Development of explicit criteria for cholecystectomy

J M Quintana, J Cabriada, I López de Tejada, M Varona, V Oribe, B Barrios, I Aróstegui, A Bilbao

Objective: Consensus development techniques were used in the late 1980s to create explicit criteria for the appropriateness of cholecystectomy. New diagnostic and treatment techniques have been developed in the last decade, so an updated appropriateness of indications tool was developed for cholecystectomy in patients with non-malignant diseases. The validity and reliability of panel results using this tool were tested.

Methods: Criteria were developed using a modified Delphi panel judgement process. The level of agreement between the panellists (six gastroenterologists and six surgeons) was analysed and the ratings were compared with those of a second different panel using weighted kappa statistics.

Results: The results of the main panel were presented as a decision tree. Of the 210 scenarios evaluated by the main panel in the second round, 51% were found appropriate, 26% uncertain, and 23% inappropriate. Agreement was achieved in 54% of the scenarios and disagreement in 3%. Although the gastroenterologists tended to score fewer scenarios as appropriate, as a group they did not differ from the surgeons. Comparison of the ratings of the main panel with those of a second panel resulted in a weighted kappa statistic of 0.75.

Conclusions: The parameters tested showed acceptable validity and reliability results for an evaluation tool. These results support the use of this algorithm as a screening tool for assessing the appropriateness of cholecystectomy.

Gallstones are a common problem in developed countries. Cholecystectomy is the procedure used most frequently to treat gallstones, although there are variations in its use in different geographical areas. Inexplicable variations in surgical rates, identification of inappropriate care, and escalating healthcare costs raise questions about potential underuse or overuse of many medical and surgical procedures, including cholecystectomy. Central to this question is the determination of what constitutes appropriate indications for a given procedure. Unfortunately, there are no rigorous scientific data on efficacy and effectiveness to justify medical practice, so other factors must be used to determine criteria for appropriateness. Uncertainty as to the indications for cholecystectomy has been reported by a number of researchers—for example, patients with an asymptomatic calculus, those with symptoms suggestive of cholelithiasis but no diagnostic evidence, and those with polyps or cholesterolosis.

A method that combines expert opinion with available scientific evidence developed by investigators at the RAND Corporation and the University of California at Los Angeles has been used to evaluate the appropriateness of a number of medical and surgical interventions. Previous studies of cholecystectomy using the RAND methodology were performed in the early 1980s in the US and in the late 1980s in Israel and the UK to provide managers and clinicians with information that would help them decide whether or not cholecystectomy was an appropriate treatment for a specific patient. Important changes have since occurred in this field—for example, the introduction of echography and other imaging tests has improved the diagnosis of patients with gallbladder disease or digestive symptoms and new pharmacological and invasive modalities such as endoscopic retrograde cholangiopancreatography (ERCP) and surgical procedures such as laparoscopy have changed the indications for cholecystectomy. The criteria developed in the 1980s are therefore no longer useful.

The validity of any tool is important. Studies to date have investigated the relation between the literature and the ratings, the reliability of the ratings, and the face, content, and construct validity. Despite this, criticisms of the appropriateness method include the low sensitivity of the results to changes over time, to the selection of experts, to the composition of the panel, and to the validity and reliability of the results.

This study was undertaken to provide an updated review of the indications for cholecystectomy in patients with non-malignant disease using the RAND methodology and to study its validity and reliability.

METHODS

Explicit criteria development

The criteria for measuring the appropriateness of the use of cholecystectomy were developed using a previously described explicit method—the RAND appropriateness method—as follows. An extensive literature review was performed to summarize existing knowledge concerning the efficacy, effectiveness, risks, costs, and opinions about the use of cholecystectomy to treat non-malignant diseases. From this review a comprehensive and detailed list of mutually exclusive and clinically specific scenarios (indications) was developed in which cholecystectomy might be performed. This list contained 210 indications for cholecystectomy in patients with no special circumstances such as pregnancy or comorbidities such as diabetes or immunosuppression, which were considered separately. Each indication was specified in sufficient detail that patients within a given indication were reasonably homogeneous.

The indications for cholecystectomy included the following variables:

- Patients with symptomatic cholelithiasis without complications: age (<76 years and >75 years), gallbladder and common bile duct imaging studies (macrolithiasis or microlithiasis, thickened gallbladder, dilated bile duct), surgical risk (low or high, based on the American Society of Anesthesiologists [ASA] criteria).
Patients with complications: if cholecystitis, age and surgical risk; if cholelithiasis, age, surgical risk, presence of current cholelithiasis and previous non-surgical procedures performed as ERCP (successful or not); if cholangitis, same as above; if pancreatitis, same as above plus presence of current cholelithiasis and previous non-surgical procedures.

Asymptomatic patients with cholelithiasis: four clinical situations—silent calculus, incidental finding on an imaging test, incidental findings before a surgical intervention in the area, or incidental findings within a surgical intervention in the area. The variables included were age, surgical risk, and previous surgical procedures performed in the area (yes or no).

Patients with symptoms suggestive of cholelithiasis without cholelithiasis: symptoms suggestive of biliary colic, alithiasic cholecystitis, idiopathic pancreatitis. Age and surgical risk were included in all.

Miscellaneous category that included imaging findings: porcelain gallbladder, polyps (≥1 cm or <1 cm), and cholesterolosis, the latter two groups with or without the presence of symptoms.

Table 1

<table>
<thead>
<tr>
<th>Panel agreement</th>
<th>Appropriateness judgement</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Appropriate</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Agreement</td>
<td>86 (80.4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>21 (19.6%)</td>
<td>49 (89.1%)</td>
</tr>
<tr>
<td>Disagreement</td>
<td>0 (0%)</td>
<td>6 (10.9%)</td>
</tr>
<tr>
<td>Total</td>
<td>107 (50.9%)</td>
<td>55 (26.2%)</td>
</tr>
</tbody>
</table>

The percentages in parentheses are for indications rated in each category from the total number of indications rated by the main panel in each appropriateness category, with the exception of the total row.
were collated and presented to the panellists at a 1 day meeting. Each panellist received the anonymous ratings of the other panellists as well as a reminder of his own ratings. After extensive discussion, panellists revised the indications according to the definitions. During the first round each panellist rated 192 separate indications and during the second round 210 indications because a new age category was added in some cases.

To study the reliability of panel 1, a second panel (panel 2) of six surgeons and six gastroenterologists was created by contacting other reputable specialists from different areas throughout the Basque Country. The same documents that were supplied to panel 1 and the 210 original indications were sent to them to be rated in just one round. The final scores were then compared with those of panel 1.

Data relating to some of the algorithm variables were collected for 780 patients from their medical records and the number of theoretical scenarios used was calculated for each of the main diagnostic groups.

**Statistical analysis**

For each panellist the mean appropriateness rating of all indications, the percentage of all indications rated for appropriateness as 1–3, 4–6, and 7–9 in all rounds, and the mean change between the two rounds were estimated. Decision algorithms which should permit rapid estimation of appropriateness in practice were compiled from the final results. Weighted kappa statistics and 95% confidence intervals (CI) were calculated to test agreement between different panels. All statistical analyses were performed using SAS for Windows version 8.

---

**RESULTS**

**Panel results**

After the second round 50.9% of the 210 scenarios were considered appropriate, 26.2% uncertain, and 22.9% inappropriate. Agreement was reached on 53.8%, mainly in the appropriate category (80.4% agreement of all appropriate scenarios, table 1).

Changes in the panellists’ scoring from round 1 to 2 were minor, except for those who had the lowest and highest scores during the first round who moved their scores to the group mean in the second round (table 2). No differences were found in the scoring between surgeons and gastroenterologists with mean (SD) scores at the second round of 5.98 (0.44) and 5.55 (0.75), respectively.

A second panel of six surgeons and six gastroenterologists (panel 2) scored all the indications; the reliability of this panel compared with panel 1 gave a kappa statistic of 0.75 (95% CI 0.68 to 0.82). No disagreement was found between the two panels for appropriate/inappropriate categories, but panel 2 tended to rate fewer indications as appropriate than panel 1 (table 3). The percentage of uncertain indications was higher in panel 2 than in panel 1, while the percentage of indications considered inappropriate was similar.

**Synthesis of panel results**

Five groups of scenarios were analysed (fig 1):

1. patients with symptomatic cholelithiasis without complications: the panel considered a cholecystectomy to be appropriate in patients with ASA grade I–III; a thickened gall-bladder and a dilated common bile duct and age <76 years; or an ASA grade IV and age <76 years with no previous surgical interventions in the supramesocolic area (fig 2);
(2) patients with symptomatic cholelithiasis with complications which included cholecystitis, pancreatitis, cholecdocholithiasis, and cholangitis (fig 3);

(3) asymptomatic cholelithiasis (fig 4);

(4) symptoms suggestive of biliary disease without cholelithiasis on imaging: in these patients and in those with a diagnosis of idiopathic pancreatitis cholecystectomy was never an option. In those with a diagnosis of alithiasic cholecystitis aged ≤76 years the intervention was judged to be appropriate (fig 5);

(5) other imaging findings (those related to porcelain gallbladder, polyps, and cholesterolosis) were analysed as separate scenarios (fig 6).

Field study
Of the 210 theoretical scenarios scored by the panel of experts, only 44 (21%) were used in the 780 medical records reviewed in the field study. Of those with cholelithiasis without complications, nine of 30 scenarios (30%) were used for 455 patients (58.3%); in those with a complication (25.8% of the sample) 22 of 126 theoretical scenarios (17.5%) were used; seven of 30 scenarios (23.3%) were asymptomatic (12.3% of patients); 0.5% of patients had symptoms but no cholelithiasis (2/18 scenarios, 1.1%); and in the miscellaneous group (cholesterolosis, polyp, or porcelain gallbladder) eight of 42 scenarios (19%) were used for 3.1% of patients.

Figure 2 Appropriateness of cholecystectomy in patients with symptomatic cholelithiasis without complications.

Figure 3 Appropriateness of cholecystectomy in patients with symptomatic cholelithiasis with complications.
DISCUSSION
Cholecystectomy has been used by a number of authors to study the variability or the appropriateness of its use. In this study we have included updated and clear algorithms to provide clinicians with a rapid decision making process for specific patients. The main complaints which can result in cholecystectomy are included in the explicit criteria which were developed using the RAND appropriateness method. Our findings support, in part, the validity and reliability of the tool. The aim is to reduce inappropriate interventions, to increase appropriate procedures, and thus to improve the quality of care provided to patients with gallbladder disease.

The previous RAND cholecystectomy panels used in the 1980s in the US and Israel included nine expert members (three gastroenterologists, two surgeons, two internists, one family physician, and one radiologist). The UK group had two panels—one mixed panel with six specialists (two gastroenterologists, one surgeon, one internist, one family physician, and one radiologist) and one single specialist panel with eight surgeons. The US group developed 192 scenarios, the Israeli panel included 11 chapters and 266 scenarios, while the UK had 272 scenarios but shared similar variables. Although the three groups used a similar methodology, they differed slightly in the variables and categories included. The US panel disagreed on 20.4% of the scenarios and agreed on 53.1%, the Israeli panel disagreed on 35.3%, while the UK mixed panel disagreed on 15% and the single specialist panel on 8%. The experts on the Israeli panel agreed

Figure 4 Appropriateness of cholecystectomy in patients with asymptomatic choledolithiasis.

Figure 5 Appropriateness of cholecystectomy in patients with suggestive symptoms but without choledolithiasis.

Figure 6 Appropriateness of cholecystectomy in other specific situations (findings on image tests).

Exceptions: (1) If age ≤ 45 years and ASA I to III, then Appropriate. If age > 75 years and ASA IV, then Inappropriate; (2) If previous surgery in the area, then Uncertain.
on 46.7% and those on the two British panels agreed on 67% and 61%, respectively. Finally, the Israeli experts determined that 47.6% of the scenarios were appropriate and 46.8% were inappropriate and in the UK study 13–19% considered the scenarios to be appropriate and 27–50% to be inappropriate, depending on which panel criteria were used.

The variables included in the algorithms used in these three studies and ours are not comparable. As the result of technological changes in the last 10–20 years, we introduced variables such as the use of ERCP or new imaging tests such as echography which were not included in the earlier algorithms. We attempted to incorporate new evidence and new scenarios into our algorithm to present updated explicit criteria for the indications for cholecystectomy. An equal number of gastroenterologists and surgeons directly involved in the treatment decision making in Spain were included on the panel.

Our results showed that our main panel reached an acceptable level of agreement after the second round; they were uncertain in 26% of the theoretical scenarios, which is similar to or lower than in many RAND studies. The reliability of the two panels was good (weighted kappa 0.75), taking into account that panel 2 did the scoring in just one round which would have limited the agreement and created more uncertain scenarios. As noted elsewhere, the use of a panel of experts has some limitations, including the composition of the panel. In most studies in which mixed panels have been used, clinicians who do not perform a procedure tend to score lower than those who do perform it.10 21 29 30 31 32 In our case, although some gastroenterologists scored lower than surgeons, as a group there were no differences between them.

We compared the results of our panel of experts with the judgement of a group of specialists and the evidence in the literature over the last 10 years since the latest RAND panels. Our panel tended not to recommend cholecystectomy mainly for patients with high surgical risk (ASA IV)11 and/or older patients and, in cases where applicable, if ERCP was successful. However, there is a lack of evidence as to the appropriate treatment for some common clinical situations such as asymptomatic cholelithiasis and polyps. Even so, our panel criteria matched those reflected in some reports on asymptomatic cholelithiasis12 and polyps,13 and special clinical situations such as alithiasic cholecystis.14 Most authors tend to recommend no intervention in those with asymptomatic cholelithiasis, although our panel did recommend cholecystectomy for those aged <45 years or when the condition was found incidentally. This contradiction probably reflects current clinical practice without any evidence to support the recommendation. The case of those patients with cholelithiasis-like symptoms without lithiasis is controversial except for those with a diagnosis of alithiasic cholecystis in whom the intervention is recommended. Except for these clinical situations, the criteria of our panel were comparable to those of the current evidence.

Different studies have pointed out the limitations of the RAND method which include lack of evidence in the field, quality of the literature review given to the panel, scope of the criteria, definition of appropriateness, panel composition, reliability and management of the panel discussion, validity (sensitivity and specificity, predictive validity) and reliability of their conclusions, sensitivity to changes over time, unsolved or uncertain scenarios, and usefulness of their results.11–13 We tried to avoid some of these known limitations. An up to date review of the literature was given to the panel members; researchers who were experts on the RAND method performed all the panel processes; the results from the main panel were compared with those from a second panel and, finally, our panel criteria matched most of the criteria found in the literature which supported the criteria validity. As mentioned previously, the RAND method is subject to changes over time. New evidence leads to new indications, contraindications, or better treatment options. Nevertheless, our criteria were designed to provide clinicians and managers with an up to date review of the literature and evaluation by a group of experts.

In addition to the previous specific limitations of the RAND method, our study presents some more. With the algorithms used we tried to provide the clinician and manager with a tool that might help them in clinical decision making. Nevertheless, the clinical condition of a particular patient can add relevant information not included in our algorithm that may lead to a different decision. To create the indications, categorical variables had to be created. The cut off point of some of them, such as age, could be a source of disagreement. We based our categorizations on the current literature and on the agreement of the whole panel. Our panel had to decide whether or not cholecystectomy was appropriate for a particular patient. We did not ask the panel members to decide between different surgical options currently available such as open cholecystectomy versus laparoscopy, as it is known that asking the panel to make two decisions at once (whether or not to perform an intervention and the type of intervention) creates more confusion.15 We performed some validity and reliability studies, although more studies on the validity of the tool could have been performed.

Other groups have developed recommendations using other methods such as consensus groups.16 The advantage of our method is that it is more precise and can be applied to individual patients. Nevertheless, this method has not been recommended for that purpose, but as a screening method to determine the degree of appropriateness of one procedure.

Explicit criteria may have different uses—for example, to develop practice guidelines, provide feedback to clinicians, or as a utilization review tool. Some authors have suggested that this method may be useful for comparing levels of appropriateness between different patient populations, to study variability, but not for direct care of individual patients. The value of the criteria developed will be judged based on the changes in clinical practice promoted by them, either through the development of clinical guidelines or through other techniques that may contribute to an improvement in the quality of care provided. To be useful this tool has to be viewed as work in progress that needs continuous improvement in response to technological and social changes.

The explicit criteria developed by our panel of experts and the algorithms presented in this study can enhance the quality of patient care in different ways; by using them as practice guidelines; by applying them in real situations, either retrospective or prospectively; to evaluate the appropriateness of clinical practice; or to compare different hospitals or services or improvements over time. The ultimate purpose in all cases is to help to reduce inappropriate variations in clinical practice.
ACKNOWLEDGEMENTS

This study was partially supported by a grant from the Fondo de Investigación Sanitaria (98002-03). Amaia Bilbao received a grant from the Department of Health of the Basque Government. We also thank the following people for their contributions to this study: Drs Aguado, Arenas, Atín, Baile, Casanova, Hinojosa, Lanas, Monés, Pons, Robles, Suárez Alzamora, Valdivieso, and Zaballa. The following also collaborated in parts of the study: Drs Rodríguez Montes, Iturburu, Olabarrioz, Sarabia, P Múgica, Gardeazabal, J A Múgica, Bernal, Bujanda, Castiella, Múgica, De las Heras. The authors acknowledge Y Etxeberria and I Lafuente for their contribution to the development of the panel of experts and data introduction.

Authors’ affiliations

J M Quintana, A Bilbao, Unidad de Investigación, Hospital de Galdakao, Galdakao, Vizcaya, Spain
J Cabriada, Servicio de Digestivo, Hospital de Galdakao
I Lopez de Tejada, Servicio de Cirugía, Hospital de San Elay, Vizcaya, Spain
M Varona, Servicio de Urgencias, Hospital de Basurto, Bilbao, Vizcaya, Spain
V Oribe, Servicio de Digestivo, Hospital de Basurto
B Barrios, Servicio de Cirugía, Hospital de Basurto
I Aróstegui, Departamento de Matemática Aplicada, Estadística e Investigación Operativa, Universidad del País Vasco, Lejona, Vizcaya, Spain

REFERENCES