

Development of appropriateness criteria for colonoscopy: comparison between a standardized expert panel and an evidence-based medicine approach

ANNE NICOLLIER-FAHRNI¹, JOHN-PAUL VADER¹, FLORIAN FROEHLICH², JEAN-JACQUES GONVERS² AND BERNARD BURNAND¹

¹Institute of Social and Preventive Medicine, University of Lausanne, ²Medical Outpatient Clinic, University Hospital, Lausanne, Switzerland

Abstract

Objectives. To assess the degree of agreement between appropriateness criteria for the use of colonoscopy developed by a standardized expert panel method and evidence from published studies.

Design. Descriptive, agreement study.

Setting. Multidisciplinary panel; primary care practice in Switzerland.

Participants. Nine national experts; 577 primary care patients referred for colonoscopy; 154 published papers.

Interventions. Evaluation of the appropriateness of 402 possible clinical indications for colonoscopy, based on a comprehensive review of the literature.

Main outcomes measures. Proportion of agreement (weighted kappa), between panel- and literature-based appropriateness categories (appropriate, uncertain, inappropriate) for theoretical and actual indications encountered.

Results. Nineteen of 402 indications rated by the panel could be based on the evidence retrieved from eight randomized clinical trials. A 68% agreement (kappa = 0.52) was found between panel- and study-based criteria. The addition of an uncontrolled trial and seven observational studies yielded a 71% agreement (kappa = 0.63). Agreement was similar when examining 577 actual cases: 69% agreement, kappa = 0.47. Agreement between panel-based indications and published evidence was not influenced by the perceived comprehensiveness and the apparent quality of the published reports.

Conclusions. Evidence for the appropriateness of most indications for colonoscopy could not be derived directly from the published literature. Agreement between appropriateness criteria developed by an expert panel and evidence from published studies was moderate to good, where available. New approaches should be sought in order to systematically integrate complementary evidence obtained from clinical trials and expert panels into practice guidelines.

Keywords: appropriateness, clinical trials, colon diseases, literature review, RAND appropriateness method, recommendation, standardized expert panel method

Health care systems in many countries are undergoing dramatic and rapid transformations in an effort to control health care expenditures. Simultaneously, efforts to improve quality of care have been aimed at prioritizing evidence-based procedures. However, a high level of evidence, typically obtained from randomized controlled clinical trials, does not exist for from 50 to 90% of medical procedures [1–4]. Even the

clinical guidelines for medical procedures based on evidence from clinical trials may vary substantially [5]. Whatever scientific evidence for a specific medical procedure already exists, determining the appropriateness of use of health care interventions will always depend on human interpretation of the—at the time potentially contradictory—literature that is available.

Address reprint requests to Bernard Burnand, IUMSP, 17 Bugnon, CH-1005 Lausanne, Switzerland.
E-mail: bernard.burnand@inst.hospvd.ch

The RAND appropriateness method (RAM) [6] combines a detailed review of the literature with an expert opinion group process. The RAM could thus represent one way of integrating clinical expertise and more formal clinical evidence to aid physicians and patients in making decisions about individual care. Even though it is one of the most widely accepted methods, the RAM has been criticized on various grounds [7,8]. In particular, questions have been raised about the unknown extent to which it represents the best available evidence from published clinical trials, or whether it simply reproduces current practice.

This study examined the level of agreement between clinical evidence from published clinical trials and the recommendations of an expert panel, which had developed appropriateness criteria for the use of colonoscopy using the RAM.

Materials and methods

In October 1994, a national multidisciplinary panel of nine experts from Switzerland used a standardized expert panel method, RAM [6], to develop criteria for the appropriateness of performing colonoscopy. The panel was composed of nationally recognized authorities from relevant specialties: five gastroenterologists, two internists, one general practitioner and one surgeon. Four out of the five gastroenterologists were actually very experienced in performance of the procedure. An inventory of all potential specific clinical scenarios (indications) ($n = 402$) for which performance of colonoscopy might be considered was prepared. This inventory was based on an extensive literature review which included articles published up to 1993. The descriptions of the indications were sufficiently detailed so that patients presenting with a particular indication were reasonably homogeneous, in the sense that performing the procedure would be equally appropriate or inappropriate for all patients in that group.

The literature review was prepared on the basis of a Medline search covering the period 1976–1993. The keywords used were: ‘gastrointestinal endoscopy and gastrointestinal diseases’ crossed with ‘efficacy, outcomes, clinical trials, meta-analysis and complications’. A search was also performed on the keywords ‘inflammatory bowel diseases, rectal bleeding, gastrointestinal hemorrhage, colorectal polyps, abdominal pain, anemia, occult blood, diarrhoea’ associated with ‘surveillance’ or ‘screening’ and crossed with ‘gastrointestinal endoscopy’. In addition, relevant studies were proposed by the panel of experts themselves. Priority was given to clinical trials and observational studies with the aim of identifying the best evidence linking performance of colonoscopy to achievement of a favorable outcome. For this study, 154 articles related to the use of colonoscopy were included, out of over 1500 articles considered.

The major categories of indications are shown in the first column of Table 1. Other elements used to render the indications clinically specific included: patient age, symptoms, risk factors for colorectal cancer, previous lower gastrointestinal evaluations and treatments, and previously

documented pathology of the gastrointestinal system. Examples of detailed clinical scenarios are presented in Table 4.

Each panelist was provided with the literature review and the catalogue of indications. The nine experts were asked to rate the appropriateness of each indication on a 9-point scale, ranging from 1 (extremely inappropriate) to 9 (extremely appropriate). The following definition of appropriateness of an indication was used by the experts: the indication to perform a medical procedure is appropriate when the expected health benefit (for example, increased life expectancy, pain relief, reduction in anxiety, improved functional capacity) exceeds the possible negative consequences (for example, mortality, morbidity, anxiety of anticipation of the procedure, pain produced by the procedure) by a sufficiently wide margin that the procedure is worth doing. After an initial rating performed at home, the panelists then convened and were provided with reports showing their own initial ratings and the anonymous distribution of other panelists’ ratings. Indications were discussed in depth and panelists then individually re-rated all indications. The 9-point scale was consolidated into three categories (inappropriate, uncertain, appropriate) by using the median rating and the degree of agreement among the panelists. The indication for colonoscopy was considered appropriate when the median was between 7 and 9 without any disagreement and inappropriate if the median was between 1 and 3 without any disagreement. The indications were categorized as uncertain if the median was between 4 and 6 or if panel members disagreed. For such a panel of nine experts, disagreement was defined as occurring when at least two members rated an indication in the 1 to 3 range and at least two others in the 7 to 9 range.

The design of each study quoted in the literature review was examined by two reviewers (A.N.-F. and B.B.). In case of disagreement between the two reviewers, a consensus decision was achieved after a re-reading of the study and discussion. Clinical trials, observational cohort and case-control studies, as well as studies of the accuracy of colonoscopy were further analysed to identify which indications for colonoscopy were included in each study. When appropriate, results of studies that had examined the use of sigmoidoscopy (in general the older studies) were extrapolated to the use of colonoscopy. However, this approach was used only when one could logically project that the outcome of the use of sigmoidoscopy would also apply to the use of colonoscopy. As it turns out, this was basically done for colorectal cancer screening studies.

In order to match the indications rated by the panel and those retrieved from the literature, we used the inclusion and exclusion criteria of the published studies and patient characteristics (for example, age, gender). Matching was considered possible whenever the profile of the majority of the patients in the study, or a clearly identified subgroup, corresponded to a specific indication rated by the panel. In several instances a perfect match between the clinical scenarios and patient profile in the corresponding study was possible. Most often, matching could be made with groupings of indications or after applying some form of logical deduction

Table 1 Categories and number of indications for colonoscopy developed by a Swiss panel; number of comparable indications identified in published studies; and appropriateness ratings for the panel indications and published indications categories (for the purpose of the appropriateness comparisons, all 48 relevant indications from the literature were used)

Category of indications (Swiss panel)	Number of indications evaluated by the panel	Comparable published indications Clinical studies ($n = 16$) ² RCTs ($n = 8$) ²	Appropriateness categories ¹					
			Swiss panel			Relevant studies		
			A	U	I	A	U	I
Unexplained positive faecal occult blood test or iron-deficiency anaemia	52	4	2	2	0	2	0	2
Haematochezia	22	4	3	0	1	2	2	0
Uncomplicated lower abdominal pain or change in bowel habits	104	0	-	-	-	-	-	-
Diarrhoea of 3 weeks or more duration	22	0	-	-	-	-	-	-
Known inflammatory bowel disease	34	3	0	0	3	0	0	3
Surveillance after colonoscopic polypectomy	15	12	6	1	5	8	0	4
Screening for colorectal cancer in asymptomatic individuals	48	10	4	3	3	6	1	3
Screening for colorectal cancer in patients with ulcerative colitis	31	15	9	4	2	9	0	6
Screening for colorectal cancer in patients with Crohn's disease or asymptomatic patients with history of colorectal cancer	27	0	-	-	-	-	-	-
Miscellaneous (e.g. foreign body removal, palliative treatment of stenosing neoplasm)	47	0	-	-	-	-	-	-
Total	402	48	24	10	14	27	3	18

A, appropriate; U, uncertain; I, Inappropriate; RCT, randomized controlled trial.

¹Dashes indicate where no study was available for this category of indications.

² ≥ 1 match with the panel indications.

Table 2 Concordance between the Swiss panel’s appropriateness ratings and the ratings reported in the randomized controlled trials for 19 indications for colonoscopy

Expert panel ¹	Randomized controlled trials ²			
	Appropriate	Uncertain	Inappropriate	Total
Appropriate	8	1	1	10
Uncertain	0	0	1	1
Inappropriate	2	1	5	8
Total	10	2	7	19
68% agreement; kappa = 0.52 (0.17–0.86) ³				

¹Swiss Panel on the Appropriateness of Use of Colonoscopy, 1994. Experts’ opinion based on knowledge, experience, and appreciation of published evidence reported in the literature review, according to the RAND appropriateness method.

²Evidence from the eight randomized controlled trials reviewed for this study.

³95% Confidence Interval.

Table 3 Concordance between the Swiss panel’s appropriateness ratings and the ratings reported in published clinical trials and observational studies for 48 indications for colonoscopy

Expert panel ¹	Published studies ²			
	Appropriate	Uncertain	Inappropriate	Total
Appropriate	22	1	1	24
Uncertain	3	1	6	10
Inappropriate	2	1	11	14
Total	27	3	18	48
71% agreement; kappa = 0.63 (0.45–0.80) ³				

¹Swiss Panel on the Appropriateness of Use of Colonoscopy, 1994 (see Table 2).

²Eight randomized controlled trials, one non-randomized trial, five cohort and two case-control studies.

³95% Confidence Interval.

(for example, if according to a study’s results it is appropriate to perform a colonoscopy every 3–5 years in a non-symptomatic patient aged ≥ 40 years, it would also be deemed appropriate if the previous colonoscopy had been performed >5 years previously). Categorization into appropriate, uncertain and inappropriate indications was also made for indications retrieved from the selected studies, based on the results and the discussion section of the published report. Appropriate meant that outcome was better when colonoscopy was performed; inappropriate was used in the absence of benefit or because of the predominance of adverse effects. The category uncertain was used when the results themselves of the study under evaluation were considered to be uncertain or in case of disagreement of results between two or more studies of the same indication. For this last category, we did not attribute a different weight to the various studies according to design. Only clinical scenarios that could be matched with a clinical trial, an observational study or a study of the accuracy of colonoscopy were considered for comparative analysis.

The quality of the studies was neither formally assessed nor used later in the analyses, as only studies of sufficient quality were included in the literature review. However, in

order to test whether the degree of agreement between panel- and literature-based appropriateness criteria corresponded to the apparent quality of the published trials, we stratified according to comprehensiveness and quality, using the 21 items of the CONSORT Statement [9] and the criteria proposed by Jadad *et al.* [10]. Given the small number of relevant publications available, only the CONSORT criteria allowed us to construct meaningful strata for this analysis. The CONSORT items were used independently by two reviewers (A.N.-F. and B.B.) with good agreement between them (kappa = 0.71). A consensus was easily arrived at after examination of cases where there was disagreement. A simple score was calculated which allowed us to dichotomize indications derived from published trials with a higher versus lower apparent quality.

In additional analyses, the literature-based indications were compared to 430 indications rated by a panel with a similar composition, held in March 1994 in the US. The Swiss panel did in fact use an identical catalogue of indications to the US panel as well as the same review of the literature. In addition, because in actual practice some of the indications considered by the panel are much more frequent than others, differing results in the comparison between appropriateness-

Table 4. Details of indications for which there was lack of agreement between the panel- and literature-based appropriateness ratings for matched indications

Clinical indication	Appropriateness	
	Expert panel	Published studies ¹
Surveillance colonoscopy after colonoscopic polypectomy in patients without inflammatory bowel disease, familial polyposis, or a personal history of colorectal cancer		
For a single small (<1 cm) tubular adenoma without high-grade dysplasia; interval since last colonoscopy >3 years (two indications)	Inappropriate	Appropriate ²
For any adenoma with high-grade dysplasia, or villous or tubulo-villous histology; interval since last colonoscopy <3 years	Appropriate	Inappropriate ²
Colonoscopy for positive faecal occult blood test in young asymptomatic patients (<50) without risk factors for colorectal cancer, and no known polyps or inflammatory bowel disease		
In absence of other lower gastrointestinal examination	Uncertain	Inappropriate ³
Colonoscopy in young patients (<50) with haematochaezia, but without risk factors for colorectal cancer, without haemodynamic compromise or known inflammatory bowel disease, and no evaluation done		
Presumed anorectal source	Inappropriate	Uncertain ⁴
Not presumed anorectal source	Appropriate	Uncertain ⁴

¹Indications from randomized controlled trials only.

²As reported by Winawer *et al.* [11].

³As reported by Kronborg *et al.* [15].

⁴As reported by Rex *et al.* [17].

and literature-based criteria may occur depending on whether they are based only on theoretical or real-patient cases. We therefore also considered the indications for colonoscopy encountered in real patients referred for colonoscopy in a primary care setting ($n = 577$) by 22 general practitioners.

The proportion of agreement and weighted kappa coefficients were calculated for both theoretical indications and those encountered in practice. A bootstrap approach was used for the calculation of 95% confidence intervals for kappa. These analyses were conducted for all relevant studies, and separately for the randomized controlled trials (RCTs). We used Stata 7.0 (Stata Corporation, USA) for statistical analyses.

Results

The literature review comprised 154 published papers, including 11 clinical trials, 14 observational studies, and 11 studies of the accuracy of colonoscopy and sigmoidoscopy. Forty-eight (12%) of the 402 specific indications for colonoscopy rated by the Swiss panel could be matched to a corresponding indication from at least one of 16 published studies (eight RCTs [11–18], one non-randomized trial [19], five cohort studies [20–24] and two case-control studies [25,26]) (Table 1). Nineteen specific indications could be identified in at least one RCT. A 68% agreement was found between appropriateness criteria and criteria derived from RCTs (kappa = 0.52) (Table 2). A similar degree of agreement (71%

agreement, kappa = 0.63) was found for the 48 indications, which could be compared to all 16 relevant studies (Table 3). The comparison of indications from the studies using the appropriateness criteria developed by the US panel led to similar results; the proportion of agreement was 78% (kappa = 0.67) for the 18 indications identified in the RCTs and 79% (kappa = 0.74) for 47 indications for all the relevant studies. The indications for which there was a lack of agreement between the RCTs and the panel are presented in detail in Table 4.

The degree of agreement between the indications rated by the Swiss panel and those retrieved from RCTs did not differ when based on the 577 real cases. Fifty-one cases (9%) could be compared using the two sets of criteria (12 indications). The proportion of agreement between panel- and literature-based indications was 69% (kappa = 0.47).

The degree of agreement between the two sets of indications was not different when the comprehensiveness and the apparent quality of the published reports were considered. The proportion of agreement was 77% (kappa = 0.54) for 13 indications retrieved from the three trials with higher CONSORT scores and 73% (kappa = 0.48) for 11 indications from the other five trials.

Discussion

This study indicates that, when considering all possible indications for the use of colonoscopy, the rationale for its

utilization could be based on evidence from published randomized clinical trials in only 5% of cases. The proportion was still <10% when calculated from indications to perform colonoscopy in actual primary care patients referred for the procedure. The inclusion of observational studies, which are in general more vulnerable to bias and error, did not significantly increase the number of indications which could be based on that additional evidence. These results alone underscore the need to further increase the evidence base for many clinical indications for colonoscopy. Although it is not easy to conduct effectiveness studies, it is imperative that physicians and patients be able to rely on more solid information to help them make informed decisions about care. It is also important to have such information to improve use of resources in health care. However, one should also realistically acknowledge that we would never have at our disposal all the high quality studies that we would like to have. Therefore, we should also improve alternative approaches to determining the appropriateness of care, such as standardized expert panel methods.

Evidence-based medicine (EBM) seeks to formulate relevant questions and propose answers for individual patient care [27,28]. Although it does not follow the rules developed by the EBM movement, the catalogue of clinical scenarios used in the expert panel process could be considered as the sum of all the possible indications for which colonoscopy could be indicated. Thus, a question could be formulated based on each clinical scenario to look for published evidence to evaluate the degree of appropriateness of use of colonoscopy in such patients. For the vast majority of these indications, no RCTs—nor any studies of sufficient quality—exist and the results of an expert panel may well represent ‘the next best external evidence’ [28].

We acknowledge, however, that this analysis is based on a literature review and a RAM panel that took place several years ago. A limited overview of more recently published studies of the potential usefulness of diagnostic colonoscopy was performed. A few prospective cohort studies and clinical trials have been published since this RAM panel. Several studies provided updated information for indications for which information already existed in 1994 (for example, faecal occult blood test for colorectal cancer screening, post-polypectomy follow up), but not much has been published about the usefulness of colonoscopy in patients with abdominal pain or changes in bowel habits, for instance. Therefore, we feel that, had the analysis been conducted anew more recently, we would not expect much change in our findings and conclusions.

The low proportion of indications based on published evidence is explained in part by the large number of clinically specific indications for appropriate use of colonoscopy developed by the panel process. Although one of the purposes of the RAM is to obtain clinically specific and meaningful indications, the catalogue of all potential indications for colonoscopy was built with the aim of being virtually comprehensive. Actually, the indications, which would a priori be considered inappropriate by a large majority of clinicians was also included. In addition, for one RCT and eight

observational studies, matching indications between the panel and the literature was not possible. This was either because the study population included heterogeneous groups of patients with different susceptibility to disease (for example, colorectal cancer), or because information was lacking regarding inclusion and exclusion criteria or subgroup analyses. The difficulty encountered in using these studies as a source of clinical evidence for one specific clinical scenario would probably be similar to that encountered by a physician looking for evidence on the care of an individual patient. In the application of EBM in daily practice, such interpretation of published evidence is however sometimes possible. Another reason for the low number of literature-based indications was the fact that published studies providing lower-quality evidence (for example, small case series) were not taken into consideration in this analysis.

Based on kappa, the agreement between literature- and panel-based indications was moderate, despite proportions of agreement, which were generally in excess of 70%, a problem due to small numbers and imbalances in the agreement table [29]. This less-than-optimal agreement could indicate that the panelists paid insufficient attention to published evidence summarized in the literature review when rating appropriateness of indications. Alternatively, although they were aware of the results of the published trials, they could have preferred their own clinical expertise in certain situations, as proposed by the EBM approach [28]. Additional elements other than the results of published trials could have influenced their judgement; for instance, a perceived low effectiveness of the procedure in spite of proven efficacy or the apparent poor quality of some published studies. Furthermore, it is probable that panel members, although experts in their own clinical field, may not master all the skills and knowledge necessary for the critical appraisal of evidence from the literature. Other potential reasons for lack of agreement between experts’ judgement and published evidence may include time-lag between the realization of a study and judgement concerning its role in current practice or the rate of diffusion and frequency of use of a technology in a country. Last but not least, the definition of the appropriateness categories was different for the literature- and the panel-based indications, especially for the ‘uncertain’ category, thus precluding full agreement.

In their study of the development of RAM clinical indications for carotid endarterectomy, Merrick *et al.* found, in a simple analysis, that panel ratings were consistent with those of the literature [30]. However, comparing the same RAM clinical indications with the results of research synthesis, produced by a systematic quantitative summary of research findings, Wortman *et al.* observed that although there were areas of agreement between the two methods, differences occurred which were considered to be significant [31]. Lomas *et al.* examined the role of evidence in a consensus conference process in Canada. They found that, for evidence-based scenarios, as opposed to non-evidence-based scenarios, consensus was substantially greater among panelists before the conference and that improvement in consensus after the conference was also much higher [32].

Many of the hypotheses formulated above about the lack of full agreement between panelists' judgement and the evidence retrieved from the literature are speculative and the reasons therefore remain largely unknown. This uncertainty about the degree of agreement between panel judgement and published evidence has been regarded as a weakness of the RAM [7]. Indeed, although the method implies the examination of specific indications for a medical procedure, the role, place and significance of published evidence is not currently systematically introduced into the process, beyond provision of the literature review to the panelists, and frequent reference to published studies by moderators and experts at the panel meeting. Choosing, summarizing and presenting scientific evidence in a way in which it can be assimilated and explicitly used is a key component in the process of developing guidelines [33]. This task is however difficult, as illustrated by the use of randomized trials and meta-analyses in developing guidelines in England [34].

In conclusion, only a small proportion of indications for colonoscopy could be linked to published clinical evidence. This underscores the limits of relying solely on an evidence-based approach, in the strict sense of the term. Evidence is poor for many health care interventions, and there is uncertainty about which diagnostic or therapeutic approach to use. In these situations, there is a need for a complementary approach such as a standardized panel method in the development of recommendations to assist clinicians and patients in decision-making. Where evidence did exist, there was a lack of complete agreement between experts and published evidence of ~20–30%. This lack of full agreement between literature- and panel-based indications should represent the object of future studies. Further improvements in the process of developing appropriateness criteria and related practice guidelines should better combine published evidence and complementary judgement from multidisciplinary expert panels. In particular, existing published evidence should be more systematically presented and integrated into the process, thus providing a more powerful tool to obtain both evidence-based, as far as are available, and clinically specific guidelines for the appropriate use of medical procedures.

Acknowledgements

We thank Robert W. Dubois MD PhD, currently at Zynx Health, a Cerner Corporation, USA, for sharing with us the results of the 1994 USA Appropriateness Panel on Colonoscopy. We thank Guy van Melle PhD, of the Institute of Social and Preventive Medicine, University of Lausanne, for statistical assistance. This study was supported by the Swiss National Science Foundation (grant number 32-40522.94), the Swiss National Science Foundation (grant number 32-057244.99), the Swiss Federal Office of Education and Science (grant number 95.0306.2), the Stanley Thomas Johnson Foundation (Bern), the Sandoz Research Foundation (Basle), and the Helmut Horten Foundation (Madonna del Piano).

References

1. Smith R. Where is the wisdom...? The poverty of medical evidence. *Br Med J* 1991; **303**: 798–799.
2. Brook RH. Quality of care: do we care? *Ann Intern Med* 1991; **115**: 486–490.
3. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet* 1995; **346**: 407–410.
4. Gill P, Dowell AC, Neal RD, Smith N, Heywood P, Wilson AE. Evidence based general practice: a retrospective study of interventions in one training practice. *Br Med J* 1996; **312**: 819–821.
5. Fahey TP, Peters TJ. What constitutes controlled hypertension? Patient based comparison of hypertension guidelines. *Br Med J* 1996; **313**: 93–96.
6. Brook RH, Chassin MR, Fink A, Solomon DH, Koseoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986; **2**: 53–63.
7. Hicks NR. Some observations on attempts to measure appropriateness of care. *Br Med J* 1994; **309**: 730–733.
8. Phelps CE. The methodologic foundations of studies of the appropriateness of medical care. *N Engl J Med* 1993; **329**: 1241–1245.
9. Begg C, Cho M, Eastwood S *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT Statement. *J Am Med Assoc* 1996; **276**: 637–639.
10. Jadad AR, Moore RA, Carroll D *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Contr Clin Trials* 1996; **17**: 1–12.
11. Winawer SJ, Zauber AG, O'Brien MJ *et al.* Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. *N Engl J Med* 1993; **328**: 901–906.
12. Landi B, N'Guyen Anh T, Cortot A *et al.* Endoscopic monitoring of Crohn's disease treatment: a prospective, randomised clinical trial. *Gastroenterology* 1992; **102**: 1647–1653.
13. Hardcastle JD, Thomas WM, Chamberlain J *et al.* Randomised, controlled trial of faecal occult blood screening for colorectal cancer. Results for first 107 349 subjects. *Lancet* 1989; **1**: 1160–1164.
14. Kewenter J, Björk S, Haglind E, Smith L, Svanik J, Ahrén C. Screening and rescreening for colorectal cancer. A controlled trial of fecal occult blood testing in 27,700 subjects. *Cancer* 1988; **62**: 645–651.
15. Kronborg O, Fenger C, Olsen J, Bech K, Sondergaard O. Repeated screening for colorectal cancer with fecal occult blood test. A prospective randomized study at Funen, Denmark. *Scand J Gastroenterol* 1989; **24**: 599–606.
16. Mandel JS, Bond JH, Church TR *et al.* Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993; **328**: 1365–1371.
17. Rex DK, Weddle RA, Lehman GA *et al.* Flexible sigmoidoscopy plus air contrast barium enema versus colonoscopy for suspected

- lower gastrointestinal bleeding. *Gastroenterology* 1990; **98**: 855–861.
18. Hardcastle JD, Armitage NC, Chamberlain J, Amar SS, James PD, Balfour TW. Fecal occult blood screening for colorectal cancer in the general population. *Cancer* 1986; **58**: 397–403.
 19. Winawer SJ, Flehinger BJ, Schottenfeld D, Miller DG. Screening for colorectal cancer with fecal occult blood testing and sigmoidoscopy. *J Natl Cancer Inst* 1993; **85**: 1311–1318.
 20. Olsen HW, Lawrence WA, Snook CW, Mutch WM. Review of recurrent polyps and cancer in 500 patients with initial colonoscopy for polyps. *Dis Colon Rectum* 1988; **31**: 222–227.
 21. Atkin WS, Morson BC, Cuzick J. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med* 1992; **326**: 658–662.
 22. Neugut AI, Johnsen CM, Forde KA, Treat MR. Recurrence rates for colorectal polyps. *Cancer* 1985; **55**: 1586–1589.
 23. Lennard-Jones JE, Morson BC, Ritchie JK, Shove DC, Williams CB. Cancer in colitis: assessment of the individual risk by clinical and histological criteria. *Gastroenterology* 1977; **73**: 1280–1289.
 24. Choi PM, Nugent FW, Schoetz DJ, Jr, Silverman ML, Haggitt RC. Colonoscopic surveillance reduces mortality from colorectal cancer in ulcerative colitis. *Gastroenterology* 1993; **105**: 418–424.
 25. St John DJB, McDermott FT, Hopper JL, Debney EA, Johnson WR, Hughes ESR. Cancer risk in relatives of patients with common colorectal cancer. *Ann Intern Med* 1993; **118**: 785–790.
 26. Selby JV, Friedman GD, Quesenberry CP, Weiss NS. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992; **326**: 653–657.
 27. Rosenberg W, Donald A. Evidence based medicine an approach to clinical problem-solving. *Br Med J* 1995; **310**: 1122–1126.
 28. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine. How to Practice and Teach EBM*. New York: Churchill Livingstone, 1997.
 29. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol* 1990; **43**: 543–549.
 30. Merrick NJ, Fink A, Park RE et al. Derivation of clinical indications for carotid endarterectomy by an expert panel. *Am J Public Health* 1987; **77**: 187–190.
 31. Wortman PM, Smyth JM, Langenbrunner JC, Yeaton WH. Consensus among experts and research synthesis. A comparison of methods. *Int J Technol Assess Health Care* 1998; **14**: 109–122.
 32. Lomas J, Anderson G, Enkin M, Vayda E, Roberts R, Mackinnon B. The role of evidence in the consensus process. Results from a Canadian consensus exercise. *J Am Med Assoc* 1988; **259**: 3001–3005.
 33. Murphy MK, Black NA, Lamping DL et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998; **2**: 1–90.
 34. Freemantle N, Mason J, Eccles M. Deriving treatment recommendations from evidence within randomized trials. The role and limitation of meta-analysis. *Int J Technol Assess Health Care* 1999; **15**: 304–315.

Accepted for publication 29 August 2002