}) and home blood/urine glucose monitoring.

6. Discussion

One of the major problems encountered in the evaluation of decision-support systems in the domain of insulin-treated diabetes mellitus is the absence of a consistent patient management strategy, which could be used as a 'golden standard' for comparison between specialists and computer systems. Even then, the task of evaluation is not straightforward, as insulin therapy constitutes a highly complicated field of knowledge, with no clear-cut solutions. Finally, as medical experts provide different opinions concerning the selection of insulin regime and dosage for a given patient, the task of knowledge acquisition proves to be a very challenging one.

In order to address these problems, we applied a modified DELPHI procedure, using a panel of five diabetologists and DIABETES, a decision-support system in the domain of insulin therapy. Several points have been raised for discussion.

Concerning **subject selection**, the only criterion for inclusion of a patient case was that the patient had to be treated with one of the four insulin regimes supported by DIABETES. In any other case, it would be impossible to evaluate DIABETES' performance. We had to limit our cohort size to 100 patients, as, otherwise, the number of answers and comments generated by the multiple rounds of the DELPHI process would be very large and, therefore, difficult to handle and analyse. Finally, we decided to use a randomisation procedure concerning the selection of cases rather than provide the experts with pre-selected "grey" areas. This approach was preferred, not least because it was not easy to locate such patients. In any case, what is "grey" for one expert may be relatively straightforward for another. Therefore, the really difficult areas can be indicated only by contradiction among experts. The sample should be of adequate size for "grey" cases to be likely.

Concerning **methodology**, we decided to provide the specialists with DIABETES' answers and comments as early as round 1, instead of "hiding" them among round 2 responses. This decision was made in order to focus the experts' attention onto weak aspects of the system. We did not believe that we would confront an "anchoring" phenomenon, as the participating diabetologists were very experienced and they would not change their patients' management due to a computer program suggestion.

The kappa statistic is an efficient method of quantification of agreement between two specialists [45] or between a specialist and a computer system. On the other hand, if there is no agreement between the experts, they must be included in the DELPHI process. Lack of agreement will make kappa statistics inapplicable, as there is evidence to suggest that the analysis of more than two experts is a complex and controversial subject

[46]. Therefore, we had to introduce a new scoring system. We selected a simple one, that counts all the possible “pairs of agreement” among the five experts. According to this system (table 5), if three experts agree, the overall agreement across the group is higher if the other two agree to something different than if they disagree with each other. The logic behind this approach is that the opinion of an experienced diabetologist can not be considered as “wrong”, but, simply, as a “different option”. Therefore if there are two instead of three different opinions, the degree of agreement must be higher. A similar system was introduced for the quantification of agreement between experts and DIABETES.

Concerning **results**, we showed a statistically significant increase in the degree of agreement among experts, from 67.4% to 84.2%. Expanding the discussion made in the previous paragraph, we do not believe that this was due to a radical change in experts’ views, but, rather, reflects the fact that, according to their opinion, other therapeutic options can be followed as well, with equally satisfactory results. Further evidence for this claim is the fact that, in round 2, there was no decrease in agreement among experts, in any one of the 100 cases. From the study’s point of view, this is still a good result, as now there are fewer options to be included to the decision support system. An interesting idea would be to repeat round 1, with different but similar cases, in order to find out if the increase of the level of agreement, achieved with the DELPHI process, will remain so. But, the answer to this question may well be into the negative as, probably, the changes in expert’s opinion are not “permanent”, but only indicative of alternative management pathways.

With the application of a two-round DELPHI method we managed to collect 3500 comments from 5 specialists, concerning 100 cases of insulin-treated diabetic patients. The procedure was not time-consuming, as it took two weeks for completion of the two-round routine. It did not involve any interview or movement of specialists from their usual places of work. Following an oral communication with the experts, after completion of the study, they reported that it was much easier for them to deal with specific cases or comment on other experts’ opinions than trying to postulate general rules that describe their decision-making process. Therefore, although in this study we did not compare DELPHI with other knowledge acquisition techniques, we feel that the method constitutes a useful alternative tool concerning knowledge extraction, in domains similar to the one studied.

The scoring system provided a method for quantification of agreement between specialists and DIABETES, which aided the system’s performance evaluation, given that a human expert is considered as the “golden standard” in decision-making. The group of specialists agreed with DIABETES’ suggestions at a level

of 54%, which was considerably lower than that among the specialists themselves (84%). However, the fact that the specialists' level of agreement is only 84% indicates that there is no 'global optimal' decision in the context of insulin administration. This also indicates that both the knowledge base and the inference engine of DIABETES have not reached the level of "maturity" of a human expert and have to be improved.

Towards this end, the experts' comments accompanying their treatment suggestion were most valuable in identifying the weak points of the DIABETES system, as well as the areas of relative disagreement among the diabetologists. What remains is for these suggestions to be included in an improved DIABETES system.

7. Conclusions

The DELPHI method constitutes an effective procedure for reaching consensus opinion among specialists. This is particularly useful in “grey” fields, such as the treatment of insulin-dependent diabetes mellitus, where discrepancy exists among the specialists’ opinions.

The same method can be used for decision-support system evaluation, where the computer software can be included in the procedure as “another expert”. This allows quantification of the degree of agreement between the human experts and the system. Finally, the DELPHI method constitutes a knowledge acquisition tool, for “grey” areas of a knowledge base.

Using a modified DELPHI approach, it has been shown that the DIABETES system harmonises its decisions with those of specialists to a level of 54% and therefore further improvements to the system’s knowledge base and inference engine are required.

References

1. Unger R, Foster D. (1992) Diabetes mellitus. In : Wilson J, Foster D (eds). *Williams textbook of endocrinology*, 8th ed. (Philadelphia, WB Saunders), 1255-1334.
2. Karam J, Forsham P. (1994) Pancreatic hormones and diabetes mellitus. In : Greenspan F, Baxter J (eds). *Basic & Clinical Endocrinology*, 4th ed. (London:Appleton & Lange), 571-6.
3. Schiffrin A, Mihic M, Leibel B, Albisser A. (1985) Computer-assisted insulin dosage adjustment. *Diabetes Care*, **8**:545-552.
4. Pernich N, Rodbard D. (1986) Personal computer to assist with self-monitoring of blood glucose and self-adjustment of insulin dosage. *Diabetes Care*, **9**:61-69.
5. Albisser A, Beyer J. (1986) Meeting report : first international symposium on computer systems for insulin adjustment in diabetes mellitus. *Diabetes Care*, **9**:208-209.
6. Wise P. (1991) Computers in Diabetes Management. In : Pickup JC, Williams G (eds). *Textbook of diabetes*, vol II. (Oxford, Blackwell Scientific), 854-860.
7. Ambrosiadou B, Goulis DG, Pappas C. (1996) Clinical evaluation of the DIABETES expert system for decision support by multiple regimen insulin dose adjustment. *Computer Methods and Programs in Biomedicine*, **49**:105-115.
8. Dalkey NC. (1969) An experimental study of group opinion: The Delphi method. *Futures* **2**:408-426.
9. Fink A, Kosecoff J, Chassin M, Brook RH. (1984) Consensus methods: characteristics and guidelines for use. *American Journal of Public Health*, **74**:979-983.
10. Begg CB, Metz CE. (1990) Consensus diagnoses and "gold standards". *Medical Decision Making*, **10**:29-30.
11. Whitman NI. (1990) The Delphi technique as an alternative for committee meetings. *Journal of Nursing Education*, **29**:377-379.
12. Peiro Moreno S, Portella Argelaguet E. (1993) Consensus doesn't always mean agreement: limitations of consensus methods in health services. *Gac-Sanit*, **7**:294-300.
13. Jones J, Hunter D. (1995) Consensus methods for medical and health services research. *BMJ*, **311**:376-380.

14. Pearson SD, Margolis CZ, Davis S, et al. (1995) Is consensus reproducible? A study of an algorithmic guideline development process. *Medical Care*, **33**:643-660.
15. Milholland A, Wheeler S, Heieck J. (1973) Medical assesement by the Delphi group opinion technic. *New England Journal of Medicine*, **288**:1272-1275.
16. Rockwell MA. (1973) The Delphi procedure: knowledge from goat viscera? *New England Journal of Medicine*, **288**:1298-1299.
17. Kumaran KM, Lemieux M, Satchell G. (1976) Problem solving with the Delphi technique. *Dimens-Health-Serv*, **53**:34-35.
18. Goodman CM. (1987) The Delphi technique: a critique. *J-Adv-Nurs*, **12**:729-734.
19. Duffield C. (1988) The Delphi technique. *Aust-J-Adv-Nurs*, **6**:41-45.
20. Jairath N, Weinstein J. (1994) The Delphi methodology (Part one): A useful administrative approach. *Can-J-Nurs-Adm*, **7**:29-42.
21. Jairath N, Weinstein J. (1994) The Delphi methodology (Part Two): A useful administrative approach. *Can-J-Nurs-Adm*, **7**:7-20.
22. Vermeulen LC, Jr, Ratko TA, Erstad BL, Brecher ME, Matuszewski KA. (1995) A paradigm for consensus. The University Hospital Consortium guidelines for the use of albumin, non-protein colloid, and crystalloid solutions. *Archives of Internal Medicine*, **155**:373-379.
23. Barrett-Connor E. (1984) Natural history of intestinal parasites in asymptomatic adults. *J-Fam-Pract*, **19**:635-639.
24. Reves RR, Johnson PC, Ericsson CD, DuPont HL. (1988) A cost-effectiveness comparison of the use of antimicrobial agents for treatment or prophylaxis of travelers' diarrhea. *Archives of Internal Medicine*, **148**:2421-2427.
25. Moorman PW, van Ginneken AM, van der Lei J, et al. (1994) The contents of free-text endoscopy reports: an inventory and evaluation by peers. *Endoscopy*, **26**:531-538.
26. Willems JL, Arnaud P, van Bommel JH, et al. (1985) Establishment of a reference library for evaluating computer ECG measurement programs. *Comput-Biomed-Res*, **18**:439-457.
27. Kors JA, Sittig AC, van Bommel JH. (1990) The Delphi method to validate diagnostic knowledge in computerized ECG interpretation. *Methods-Inf-Med*, **29**:44-50.

28. Swaine BR, Sullivan SJ, Sicotte D. (1994) Expert opinion in the identification of variables for the measurement of early recovery in adults with traumatic brain injury. *Brain Injuries*, **8**:323-34.
29. Alexandrov AV, Pullicino PM, Meslin EM, Norris JW. (1996) Agreement on disease-specific criteria for do-not-resuscitate orders in acute stroke. Members of the Canadian and Western New York Stroke Consortiums. *Stroke*, **27**:232-7.
30. Loughlin KG, Moore LF. (1979) Using Delphi to achieve congruent objectives and activities in a pediatrics department. *Journal of Medical Education*, **54**:101-106.
31. Elliott J. (1980) Consensus on 'rational approach' to therapy of childhood febrile seizures. *JAMA*, **244**:111-112.
32. Grubb NS, Lyle S, Brodie JH, et al. (1995) Management of infants and children 0 to 36 months of age with fever without source. *Journal of American Board of Family Practitioners*, **8**:114-119.
33. Strauss G, Chassin M, Lock J. (1995) Can experts agree when to hospitalize adolescents? *J-Am-Acad-Child-Adolesc-Psychiatry*, **34**:418-424.
34. Caron FJ, van Geijn HP, van Woerden EE, Swartjes JM, Mantel R. (1988) Computerized assessment of fetal behavioral states. *Journal of Perinatal Medicine*, **16**:365-372.
35. Larach MG, Localio AR, Allen GC, et al. (1994) A clinical grading scale to predict malignant hyperthermia susceptibility. *Anesthesiology*, **80**:771-779.
36. Onorato IM, Orenstein WA, Hinman AR, Rogers MF, Koplan JP. (1989) Immunization of asymptomatic HIV-infected children with measles vaccine: assessment of risks and benefits. *Medical Decision Making*, **9**:76-83.
37. Hollender J, Gonnella C, Parker D. (1979) Functional recovery from cancer surgery: estimation of expectations. *Arch-Phys-Med-Rehabil*, **60**:45-49.
38. Koplan JP, Farer LS. (1980) Choice of preventive treatment for isoniazid-resistant tuberculous infection. Use of decision analysis and the Delphi technique. *JAMA*, **244**:2736-2740.
39. Fletcher JW, Woolf SH, Royal HD. (1994) Consensus development for producing diagnostic procedure guidelines: SPECT brain perfusion imaging with exametazime. *Journal of Nuclear Medicine*, **35**:2003-2010.

40. Gershan JA, Freeman CM, Ross MC, et al. (1990) Fluid volume deficit: validating the indicators. The Research Committee, Greater Milwaukee Area Chapter of the American Association of Critical-Care Nurses. *Heart-Lung*, **19**:152-156.
41. Stheeman SE, van 't Hoff MA, Mileman PA, van der Stelt PF. (1995) Use of the Delphi technique to develop standards for quality assessment in diagnostic radiology. *Community-Dent-Health*, **12**:194-199.
42. Lobach DF. (1995) A model for adapting clinical guidelines for electronic implementation in primary care. *Proc Annu Symp Comput Appl Med Care*, 581-585.
43. Philips BU, Jr., Longoria JM, Sturm HV, Delaney FC, Taylor MJ. (1991) Use of a modified Delphi technique to guide coordination of cancer education in Texas medical schools. *Journal of Cancer Education*, **6**:227-233.
44. Piwernetz K, Home PD, Snorgaard O, et al. (1993) Monitoring the targets of the St Vincent Declaration and the implementation of quality management in diabetes care: the DIABCARE initiative. The DIABCARE Monitoring Group of the St Vincent Declaration Steering Committee. *Diabetic Medicine*, **10**:371-377.
45. Altman DG. (1991) Practical Statistics for Medical Research. (London, Chapman and Hall).
46. Fleiss J. (1981) Statistical Methods for Rates and Proportions, (Chichester, Wiley).

Table 1. Subject characteristics.

regime type (n=100)	number of subjects
regime 1	25
regime 2	25
regime 3	25
regime 4	25
sex (n=100)	
male	56
female	44
age (years) (mean \pm SEM)	42.3 \pm 7.2
body mass index (kg/m²) (mean \pm SEM)	28.2 \pm 5.1
diabetes type	
type 1 (IDDM)	21
type 2 (NIDDM)	79
diabetes duration (years) (mean \pm SEM)	8.5 \pm 3.2

Table 2. Round 1 – input sheet.

study case code	001
current insulin regime / units	2 / 12, 0, 8, 0
glucose values (mol/l)	
fasting	8.5
pre-lunch	10.2
pre-dinner	7.3
pre-bed	7.4
clinical symptoms	
fasting	no
pre-lunch	no
pre-dinner	no
pre-bed	no
DIABETES suggestion	
insulin regime / units	2 / 12, 0, 10, 0

Table 3. Round 1 – output sheet.

Specialist code : 4

case no.	specialist answer	agreement with DIABETES	Specialist comment
001	2/12, 0, 8, 0	N	Disagreement comment
002	3/12, 8, 8, 8	Y	agreement comment
003	1/20, 0, 8, 6	N	disagreement comment
004	2/12, 0, 8, 0	Y	agreement comment
005	1/20, 0, 8, 6	N	disagreement comment
...

Table 4. Round 2 – input and output sheet.

case no. 036

initial regime :

1/12, 0, 8, 0

choices : 1st round	frequency	round 1 comments	round 2 comments
1/12, 0, 8, 0	3	comment <1>	
		comment <2>	
		comment <3>	
1/15,0, 12, 0	2	comment <4>	
		comment <5>	
2/15, 8, 8, 8	1	comment <6>	

choice : 2nd round

Table 5. Scoring system : agreement among specialists.

1 st expert	2 nd expert	3 rd expert	4 th expert	5 th expert	points
a	a	a	a	a	10
a	a	a	a	b	6
a	a	a	b	b	4
a	a	a	b	c	3
a	a	b	b	c	2
a	a	b	c	d	1
a	b	c	d	e	0

a,b,c,d,e : different opinions

Table 6. Scoring system : agreement between specialists and DIABETES.

Number of specialists agreeing with DIABETES	points
5	10
4	8
3	6
2	4
1	2
0	0

Table 7. Agreement among specialists and between DIABETES and specialists.

	specialists		DIABETES	
	round 1	round 2	round 1	round 2
specialist 1	129	158	88	84
specialist 2	108	153	112	126
specialist 3	171	189	94	106
specialist 4	145	181	108	110
specialist 5	121	161	120	118
Total	674	842	522	544

Table 8. Additional round 2 points, according to degree of agreement.

degree of agreement	no. of changes	points	%
total agreement	20	80	48
near total agreement	27	54	32
higher agreement	34	34	20
Total	81	168	100