International Journal of Technology Assessment in Health Care, **19:1** (2003), 158–167.

ARE EXPERT PANEL JUDGMENTS OF MEDICAL BENEFITS RELIABLE?

An Evaluation of Emergency Medical Service Programs

Ragnar Hotvedt
University of Tromsø

Hans Morten Lossius
Rogaland Central Hospital

Ival Sønbø Kristiansen
University of Southern Denmark

Petter Andreas Steen
Ullevål Hospital

Eldar Søreide
Rogaland Central Hospital

Olav Helge Førde
University of Tromsø

Abstract

Objective: We have used multidisciplinary expert panels to assess the health benefits from two different emergency medical service programs in Norway. This gave the opportunity to study the reliability of the expert panel method.
Methods: Two panels assessed case reports for 18 children, and two other panels assessed case reports for 64 adult patients. The assessments of each case report were compared. These assessments were also compared with assessments of the same case reports, done by the same panels 1 and 9 years earlier.
Results: Two different panels agreed on the benefit/no benefit conclusion in at least 75% of the patients, both for children and adult patients (kappa 0.88–0.50). For groups of patients assessed to have some health benefit, the magnitude of the benefit estimates differed by 25% between the panels. When the same panels assessed the same patient groups twice, 1 and 9 years apart, their estimates of total benefit differed up to 30%. However, estimates for single patients, as well as estimates from single panel members, varied considerably more.
Conclusions: Use of multidisciplinary expert panels is a useful method for estimating health benefits on program level or for groups of patients. But assessments from single panelists, and for single patients may be seriously biased.

The research has received financial support from the Norwegian Ministry of Health and Social Affairs, the Norwegian Air Ambulance Foundation, and the Laerdal Foundation for Acute Medicine.

Keywords: Reliability, Reproducibility of results, Research designs, Group processes, Emergency medical services

Randomized controlled trials (RCTs) are considered the gold standard for documenting benefit of medical interventions. Recently this conventional wisdom has been challenged by studies that claim well-designed observational studies yield similar evidence (3;5), and in some settings RCTs are not feasible. As one alternative, health benefits may be assessed by observational designs as consensus methods, such as careful external evaluation through multidisciplinary expert panels (12). Expert panel methods have most commonly been used to examine the appropriateness of clinical interventions, in education, and to identify research priorities (11;32).

The documentation of the validity and reliability of expert panels used in such settings is, however, limited. The reliability of the method has been reported to vary considerably. Poor reliability has been found in studies on criteria for emergency cesarean section (2) and assessment of quality of care (8), while fair to good reliability has been demonstrated in studies on appropriateness of coronary angiography (16), coronary revascularization and hysterectomy (31), and on benefit from hospitalization (7).

Emergency care is one of the areas of health care where sufficient evidence from rigorously conducted empirical studies such as RCTs does not exist—or may ever exist, for practical and ethical reasons. Alternatively, expert panel methods have been used in this area, for example, in trauma-related preventable death studies. The reliability of the method has varied from poor (25;30;33) to almost excellent (22;26). In a recent study on medical treatment and transport in prehospital emergency care, Nijs and coworkers (28) found that the agreement among experts was poor.

We have earlier used the expert panel method to evaluate the health benefits from a helicopter emergency medical service (EMS) program in Tromsø, Norway (10). Recently, an almost identical study design was used to assess the benefits from another air and ground EMS program operating in a different region (Stavanger) of Norway (21). This gave us the opportunity to study the reliability of the expert panel method. The aim of the present study was to compare the assessments of the same patients in different panels, to compare the assessments of the same patients at different times in the same panels, and to study the inter- and intrapanelist variation in the assessment of health benefits.

METHODS

Selection of Patients

The present study is a side project from two studies on the health benefits from anesthetist-manned air and ground EMS programs (10;21). The studies were carried out in Tromsø and Stavanger, where the benefits from 370 and 1,106 patient transports were assessed during the study period. In these studies, consulting anesthetists prepared case reports for each transport, and preliminarily grouped them as having potential benefit or not. Expert panels then assessed health benefits for the patients with a potential benefit in a two-round formal nominal group process. The case reports in the two studies included the same key elements of information (time intervals, symptoms, signs, medical interventions, clinical course, etc.), with the Stavanger study presenting the information on a standardized form.

The patients included in the present study were stratified samples from the Tromsø and Stavanger patient populations. From each population, a number of patients judged by the

panels to have no benefit were randomly sampled, with about half of them among patients for whom the panels reached the no-benefit conclusion in the second round. Further, about half of the patients judged to have a benefit were also randomly sampled, ensuring that one-third of them were drawn among those with the highest benefit. Altogether, 82 patients were included in the present study: 34 from Tromsø and 48 from Stavanger. Of these 82 patients, 47 patients were sampled from patients judged by the panels to have a benefit, and 35 from patients judged to have no benefit. The purpose of the stratified sampling procedure was to make sure that we included several patients with debatable benefit from the EMS program, as well as a number of patients who had a major benefit.

Expert Panels

We used the same panels as in the previous studies (10;21): two panels in Tromsø and two in Stavanger. In both studies, one panel assessed possible benefits for children less than 15 years of age and pregnant women, and a second panel assessed possible benefits for all other (adult) patients. The panels for children and pregnant women consisted of an anesthetist, a paediatrician, an obstetrician, a general practitioner, and an epidemiologist. In Stavanger, this panel also included a surgeon. The panels for adults included an anesthetist, a surgeon, an internist, a general practitioner, and an epidemiologist. In Tromsø the two panels had no common members, while in Stavanger the two panels shared four of the members.

Health Benefit Assessments

The panels made their judgments on the basis of the same case reports as in the previous studies. We also provided the panels with the same body of relevant evidence from the literature, asking them to integrate this into their decision-making process. The panels followed the same assessment procedure: in round 1, the panelists individually stated their estimates of benefit for each patient on a form. In round 2, the panels convened, and the panelists revealed their estimates and argued for them. Following the discussion, the panel members stated their final assessments on the form. There was no explicit requirement that the panels should achieve consensus in the second round.

Health benefits from the ambulance programs were estimated as life-years gained; for each patient, remaining life expectancy was obtained from Norwegian life tables (19). We used the same life tables as used in the previous studies (10;21). Life-years gained were calculated by multiplying the remaining life expectancy by the probability (assessed by the panel members) that these life-years were attributable to the anesthetist-manned EMS program, compared with possible outcome from a realistic medical alternative. Evaluation of the realistic alternative was based on actual state of readiness, actual qualifications, and calculated time lags for general practitioners, ground ambulance services, and any other current service to be involved in the care of each individual patient, assuming the studied EMS program was not accessible.

The Tromsø and Stavanger panels assessed all 82 case reports. Thus, both the Tromsø and the Stavanger panels assessed their "own" case reports (34 and 48 case reports, respectively) in the year 2000. Previously, the same panels had evaluated the same case reports. For the Tromsø panels, this took place 9 years earlier (1991) (10), while for the Stavanger panels the time interval was 1 year (1999) (21). When the panels assessed the case reports for the second time, their earlier estimates were concealed.

In addition, the 34 case reports from Tromsø were sent to the Stavanger panels, while the 48 Stavanger case reports conversely were sent to the Tromsø panels. In this way, two different panels evaluated the same $(34 + 48)$ 82 case reports.

Table 1. Patients Assessed by the Panels According to Age and Diagnostic Groups

Diagnostic groups	Adults, 15 + years			Children, 0–14 years		
	n	Female (%)	Mean age (range)	n	Female (%)	Mean age (range)
Infection	1	0	16	5	20	3.6 (0–10)
Cardiovascular disease	36	11	61.8 (32–81)	0		
Neonatal problems	—	—	—	1	0	0.0
Trauma	18	11	36.9 (17–80)	8	38	7.4 (0–12)
Intoxication	4	50	33.5 (20–60)	0		
Respiratory problems	4	50	59.3 (26–79)	2	0	7.0 (1–13)
Other	1	100	75	2	0	8.5 (6–11)
Total	64	17	52.3 (16–81)	18	28	6.0 (0–13)

Statistics

To compare the panels' classification of the patients as having benefit or not, kappa statistics were used. Inter- and intrapanel comparisons of the assessment on individual patients were displayed with simple plots and Bland Altman plots (4).

RESULTS

Included in this study were 82 patients: 25 adults and 9 children from Tromsø, and 39 adults and 9 children from Stavanger. All the Tromsø patients were transported by helicopter. In Stavanger 24 patients were transported to the hospital by helicopter, while 24 went by ground ambulance.

Among the 18 patients aged under 15 years, trauma and infection were the most frequent diagnoses (Table 1). The diagnostic group "Other" included one child with epileptic convulsions and one with febrile convulsions. Among the 64 patients aged 15 and older, cardiovascular disease and trauma dominated.

Interpanel Comparisons

If "benefit" for a patient is defined as a mean life-year gain of more than zero among the panelists, the two panels evaluating adult patients agreed on the benefit/no benefit conclusion in 48 of 64 cases (kappa = 0.50). If "benefit" is defined as a median life-year gain of more than zero, the panels agreed in 50 of 64 cases (kappa = 0.56). For children, the panels agreed in 14 of 18 (kappa = 0.58), and 17 of 18 cases (kappa = 0.89), respectively (Table 2).

For the 64 adult patients, the Tromsø panel estimated the total mean health benefit at 195 (median 192) life-years gained, related to 29 patient transports. The most pessimistic panelist estimated the gain to be 157 life-years (in 25 patients), while the most optimistic panelist estimated 223 years (in 24 patients). For the same 64 patients, the Stavanger panel estimated the total mean benefit to be 245 life-years, i.e., 26% higher than the Tromsø panel, in 31 patients. The Stavanger panel achieved consensus on this assessment. The association between the assessments of the two panels for the individual patients is shown in Figure 1 (Spearman's rho 0.591). The estimates differed considerably for four patients. Two patients with cardiac arrest had a great benefit according to the Tromsø panel, and a smaller benefit according to the Stavanger panel. For two other patients (one poisoning and one traffic accident), the estimates of the Stavanger panel were more optimistic (Figure 1).

The 18 patients less than 15 years of age had a total mean health benefit of 145 (median 156) life-years (among 11 patients), according to the Tromsø panel. The most pessimistic

https://doi.org/10.1017/S0266462303000151 Published online by Cambridge University Press

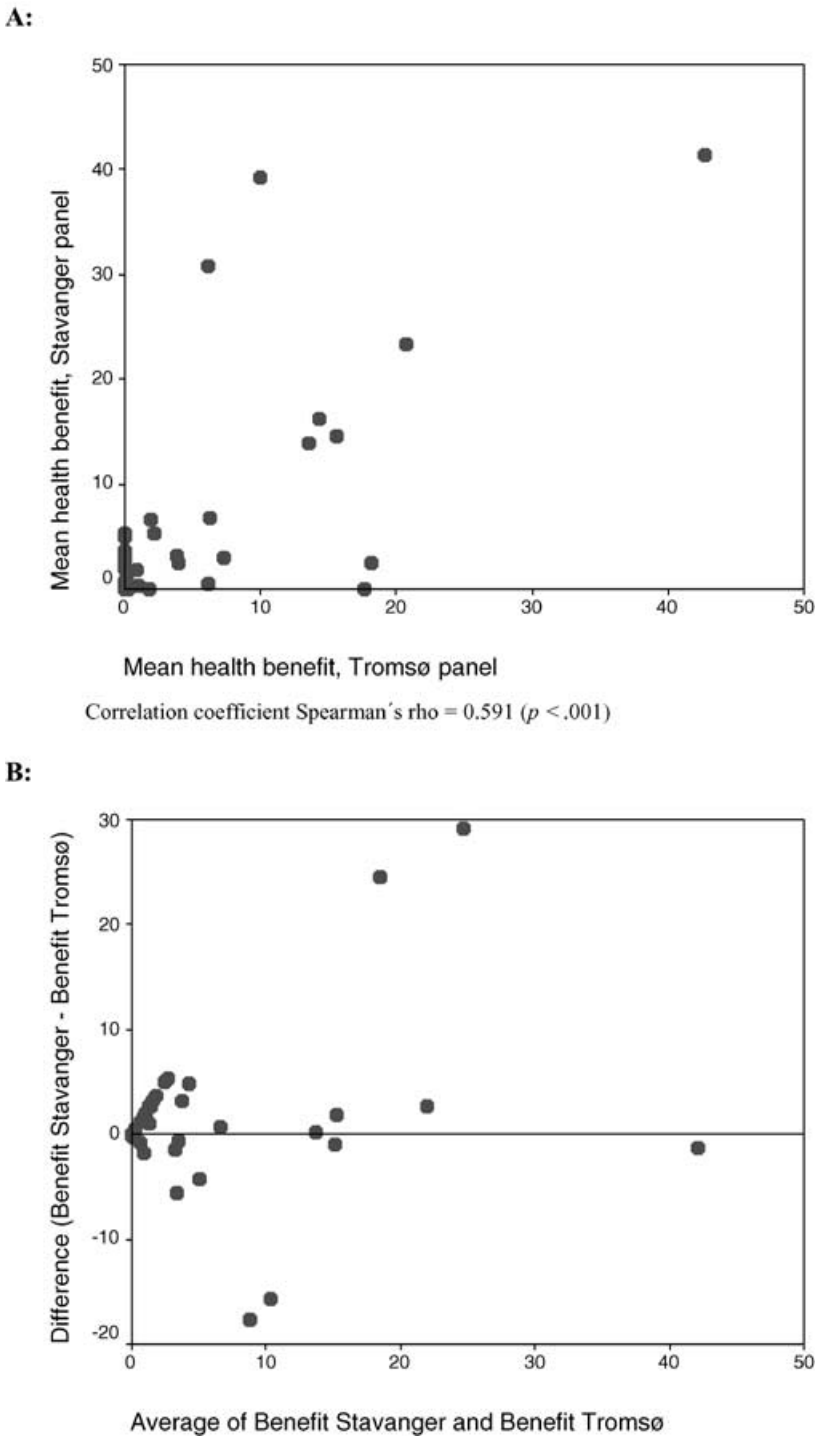

Figure 1. (A) Plot of assessment of mean health benefit (life-years gained) for patients 15 years and older (n = 64) by the two panels. (B) Difference between mean benefit estimates of Stavanger and Tromsø panels (life-years gained) plotted against average (n = 64). Each point represents one patient, except at zero, where it indicates 26 patients.

https://doi.org/10.1017/S0266462303000151 Published online by Cambridge University Press

Table 2. Agreement Between Panels Evaluating Health Benefit for (A) Patients 15 Years and Older (n = 64) and (B) for Patients Less than 15 Years (n = 18)

A Patients 15 years and older (n = 64)

	Stavanger panel		
	Benefit	No benefit	Total
Tromsø panel			
Benefit	22 (20)	7 (3)	29 (23)
No benefit	9 (11)	26 (30)	35 (41)
Total	31 (31)	33 (33)	64 (64)

B Patients less than 15 years (n = 18)

	Stavanger panel		
	Benefit	No benefit	Total
Tromsø panel			
Benefit	7 (7)	4 (1)	11 (8)
No benefit	0 (0)	7 (10)	7 (10)
Total	7 (7)	11 (11)	18 (18)

Benefit is defined as a mean life-year gain of more than 0.0 years across panel members. Numbers in parenthesis indicate results when *benefit* is defined as a median life-year gain of more than 0.0 years.

panelist estimate was 85 life-years (in 8 patients), the most optimistic 207 years (in 9 patients). The Stavanger panel judged unanimously the total mean benefit to be 27% higher, 184 life-years. In four patients (meningococcal sepsis [1], placenta previa [1], and bronchial asthma [2]), the judgments differed markedly.

Intrapanel Comparisons

Twenty-five patients 15 years and older were transported by the EMS helicopter to the hospital in Tromsø. They were evaluated twice by the Tromsø panel, 9 years apart. The panel agreed on the benefit/no benefit conclusions in 1991 and 2000, corresponding to a kappa of 0.60 ("Benefit" for a patient is defined as mean life-year gain > zero among the panelists). In 1991, the panel evaluated the total mean health benefit to be 64 life-years (most pessimistic estimate, 40; most optimistic, 81). Nine years later, the total mean benefit estimate was 78 life-years (range 60–94), representing an increase of 22%. In three patients, the assessments of the panel differed notably on the two points of time.

The Tromsø panel, evaluating patients less than 15 years of age (n = 9), assessed in 1991 that these patients had a total mean health benefit of 160 life-years (range, 127–219). In 2000, the panel estimated the total mean benefit to be 107 life years (range, 66–148), i.e., 33% lower than in 1991. The panel agreed on the benefit/no benefit conclusions corresponding to kappa = 0.77.

The Stavanger panels evaluated patients transported by EMS ground ambulance and helicopter to the hospital in Stavanger on two occasions, in 1999 and 2000. The panel evaluating patients 15 years and older (n = 39), agreed on the benefit/no benefit conclusions corresponding to kappa = 0.38. The panel concluded in 1999 that these patients had a total mean benefit of 213 life years (range, 212–213). One year later, they unanimously concluded that the total mean benefit for the same patient group was 21% lower, i.e., 169 life-years gained. One patient was considered to have a considerably higher benefit in 1999 than in 2000. This was a patient with a head injury, who was intubated at the site of a traffic accident.

The panel evaluating patients less than 15 years of age (n = 9) assessed these patients to have a total mean health benefit of 69 life-years in 1999. One year later, the total estimate was 70 life-years. Kappa value on the benefit/no benefit conclusions was 0.77.

Both Stavanger panels achieved consensus in their evaluation of the patients, with one exception. This was one adult patient in 1999, where one panel member meant that the patient had no benefit, while the other four members unanimously assessed the benefit to be 0.4 life-years.

The Panel Process

The number of patients assumed to have no benefit from the ambulance program increased from round 1 to round 2 for all panels. In 2000, the Tromsø panel evaluating adult patients (n = 64) estimated the mean benefit to be zero for 14 patients in round 1. In round 2, the panel concluded that 35 patients had no benefit. For the same patients, the Stavanger panel estimated no mean benefit for 11 patients in round 1 and for 33 patients in round 2. For children (n = 18), the Tromsø panel stated no benefit for six patients in round 1 and for seven patients in round 2. Corresponding numbers from the Stavanger panel were no benefit for seven patients in round 1 and for 11 patients in round 2. Through the discussions in the panel meetings, the total mean benefit estimates and the range between the estimates of the most pessimistic and the most optimistic panel member were generally reduced from round 1 to round 2.

In 2000, the Tromsø panel assessed the total mean benefit for the 64 adult patients to be 308 life-years (range, 149–628) in round 1. In round 2, the total mean estimate was reduced to 195 life-years (range, 157–223). Corresponding estimates for the same 64 patients for the Stavanger panel were 253 (range, 140–341) in round 1 and 245 (consensus) in round 2. The Tromsø panel evaluating children less than 15 years of age (n = 18) reduced their estimates of total mean benefit from 188 life-years (range, 78–313) in round 1 to 145 (range, 85–207) in round 2. Corresponding estimates from the Stavanger panel was 177 (range, 40–280) in round 1 to 184 (consensus) in round 2.

In 1991, the Tromsø panel evaluating the 25 adult patients reduced their total mean benefit estimate from 106 life-years (44–136) in round 1 to 64 (40–81) in round 2. The panel evaluating children (n = 9) changed their estimates from a total mean of 152 years (103–235) in round 1 to 160 (127–219) in round 2.

Corresponding estimates for the Stavanger panels in 1999 were for adult patients (n = 39) from 206 life-years (101–296) in round 1 to 213 (212–213) in round 2. For children (n = 9), the estimates changed from 78 life-years (12–163) in round 1 to 69 (consensus) in round 2.

Among the members of the panels, the epidemiologist gave the most pessimistic estimates of total benefit. This was the case for both Stavanger panels in 1999 and 2000 and for both Tromsø panels in 1991 and 2000 (except for the children panel in 2000, in which the anesthetist was the most pessimistic). The most optimistic panelist varied. In Tromsø it was the anesthetist in 1991 and the internist in 2000 for adult patients, and the anesthetist (1991) and the obstetrician (2000) for children. In Stavanger (round 1) the most optimistic member of the panel for adult patients was the surgeon in 1999 and the anesthetist in 2000. Among members of the Stavanger panel evaluating children, the obstetrician (1999) and the anesthetist (2000) were the most optimistic.

DISCUSSION

This study indicates that use of expert panels provides reasonably reliable estimates of health benefits in terms of life-years gained. Two different multidisciplinary expert panels agreed on the benefit/no benefit conclusion in at least 75% of the patients, both for adult patients and children (kappa, 0.50–0.88). Concerning the magnitude of the total health benefit for

the patient groups, the differences between the estimates of the panels were about 25%. When the same panels evaluated the same patient groups twice, 9 and 1 year(s) apart, respectively, their estimates of total benefit differed up to about 30%. However, estimates for single patients, as well as estimates from single panel members, varied considerably more.

The reliability of the expert panel method is influenced by the selection of patient cases. A false high reliability probably would have been achieved if the cases were selected among more clear-cut benefit or no benefit cases. To counteract this possibility, we used a stratified sampling procedure to secure that we included several patients for whom it was not obvious whether they benefited or not from the EMS programs.

The panels base their judgments on the information provided to them. The doctors producing them may influence the case reports. More complete information has been shown to increase the reliability of panel judgments (25;26), but may be difficult to achieve for the prehospital setting. In the present study, one of the panels asked for more information for two patients, but unfortunately this was not available. To keep the premises for judgment constant and to improve reliability (20), we also provided the panels with the same relevant research evidence from the literature.

Lack of a higher reliability of expert panel judgments should not be surprising. In the daily work of doctors, clinical disagreement is a major concern. Even skilled clinicians examining patients disagree regarding their findings. They disagree over patients' histories (9) and about physical findings (15), and consequently in their clinical decisions (2). Experienced clinicians recognize this uncertainty, but few are familiar with numerical values of probabilities. This, and the tendency of selecting the extreme values of probabilities, may have added to the variation in benefit estimates (14).

For a few patients, the panels in the present study disagreed considerably about the benefit (Figure 1). These cases were complex, with patients in critical condition with several vital functions threatened. In one, equipment failure occurred; in addition, the realistic alternative way of transport to hospital was not obvious. In two cases, the panelists disagreed whether ambulance paramedic personnel would have managed complications that probably would have happened during the alternative transport.

We studied a helicopter and car ambulance program transporting an unselected patient group. This raises the question of who is "expert" on inhomogeneous medical problems under discussion (27;29). In the panels, we included experienced clinicians from several specialties, as well as epidemiologists with background in general practice, which made more expertise available to the panelists through the discussions. This contributed to increased agreement among the panels, and reduced the magnitude and the range of the benefit estimates from round 1 to round 2, in accordance with what is shown by other authors (24;26;33).

There is clearly a potential for bias in the selection of participants for the panels. On one hand, it has been argued that the panels should include legitimate expert stakeholders to increase "ownership" of the research and to increase the probability of the results influencing clinical practice and policy (32). On the other hand, it has been claimed that the panels should be composed of members "outside" the system under study (23). In studies of appropriateness, specialists seem to favor their own procedures, and performers give higher appropriate ratings for procedures than nonperformers (1;6;13;17;18). The principle of members from outside the system was violated for one panel member (an anesthetist) in one of the Tromsø panels, and for one panel member in both Stavanger panels (an anesthetist). These members were among the most optimistic raters of benefit, possibly relying on personal knowledge about a familiar healthcare system. On the other hand, the epidemiologists were generally the most pessimistic raters in all the panels. Probably they are more used to evaluating matters in a more critical and distanced way.

The validity of expert panel methods has rightfully been much debated. A "fair" reliability is not at all any guarantee for a corresponding validity of the benefit estimates. Through the process, the panel members may conform more to group norms than expressing atypical opinions (14). There is a danger of deriving collective ignorance rather than wisdom (12) or, as in our case, collective overoptimism or nihilism with respect to the level of benefit. As demonstrated in the present study, the group process leads to lower estimates of benefit. We are fairly convinced that this process brings the group estimates closer to the truth, because the more pessimistic primary estimates very often are linked to the group members with the highest professional competence in the particular cases. However, the accuracy of expert panels in assessing health benefit can only be demonstrated in an RCT. In our study, there was no formal pressure to achieve consensus of estimates in round 2. In spite of this, the estimates of both Stavanger panels—for all patients, with one exception—converged from a considerable divergence in round 1 to consensus in round 2.

Although not optimal, we conclude that use of multidisciplinary expert panels is a useful method to study health benefits on a program level or for groups of patients in settings where RCTs or observational studies are not feasible. However, evaluations for single patients, as well as estimates from single raters, may be seriously biased.

REFERENCES

1. Ayanian JZ, Landrum MB, Normand S-LT, Guada E, McNeil BJ. Rating the appropriateness of coronary angiography: Do practicing physicians agree with an expert panel and each other? *N Engl J Med.* 1998;338:1896-1904.
2. Barretts JFR, Jarvis GJ, MacDonald HN, et al. Inconsistencies in clinical decisions in obstetrics. *Lancet.* 1990;336:549-551.
3. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342:1878-1886.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between methods of clinical measurement. *Lancet.* 1986;February 8:307-310.
5. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887-1892.
6. Coulter I, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: A comparison of a multi- and single disciplinary panel. *Health Serv Res.* 1995;30:577-591.
7. Eriksen BO, Almdahl SM, Hensrud A, et al. Assessing health benefit from hospitalization: Agreement between expert panels. *Int J Technol Assess Health Care.* 1996;12:126-135.
8. Goldman RL. The reliability of peer assessments of quality of care. *JAMA.* 1992;267:958-960.
9. Hall R, Horrokcs JC, Clamp SE, DeDombal FT. Observer variation in assessment of results of surgery for peptic ulceration. *BMJ.* 1976;1:814-816 .
10. Hotvedt R, Kristiansen IS, Førde OH, et al. Which groups of patients benefit from helicopter evacuation? *Lancet.* 1996;347:1362-1366.
11. Jones J, Hunter D. Qualitative research: Consensus methods for medical and health services research. *BMJ.* 1995;311:376-380.
12. Jones J, Hunter D. Using the Delphi and nominal group technique in health services research. In: Pope C, Mays N, eds. *Qualitative research in health care.* 2nd ed. London: BMJ Books, BMJ Publishing Group; 1999:40-49.
13. Kahan JP, Park RE, Leape LL, et al. Variations by speciality in physicians ratings of the appropriateness and necessity of indications for procedures. *Med Care.* 1996;34:512-523.
14. Kahneman D, Slovic P, Tversky A. *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press; 1982.
15. Koran LM. The reliability of clinical methods, data and judgment. *N Engl J Med.* 1975;293:642-646.
16. Landrum MB, McNeil BJ, Silva L, Normand SL. Understanding variability in physician ratings of the appropriateness of coronary angiography after acute myocardial infarction. *J Clin Epidemiol.* 1999;52:309-319.

17. Leape LL, Hilborne LH, Schwartz JS, et al. The appropriateness of coronary artery bypass graft surgery in academic medical centers: Working Group of the Appropriateness Project of the Academic Medical Center Consortium. *Ann Intern Med.* 1996;125:8-18.
18. Leape LL, Park RE, Kahan JP, Brook RH. Group judgments of appropriateness: The effect of panel composition. *Qual Assur Health Care.* 1992;4:151-159.
19. Life tables, 1983–84. In: *Statistical yearbook 1986*. Oslo-Kongsvinger: Central Bureau of Statistics of Norway; 1986.
20. Lomas J, Anderson G, Enkin M, et al. The role of evidence in the consensus process. *JAMA*. 1988;259:3001-3005.
21. Lossius HM, Søreide E, Hotvedt R, et al. Pre hospital advanced life support provided by specially trained physicians: is there a benefit in terms of life-years gained? *Acta Anaesthesiol Scand*. 2002;46:771-778.
22. Maio RF, Burney RE, Gregor MA, Baranski MG. A study of preventable trauma mortality in rural Michigan. *J Trauma*. 1996;41:83-90.
23. Mann CN, Mullins RJ, MacKenzie EJ, Jurkovich GJ, Mock CN. Systematic review of published evidence regarding trauma system effectiveness. *J Trauma*. 1999;47:S25-S33.
24. MacKenzie EJ. Review of evidence regarding trauma system effectiveness resulting from panel studies. *J Trauma*. 1999;47:S34-S41.
25. MacKenzie EJ, Steinwachs DM, Bone LR, Floccare DJ, Ramzy AI. Inter-rater reliability of preventable death judgments. *J Trauma.* 1992;33:292-303.
26. McDermott FT, Cordner SM, Tremayne AB. Reproducibility of preventable death judgments and problem identification in 60 consecutive road trauma fatalities in Victoria, Australia. *J Trauma*. 1997;43:831-839.
27. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998;2:1-83.
28. Nijs HGT, Bleeker JK, Van Der Waal MAE, Casparie AF. Need for consensus development in prehospital emergency medicine: Effect of an expert panel approach. *Eur J Emerg Med.* 1998;5:329-334.
29. Rowe G, Wright G, Bolger F. Delphi: A re-evaluation of research and theory. *Technological Forecasting and Social Change.* 1991;39:235-251.
30. Sampalis JS, Boukas S, Nikolis A, Lavoie A. Preventable death classification: Interrater reliability and comparison with ISS-based survival probability estimates. *Accid Annal Prev.* 1995;27:199-206.
31. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med.* 1998;338:1888-1895.
32. Vella K, Golfrad C, Rowan K, Bion J, Black N. Use of consensus development to establish national research priorities in critical care. *BMJ.* 2000;320:976-980.
33. Wilson DS, McElligot J, Fielding LP. Identification of preventable trauma deaths: Confounded Inquiries? *J Trauma*. 1992;32:45-51.