The use of diagnostic laparoscopy in patients with suspected appendicitis

van den Broek, W.T.

Citation for published version (APA):
van den Broek, W. T. (2002). The use of diagnostic laparoscopy in patients with suspected appendicitis
Chapter 6

Scoring and diagnostic laparoscopy for suspected appendicitis in adults

W.T van den Broek 1
A.B. Bijnen 1
A. Rijbroek 1
D. J. Gouma 2

1 Department of Surgery, Medical Center Alkmaar, Alkmaar, The Netherlands
2 Department of Surgery, Academic Medical Center, Amsterdam, The Netherlands

European Journal of Surgery, in press
ABSTRACT

Objective: To develop a reproducible scoring system to identify the subgroup of patients who present with a doubtful diagnosis of appendicitis and who would benefit from diagnostic laparoscopy.

Design: Prospective observational study.

Setting: Regional teaching hospital, The Netherlands.

Subjects: 577 consecutive patients during the period 1994-5, and 343 who presented during the period 1996-7.

Interventions: The variables that seemed to be predictive of acute appendicitis were abstracted from the earlier group, a logistic regression analysis applied, and score created. The score was validated on the 343 patients who presented during 1996-7, and then the groups were combined for further analysis.

Main outcome measures: Reproducibility of the scoring system obtained by comparing odds ratios (OR) of the two groups; its effectiveness judged by comparing the delayed and normal appendicectomy rates.

Results: The following variables were significantly correlated with the presence of acute appendicitis: white cell count $10 \times 10^9/L$ or more (score 3), rebound tenderness and male sex (score 2 each); and symptoms present for $<48$ hrs and temperature $38^\circ C$ or more (score 1 each). The OR for the two groups were 1.80 and 1.76, respectively, indicating that score was reproducible. With a sensitivity of 93% and a specificity of 83% it would be at least as accurate as clinical judgment. The normal appendicectomy rate would be 7% instead of 9%, and the negative exploration rates (laparoscopy and primary appendicectomy) would both be 22%. The score would also result in a lower perforation rate (2% compared with 17%).

Conclusion: The score can indicate when there is an indication for laparoscopy in patients with suspected appendicitis.

INTRODUCTION

Acute appendicitis remains a difficult diagnosis to make. When a normal appendix is found after a muscle-splitting incision in the right lower abdominal quadrant, it is usually removed to avoid diagnostic confusion in the future. In most series therefore, a normal appendix is removed in 15% to 30% of patients with suspected appendicitis \cite{1,4,19}. The diagnosis appendicitis is generally based on clinical and laboratory findings \cite{3,7,23,30}.

Despite the introduction of minimally invasive surgery, appendicectomy is still usually done as an open operation in the Netherlands as well as in many other European countries. Diagnostic laparoscopy for acute abdominal pain however, has been used for much longer \cite{6,8,16,27,28}. Routine diagnostic laparoscopy, followed by an open procedure if the appendix looks inflamed would lead to prolonged operation time and extra costs for that group of patients. Because of the relatively
high costs and possible complications of this invasive diagnostic procedure we have stated previously that laparoscopy should be used selectively in these patients in whom it can most effectively contribute in obtaining a diagnosis \(^{(28)}\).

A method should therefore be developed to define the group of patients with a doubtful diagnosis of appendicitis. We decided to create a new scoring system, using a simple and feasible set of variables, which allows us to choose cut-off points for a group of patients in whom appendicitis is unlikely, equivocal, or probable.

Several scoring systems have already been developed but they mainly discriminate only between a normal appendix and appendicitis. By applying these scoring systems, normal appendicectomy rates of between 5% and 17.5% were achieved and false negative rates (missed appendicitis, or delayed diagnosis) between 0 and \(7\%\) \(^{(2,9,10,17,21)}\). Validation of these scoring systems in other series, however, led to considerable higher normal appendicectomy rates, ranging between 13% and 37%, and false negative rates ranging between 0.9% and 13%. These scoring systems are therefore not recommended as standard tools for diagnostic decision-making in suspected appendicitis \(^{(11,14,18,37)}\).

From the other available diagnostic tools, ultrasound is highly observer dependent and not useful in ruling out appendicitis in the general population and computed tomography (CT) is not always available and has the disadvantage of exposing a population, mostly of fertile women, to a relatively high dose of radiation \(^{(5,12,13,15,20,22,23,26,29)}\).

The aim of this study was therefore to define the group of patients with a doubtful diagnosis of appendicitis that could optimally benefit from other diagnostic tools such as diagnostic laparoscopy or CT with the use of a new scoring system.

The application of the scoring system was compared with the results of our clinical practice.

**PATIENTS AND METHODS**

The study was done at the Medical Center Alkmaar, the Netherlands, a 900-bed, regional teaching hospital. Consecutive adolescent and adult patients (age 11 years or over), referred to the hospital by general practitioners for suspected appendicitis, were evaluated prospectively during the period 1994–7. Patients were divided into two groups: those patients evaluated during the period 1994-5 and those evaluated during the period 1996-7. The groups were comparable (table 1).
Table 1: Details of patients studied during the two time periods. Data are number (%) of patients.

<table>
<thead>
<tr>
<th>Sex</th>
<th>1994-5 (n=577)</th>
<th>1996-7 (n=343)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>239 (41)</td>
<td>152 (44)</td>
</tr>
<tr>
<td>Female</td>
<td>338 (59)</td>
<td>191 (56)</td>
</tr>
<tr>
<td>Appendicitis</td>
<td>327 (57)</td>
<td>217 (63)</td>
</tr>
</tbody>
</table>

Among others, the variables age, sex, duration of symptoms, rectal temperature, presence of rebound tenderness, and white cell count (WCC) were recorded. Based on clinical judgment alone the patients symptoms were categorised as unlikely, possible or probable appendicitis, and accordingly they were observed, or had a diagnostic laparoscopy or open operation. Patients were observed in the ward, or as outpatient’s basis if this was socially and practically acceptable. In case of increasing symptoms they were re-evaluated immediately or otherwise routinely after one day in the emergency ward. Depending on their symptoms, those observed patients primarily were then discharged or had a diagnostic laparoscopy or an open appendicectomy.

At laparoscopy, a normal appendix was left in place as a routine; an inflamed appendix was removed laparoscopically or by a muscle-splitting incision, depending on the experience of the surgeon. During an open procedure the appendix was always removed irrespective of the macroscopic appearance. All removed appendix were examined histologically. Data were analysed using the Statistical Package for the Social Sciences (SPSS) computer software. Negative appendectomy rates, delayed (missed appendicitis), and perforation rates were calculated. To compare these results with those of the scoring system, the patients for whom we could retrieve all variables required for the scoring system (862/920, 94%) were considered.

Creating the scoring system

From the patients seen during 1994-5, variables were obtained that correlated significantly with the prevalence of appendicitis on univariate analysis (chi-square test). Sensitivity and specificity of each variable were calculated.

We then did a regression analysis to establish how strongly the individual variables contributed to the diagnosis of appendicitis. Each significant variable was given a score based on its regression coefficient. The final score was created from the sum of these scores, from 0 (appendicitis unlikely) to 9 (appendicitis very likely).
Evaluation of the scoring system

The odds ratio (OR) is a measure that indicates the increased chance of having appendicitis for each score. The scoring system was validated by comparing OR for scores from the group 1994-5 with the odds ratios for score of independent patients in the group 1996-7. Then the data from both groups were combined.

To evaluate the scoring system it was applied retrospectively to the total number of evaluated patients (n=862). By choosing two cut-off points in our scoring system, patients could be divided into three groups: low probability - amenable to observation; intermediate probability-necessitating further investigation such as diagnostic laparoscopy; and high probability of appendicitis - justifying immediate appendicectomy through a muscle-splitting incision. The results that would have been obtained from applying the system were compared to the actual results of clinical practice, done in the same series.

RESULTS

Creating the scoring system:

The variables nausea, anorexia, tenderness on rectal examination and migration of pain were not significant for predicting appendicitis. The following three objective variables: male sex, temperature (≥ 38°C), WCC 10x10^9/L or more, and two subjective variable, rebound tenderness and duration of symptoms (≤ 48 hours) correlated significantly (p ≤ 0.01) with a diagnosis of appendicitis. Using logistic regression these variables were allotted a certain weight to form the scoring system (table 2).
Table 2:

Variables predicting appendicitis obtained after chi-square testing. Score based on regression coefficient of variable in regression analysis (n = 577, 1994-5).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>chi-square</th>
<th>p-value: univariate</th>
<th>Regression coefficient</th>
<th>p-value: logistic regression</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukocyte count: (≥ 10.10^{9}/L)</td>
<td>0.82</td>
<td>0.55</td>
<td>87.0</td>
<td>&lt; 0.01</td>
<td>1.6144</td>
<td>0.0000</td>
<td>3</td>
</tr>
<tr>
<td>Rebound tenderness</td>
<td>0.81</td>
<td>0.56</td>
<td>85.0</td>
<td>&lt; 0.01</td>
<td>1.3927</td>
<td>0.0000</td>
<td>2</td>
</tr>
<tr>
<td>Sex (male)</td>
<td>0.54</td>
<td>0.75</td>
<td>50.2</td>
<td>&lt; 0.01</td>
<td>1.1635</td>
<td>0.0000</td>
<td>2</td>
</tr>
<tr>
<td>Duration symptoms: (≤ 48 hrs)</td>
<td>0.83</td>
<td>0.35</td>
<td>24.7</td>
<td>&lt; 0.01</td>
<td>0.5417</td>
<td>0.0055</td>
<td>1</td>
</tr>
<tr>
<td>Temperature (≥ 38°C)</td>
<td>0.75</td>
<td>0.37</td>
<td>9.9</td>
<td>&lt; 0.01</td>
<td>0.3019</td>
<td>0.0473</td>
<td>1</td>
</tr>
</tbody>
</table>

Maximum score: 9

The scores were derived from the regression coefficients (RC) instead of using the exact RC, to make the scoring system workable in a practical setting.

The relative incidence of appendicitis for each score from patients in the two groups are shown in Fig. 1. As expected, the percentage with appendicitis increases as the score increases. The OR of the scores in the group 1994-5: 1.80. This means that a difference in score of 1 point contributed to a 1.8 greater chance of having appendicitis. The OR of the scores in the 1996-7 was almost identical at: 1.76, which validates it as a reproducible scoring system in this setting.
Scoring and diagnostic laparoscopy for suspected appendicitis in adults

Figure 1:
Incidences of appendicitis according to the score.

As the findings in the two groups did not differ, they have been combined for further analysis to evaluate the score's clinical efficiency.
Chapter 6

Observed results:

All evaluated variables could be retrieved from 862/920 (94%) patients and these were compared with the results of the scoring system. Clinical practice led to an overall 15% normal appendicectomy rate and a 9% delayed operation (missed appendicitis) rate. A total of 276 (32%) patients had primary diagnostic laparoscopies. There were 189 negative explorations (diagnostic laparoscopy and negative appendicectomy) (22%) performed.

In the three groups created by clinical judgment the percentage appendicitis were 16%, 63% and 89% respectively (table 3).

Table 3:

Outcome of treatment of all patients by clinical practice. Data are number (%) of patients.

<table>
<thead>
<tr>
<th>Patients</th>
<th>Appendicectomy</th>
<th>Appendicitis</th>
<th>Normal appendix removed</th>
<th>Laparoscopy alone</th>
<th>Perforation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>261 (30)</td>
<td>71 (27)</td>
<td>43 (16)</td>
<td>28 (11)</td>
<td>22 (8)</td>
</tr>
<tr>
<td>Laparoscopy</td>
<td>277 (32)</td>
<td>197 (71)</td>
<td>174 (63)</td>
<td>23 (8)</td>
<td>80 (29)</td>
</tr>
<tr>
<td>Open operation</td>
<td>324 (38)</td>
<td>324 (100)</td>
<td>288 (89)</td>
<td>36 (11)</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>862</td>
<td>592 (69)</td>
<td>505 (59)</td>
<td>87 (15)</td>
<td>102 (12)</td>
</tr>
</tbody>
</table>

Delayed operation rate: 43/505 (9)
Negative appendicectomy rate: 87/592 (15);
Total negative explorations: 189 (22).

Evaluation of the scoring system:

We used the scoring system to decide which treatment should follow: observation, laparoscopy, or appendicectomy, so we chose two cut-off points. The lowest cut-off point was established by the sensitivity (indication of the ability of the test to rule out appendicitis), as the specificity (indication of the ability of the test to diagnose appendicitis) was useful in establishing the higher cut-off point.
The sensitivity and specificity of the cut-off points that was applied to all patients are shown in Fig. 2.

Figure 2:
Sensitivity and specificity according to the cut-off point.

If the low and high cut-off points were varied it would affect the numbers of delayed diagnoses, laparoscopies done, and normal appendixes removed. Considering fig. 2, the optimal low cut-off point could be 2 (sensitivity 97%) — that is, patients with the scores 0 and 1 would be observed, in combination with a high cut-off point of 7 (specificity 95%) — that is patients with scores of 8 and 9 will have immediate appendicectomy, as this would lead to a delayed appendicectomy rate of 5/505 (1%) and a normal appendicectomy rate of 17/522 (3%). This seems an attractive approach in terms of normal appendicectomy and delayed diagnosis rates. However, the number of laparoscopies done (585/862, 68%) and therefore the total number of operative interventions, would be much
higher than when clinical judgment alone was used. However, if we chose a low cut-off point of 3 (patients with scores of 0, 1, 2, and 3 will be observed) combined with a high cut-off point of 6 (patients with scores of 7, 8, and 9 will have immediate appendicectomy), it would theoretically lead to a 36/505 (7%) delayed appendicectomy rate, 304 (35%) laparoscopies, and 59/564 (10%) normal appendicectomies (table 4). When using these cut-off points, the sensitivity of the scoring system is 93% with a specificity of 83%.

The delayed operation rate (7% compared with 9%) and percentage of laparoscopies done (35% compared with 32%) of our scoring system were comparable with the clinical judgment. The scoring system could lower the normal appendicectomy rate from 15% to 10%, but we must realise that removal of a normal appendix after laparoscopy and the treatment of patients who were initially observed were not taken into account in the evaluation of the scoring system. The normal appendicectomy rate would therefore be slightly higher. The total number of negative explorations (diagnostic laparoscopy and primary muscle splitting incision) was 22% for both scoring system and clinical practice, though the scoring system would lead to a lower perforation rate in those patients who were primarily observed (2% compared with 17 %), tables 3 and 4.

In the three groups created by our cut-off points of 3 and 6, the percentages with appendicitis were 18%, 57%, and 83% respectively.

Table 4:

Theoretical results of outcome of treatment according to the scoring system, using cut-off points 3 (low) and 6 (high). Patients with scores 0-3 will be observed, those with scores 7-9 will have immediate appendicectomy through a muscle-splitting incision, and those with scores 4-6 will have diagnostic laparoscopy.

Data are number (%) of patients.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Appendicectomy</th>
<th>Appendicitis</th>
<th>Normal appendix removed</th>
<th>Laparoscopy alone</th>
<th>Perforation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>203 (24)</td>
<td>36 (18)</td>
<td>36 (18)</td>
<td>-</td>
<td>-</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Laparoscopy</td>
<td>304 (35)</td>
<td>173 (73)</td>
<td>173 (57)</td>
<td>-</td>
<td>131 (43)</td>
<td>37 (12)</td>
</tr>
<tr>
<td>Open operation</td>
<td>355 (41)</td>
<td>355 (100)</td>
<td>296 (83)</td>
<td>59 (17)</td>
<td>-</td>
<td>54 (15)</td>
</tr>
<tr>
<td>Total</td>
<td>862</td>
<td>564 (65)</td>
<td>505 (59)</td>
<td>59 (7)</td>
<td>131 (43)</td>
<td>93 (11)</td>
</tr>
</tbody>
</table>

Delayed operation rate: 36/505 (7%); Negative appendicectomy rate: 59/564 (10%); Total negative explorations: 190 (22)
The distribution of appendicitis for men and women with different scores is shown in Fig. 3. Male sex contributes 2 points in our scoring system, so ranges the score in men between 2 to 9 and in women between 0 to 7. For scores of 3 to 7 distribution of appendicitis among men and women is almost identical. The OR are also comparable (for men 1.67 and for women 1.79).

Figure 3:
Percentages of men and women with appendicitis according to the score.
Men’s score ranged from 2-9 and women’s from 0-7.
DISCUSSION

The indiscriminate use of laparoscopy for patients with possible appendicitis might lead to a considerable increase of costs in those clinics where laparoscopic appendicectomy is not routine. We have shown that improved results can be obtained by using laparoscopy selectively, as there are subsets of patients in whom the diagnosis of appendicitis is almost certain, and some in whom it is very unlikely. Our results show that such subsets can be identified by good clinical judgement but also by a simple, reproducible, scoring system.

Scoring systems have previously been devised by others to decide whether or not to operate. In the present study the scoring system was used to define a group of patients in whom additional laparoscopy can be helpful. The correlation between score and likelihood of appendicitis allows us to choose cut-off points to increase the diagnostic yield of laparoscopy. However, a caveat is that the availability of laparoscopy might invite us to take a more aggressive approach for patients with abdominal complaints, as observation heightens the chance of perforation; on the other hand laparoscopy has its own morbidity. It therefore seems reasonable option to choose the lower cut-off point on a similar level to that of clinical judgement. In doing so, laparoscopy would be used mainly to prevent normal appendicectomies. The number of laparoscopies will then be regulated by the percentage of normal appendicectomies that is regarded as acceptable. Cost-efficiency studies (laparoscopies compared with risks of perforation and normal appendicectomies) should optimally decide the choice of these points, but are not yet available.

The present study has shown that application of this scoring system would have led to similar or slightly better results than achieved by clinical judgement. Scoring could therefore be of clinical help, particularly for less experienced physicians.

It is clear that the overall results of a certain policy based on cut-off points differ from the outcome of an individual patient with a certain score (Fig. 1), so a patient with a score of 3 still has a 27% chance of having appendicitis. Individual factors will therefore also have a role in the final decision of whether or not to operate.

The scoring system also showed that the indication range for laparoscopy in men and women differs, because adding 2 points to the score for men results in equal percentages of appendicitis among the different scores. When using a high cut-off point of 6, women with a score of 7 (those women with a temperature \( \geq 38^\circ C \), rebound tenderness, symptoms less than 48 hours and WCC \( \geq 10x10^9/L \) ) could also have an appendicectomy without a preceding diagnostic laparoscopy. In men diagnostic laparoscopy can also be useful under certain conditions.

The delay in diagnosis caused by observation in some patients led to the removal of 17 perforated appendixes. If our scoring system had been used, only 2 appendixes would have perforated in the observed patients. Even so, delayed diagnosis with the inherent increased risk of perforation is the price we have to pay if we want to avoid every patient with abdominal complaints having an operation.

It is still not clear if the same scoring system and probability cut-off points can be used in other hospitals, as it may depend on the prevalence of appendicitis in the patient population referred as a result of referral policies by different general practitioners. However, the model can always be used
Scoring and diagnostic laparoscopy for suspected appendicitis in adults

with probability cut-off points that are suited to that specific hospital. Ideally, others should validate our scoring system prospectively.

ACKNOWLEDGEMENT

We thank Ms ESM de Lange from the biostatistics department of the Free University Hospital, Amsterdam, the Netherlands for helping with the statistical analysis.

REFERENCES


87