Improved Alvarado Score (MANTRELS) for the Early Diagnosis of Acute Appendicitis

Alfredo Alvarado, MD

1Alma mater: Universidad Nacional de Colombia
2Certified by the American Board of Surgery and the American Board of Emergency Medicine
3International College of Surgeons, Chicago IL, USA
4Society of Laparoscopic Surgeons, Chicago IL, USA
5American College of Emergency Physicians, Irving TX, USA

*Corresponding author: Alfredo Alvarado, MD, Alma Mater: Universidad Nacional de Colombia, Certified by the American Board of Surgery and the American Board of Emergency Medicine, International College of Surgeons, Chicago IL, USA, Society of Laparoscopic Surgeons, Chicago IL, USA, American College of Emergency Physicians, Irving TX, USA, E-mail: alfredoalvari@hotmail.com

Abstract

After a careful analysis of various diagnostic scores on acute appendicitis I am introducing here an Improved Alvarado score (MANTRELS) that includes several substitute predictive factors aimed to obtain better clinical results.

Keywords

Improved Alvarado score, Alvarado score, MANTRELS score, Acute appendicitis diagnosis, Clinical approach for diagnosis of acute appendicitis

Introduction

A correct and timely diagnosis of acute appendicitis is very important in the medical practice since it could avoid complications such as perforation, abscess formation and peritonitis, and at the same time could reduce the negative appendectomy rate. It is for this reason that several diagnostic scores have been developed during the last few decades with various results. One of the first ones was the Alvarado Score (MANTRELS) developed in 1986 with the aim to reduce the useless appendectomies without increasing the risk of perforations.

Evaluation of the predictive factors used in the Alvarado score

According to Ying-Lie, et al. [1], in a meta-analysis of 19 different appendicitis scores, found that the six most relevant features were: Elevated WBC, RLQ tenderness, combination of anorexia, nausea or vomiting, rebound tenderness, migration of pain to the RLQ, and elevation of temperature.

In another study, Sandell, et al. [2] found that, among the signs, tenderness in the right iliac fossa had the greatest impact on the decision to perform appendectomy with an odds ratio (OR) of 80.3 followed by indirect tenderness with an OR of 29.1. Among the symptoms, they found that pain migration was the most important symptom with an OR of 23.6, and image diagnostics gave an OR of 4.99. All of these signs and symptoms had a p-value of < 0.001.

Wilasrusmee, et al. [3] in a systematic review of scores performance, found that rebound pain was the most common sign (76.9%) followed by right lower quadrant tenderness (61.5%), and right lower quadrant guarding or elevated temperature (53.9% for both). Ten symptoms were considered in which nausea (64.3%) followed by migration and duration of pain (46.2%) were most commonly included. They also found that leukocytosis and shift to the left were present in 76.9% and 46.2%, respectively. They noted that eight predictors of the Alvarado score had a calibration coefficient of 1.0 which is very good. They constructed a table including 14 derivation studies where it is important to point out that the original Alvarado score
showed the lowest negative appendectomy rate (7%) that contrasts with the negative appendectomy rate (16.3%) for the RIPASA score (Raja Isteri Pengiran Anak Saleha Appendicitis). They recommended that a model score should be simple, easy to apply and to interpret in order to encourage general surgeons to apply these models in clinical practice.

Comparison of the Alvarado score and the RIPASA score

In recent years, a controversy has emerged about the use of the RIPASA score in which some investigators claim that it gives better results than using the Alvarado score in the South Asian and Middle East populations. It is questionable how the RIPASA score, containing the same predicting factors of the Modified Alvarado score, can produce better clinical results. It is important to know that all the additional parameters included in the RIPASA score (Age, Gender, Duration of pain, Right lower pain, Guarding, Rovsing sign, Normal urinalysis and Foreign identification card) do not have a good statistical significance (Table 1).

In a retrospective Multivariate Analysis of Data from children, Dokumco, et al. [4] found that the male predominance was slightly higher with a p-value of 0.03 and a negative urinalysis with a p-value of 0.05 which indicates that these two parameters have a poor statistical significance as compared with other clinical and laboratory predictive factors that showed a p-value of less than 0.001.

In the development study of the RIPASA score, Chong, et al. [5] found a sensitivity of 88.46% and a specificity of 66.67% which is lower than the specificity generally reported when using the Alvarado score. In the evaluation of the RIPASA score, Chong, et al. [6] found a sensitivity of 97.5%, a specificity of 81.8%, a PPV of 86.5%, a NPV of 96.4% and a diagnostic accuracy of 91.8%. The predictive negative appendectomy rate (NAR) was 13.5% but the observed NAR was 19.4%. It is interesting to note that in this study 43.1% were male which contradicts their own claim that gives more importance to the male gender. Their perforation rate was rather low (9.1%) that could be related to the high negative appendectomy rate. Chong, et al. [7], in a comparison of the RIPASA and Alvarado score, found 92 male patients and 100 female patients which contradicts again the fact that they gave 0.5 points to the female gender only. In this study, the specificity was higher for the Alvarado score as compared with the RIPASA score (87.91% vs. 81.32%). The same way the PPV (precision value) was higher for the Alvarado score (86.3% vs. 85.3%).

In a meta-analysis, Frountzas, et al. [8] found that the RIPASA score is more sensitive (94%) than the Alvarado score (69%) but less specific (55% vs. 77%). In addition they found that the male female ratio was 1:1.3 which is the reverse of the accepted ratio of 1.4:1.

In a comparative study between Alvarado score and RIPASA score, El Hosseiny, et al. [9] found 36 female and 20 male patients. Using the Alvarado score the sensitivity and the specificity were 65.2% and 100%, respectively which is the opposite to the RIPASA score that gave a sensitivity and specificity of 100% and 75%, respectively. The PPV and the NPV for the Alvarado score were 100% and 33.3%, respectively which is again the opposite to the RIPASA score with a PPV and a NPV of 95.8% and 100%, respectively. The negative appendectomy rate was 0% for the Alvarado score and 4.17% for the RIPASA score. These data indicate that the Alvarado score is more specific, as demonstrated in many studies, and that the Alvarado score can reduce unnecessary admissions and negative appendectomy and complication rates.

In another study, to evaluate the efficacy of the Alvarado score and the RIPASA score, Goel, et al. [10] found similar results as the El Hosseiny study. The sensitivity for an Alvarado score of 6 or more was 63.3%, and for a RIPASA score of 7.5 or more was 95.6%. However, the specificity and PPV for the Alvarado score were 100% for both which is better than the specificity of 50% and PPV of 94.5% for the RIPASA score. The negative appendectomy rate was 0% for the Alvarado score and 87.91% vs. 81.32% for the RIPASA score. The area under curve (AUC) for the Alvarado score was slightly higher than the RIPASA score (0.926 vs. 0.914).

Singla, et al. [11], in a prospective study to evaluate the efficacy of the Alvarado score and RIPASA score, found that both scoring systems were equally good. Sensitivity for the Alvarado and the RIPASA scores were 64.4% and 96.7%, respectively. However, specificity for the Alvarado score was 100% and for

<table>
<thead>
<tr>
<th>Points</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration</td>
<td>1 &lt; 0.001</td>
</tr>
<tr>
<td>Anorexia</td>
<td>1 &lt; 0.001</td>
</tr>
<tr>
<td>Nausea/Vomiting</td>
<td>1 &lt; 0.001</td>
</tr>
<tr>
<td>Tenderness in the RLQ</td>
<td>2 &lt; 0.001</td>
</tr>
<tr>
<td>Rebound pain in the RLQ</td>
<td>1 &lt; 0.001</td>
</tr>
<tr>
<td>Elevation of temperature</td>
<td>1 &lt; 0.015</td>
</tr>
<tr>
<td>Leukocytosis</td>
<td>2 &lt; 0.001</td>
</tr>
<tr>
<td><strong>Total points for the Modified Alvarado score</strong></td>
<td><strong>9</strong></td>
</tr>
<tr>
<td>Shift to the left</td>
<td>1 &lt; 0.001</td>
</tr>
<tr>
<td><strong>Total points for the Alvarado score</strong></td>
<td><strong>10</strong></td>
</tr>
<tr>
<td>Age</td>
<td>0.5-1 &lt; 0.118</td>
</tr>
<tr>
<td>Male/Female</td>
<td>1-0.5 &lt; 0.118</td>
</tr>
<tr>
<td>Duration of pain</td>
<td>0.5-1 &lt; 0.221</td>
</tr>
<tr>
<td>Guarding</td>
<td>2 &lt; 0.003</td>
</tr>
<tr>
<td>Rovsing sign</td>
<td>2 &lt; 0.378</td>
</tr>
<tr>
<td>Normal urinalysis</td>
<td>1 &lt; 0.506</td>
</tr>
<tr>
<td>Right iliac fossa pain</td>
<td>0.5 &lt; 1.000</td>
</tr>
<tr>
<td>Foreign ID card</td>
<td>0.5 &lt; 0.477</td>
</tr>
<tr>
<td><strong>Total points for the RIPASA score</strong></td>
<td><strong>17.5</strong></td>
</tr>
</tbody>
</table>
the RIPASA score 60%. The PPV (precision value) for the Alvarado score was 100% and for the RIPASA score 70.7%. On histopathology examination, the negative appendectomy rate was 0% for an Alvarado score of 7 or more, and 4.5% for a RIPASA score of 7.5 or more. Over all, the negative appendectomy rate was 10%. Appendicular perforation was seen in 30% of patients which is higher than in many similar studies. Singla, et al. constructed a table where they showed the sensitivity and specificity of several studies comparing the Alvarado and RIPASA scores. In this table, the average sensitivity for the Alvarado and the RIPASA scores was 75.37% and 94.6%, respectively, but the average specificity for the Alvarado and RIPASA scores was 82.36% and 52.06%, respectively. This demonstrates that the RIPASA score generates many useless appendectomies and this has to do with its lower specificity as compared with the Alvarado score.

Karami, et al. [12], in a prospective study comparing the RIPASA, Alvarado and AIR scoring systems, found that the sensitivity and specificity of the RIPASA score were 93.18% and 91.67%, respectively. The sensitivities of the Alvarado and AIR scores were 78.41% for both. The specificity of the Alvarado and the AIR score was 100% and 91.67%, respectively. The positive predictive value (precision value) was 100% for the Alvarado score, 98.80% for the RIPASA score and 98.57% for the AIR score. The AUC of the ROC curve was 0.981 for the RIPASA score, 0.906 for the Alvarado score and 0.867% for the AIR score.

Naz, et al. [13], in a study to determine the concordance between the RIPASA, Alvarado and AIR scoring systems for the diagnosis of acute appendicitis, found that using the Kappa statistics, the Kappa value was 0.847 (p > 0.05) which means a strong agreement between both scoring systems. An Alvarado score of 7 points or more and a RIPASA score of 7.5 or more are both strongly predictive of acute appendicitis. The ROC curve showed an AUC of 0.962 for the RIPASA score and 0.938 for the Alvarado score.

Comparison of the modified Alvarado score and the RIPASA score

The RIPASA score contains exactly the same diagnostic factors of the Modified Alvarado score but contains other parameters that do not have a good statistical significance. El Maksoud, et al. [14], in a comparison between the Modified Alvarado score (MAS) and the RIPASA score, found that over all the MAS showed a poor sensitivity (59.8%) a good specificity (87.5%), an excellent PPV (96.9%) and a poor NPV (25.90%). On the other hand, the RIPASA score showed an excellent sensitivity (96%), a very poor specificity (12.5%) and a poor NPV (33.3%). In this study, it is interesting to note that, in patients above 30 years of age, the MAS score showed perfect results of 100% for all of the statistical values. The accuracy of the MAS and the RIPASA scores was 63.3% and 85.0%, respectively, but these results are due to the higher sensitivity of the RIPASA score. In this study, postoperative histopathology revealed a negative appendectomy rate of 13.3%, and a perforation rate of 8.3%. The p-values for all the additional components of the RIPASA score were as follows: Gender 0.193, age 0.928, duration of symptoms 0.008, RIF guarding 0.524, Rovsing sign 0.797, negative urinalysis 0.309, and foreign ID 0.477. All of these p-values are not statistically significant and that is the reason why the sensitivity and the AUC of the C-statistics of the RIPASA score are always higher than the sensitivity for the MAS score so they could be considered as confounding variables. In other words: Trash in, trash out.

Comparison of the Alvarado score with other scores

In a clinical comparative study of different scoring systems in acute appendicitis, Nema and Jain [15] found that the negative appendectomy rate (NAR) for the Alvarado score was 18.9% as compared with the NAR for the Izbicki and Christian scores that showed a NAR of 27.9% and 25.0%, respectively. In this study, Alvarado score had the lowest missed perforations (0%) as compared with the missed perforation rate of 8.3% for both the Izbicki and the Christian scores. Their conclusion was that the Alvarado score appears to be more accurate, simple, rapid, reliable and economic diagnostic modality in the clinical decision making.

Thompson [16], in a clinical prediction rules study in children, found that the Pediatric Appendicitis Score (PAS) has a sensitivity of 82% and a specificity of 65% as compared with the Alvarado score that has a sensitivity and a specificity of 72% and 81%, respectively. He felt that the Alvarado and PAS scores have similar qualities and that they are the most thoroughly evaluated of the clinical scoring systems (CSS) in children with an acceptable performance for risk stratification. Since a number of his staff groups are employed in both pediatric and adult hospital settings, one consistent CSS was felt to be optimal. It is for this reason that Alvarado score was incorporated into these two groups as a pediatric Appendicitis Pathways for their particular region.

Erdem, et al. [17] in a study to assess the reliability of the Alvarado, Eskelinen, Ohmann and RIPASA scoring systems, found that the Alvarado score has a sensitivity of 82%, the Eskelinen and RIPASA scores have a sensitivity of 100% and the Ohman score has a sensitivity of 96%. They also found that the Alvarado score had the highest specificity (75%) as compared with the specificity of the other scoring systems, 44% for Eskelinen, 42% for Ohmann, and 28% for RIPASA. Also, the positive predictive value (PPV) was the highest (88%) as compared with the PPV of the Ohmann (75%), Eskelinen (79%), and RIPASA scores. In addition, the Alvarado score gave the lowest negative appendectomy
rate (12%) as compared with the other scoring systems, 21% for Eskelinem, 22% for Ohmann and 25% for RIPASA scores.

Xingyie, et al. [18] in a retrospective study in China, evaluated seven scoring systems in patients with acute appendicitis. In this study 179 patients were enrolled, their ages ranged from 13 to 87 years. There was 51.4% males and 48.6% females. The sensitivity of the Alvarado score was up to 92.7% and it outperformed the other scores. For scores > 8, Alvarado score gave a sensitivity of 33.33% and a specificity of 97.96% as compared with the AIR score with a sensitivity of 20.59% and a specificity of 96.94%. For advanced appendicitis (macroscopic gangrenous appendix with or without perforation) the rate was 24.5% which is high as compared with other studies. In this study, AIR score performed equally well with the Alvarado score when analyzing high-risk groups (score > 8) and they had a high negative predictive value thereby reducing negative appendectomies.

In a prospective observational study on acute appendicitis worldwide (POSAW), Sartelli, Balocci, Saverio, et al. [19] found a male female ratio of 1.2:1 which confirms that this parameter has no statistical importance. On histopathological examination of 3.631 specimens, they found a negative appendectomy rate (NAR) of 4%, a perforation rate of 7% and a gangrenous appendicitis rate of 18.5%. In this study, 3.857 patients were evaluated using the Alvarado score and 3.751 patients were evaluated using the Andersson (AIR) score. Their conclusion was that the Alvarado, Andersson and WSES grading scores are useful methods to classify the patients since they predict and correlate with surgical or pathological diagnosis.

**Negative appendectomy and perforation rates as indicators of clinical efficiency**

The Alvarado score has been confirmed in many studies around the world as able to generate a low negative appendectomy rate (NAR) with rates below 10% and even below 5%. For instance, Menon, et al. [20] in a study of 100 children, found a negative appendectomy rate of 1.8% when using an Alvarado score of 7 points or more. Ricci, et al. [21] demonstrated that with an Alvarado score of 7 or more the negative appendectomy rate gets close to 5%, and with a score of less than 5, the true NAR raises to 30%. In this study they found that, in a retrospective cohort group, the NAR was 20.7% in comparison with a prospective group that obtained a NAR of 7.8% when using the Alvarado score.

Mariadason, et al. [22] using a stringent definition of negative appendectomy rate, demonstrated that computed tomography (CT) lowered the negative appendectomy rate (NAR) in women from 24.2% in the period 1996-2000 to 9.1% in the period 2006-2010. However, increasing the CT rate during the period of 2001-2010 had no impact on the NAR. The positive predictive values for the Alvarado score and for the CT were similar, 98.60% and 99.03%, respectively. They found a false positive rate of 1.9% for the Alvarado score which proves that this score can reduce the NAR significantly. Nevertheless, the perforation rates were statistically unchanged, ranging between 12.8% and 15.8%.

As mentioned before, El Hosseini [9] reported a negative appendectomy rate of 0% when using the Alvarado score and 4.17% when using the RIPASA score. The same way, Goel [10] reported a negative appendectomy rate of 0% for the Alvarado score and 5.5% for the RIPASA score.

According to Almaramhy [23], perforation rates are more common in children and can reach levels of 82%-92%. In the elderly, Busch, et al. [24] reported perforation rates as high as 29.7%, and Omari, et al. [25] reported a perforation rate in 41% of patients. In developing countries, such as in South Africa, Kong, et al. [26] reported that 60% of patients had perforated appendicitis. In general, perforations are inversely related to the negative appendectomy rate so it would be important that investigators in this field report always these rates because in such a manner they could demonstrate, in practical terms, the real outcome of their studies.

**Diagnosis of acute appendicitis in its late stages**

Bonadio, et al. [27] in a regression analysis in pediatric patients, identified three independent significant variables associated with perforation outcome. They determined ideal threshold values: 1) Duration of symptoms greater than one day, 2) ED-measured fever (body temperature greater than 38.0 °C) and 3) A WBC absolute neutrophil count greater than 13.000 cells/mm³. Risk for perforation was additive with each predictive variable, linearly increasing from 7%, with no variables present, to 85% when all variables were present. These variables can be used in combination with the clinical signs of an acute abdomen to make the diagnosis of a complicated appendicitis such as necrotic, perforated, appendicular abscess or frank peritonitis.

**Reasonable goals when using the improved Alvarado score (MANTRELS)**

- Negative appendectomy rate of 5% or less.
- Perforation rates of 5% or less.

**Conclusion**

Alvarado score is still very useful in the diagnosis of acute appendicitis in the early stages of the disease and is aimed to reduce the negative appendectomy and perforation rates.

Nevertheless, there is room for improvement and it is for his reason that I am introducing here an Improved
Alvarado score (MANTRELS) that, while preserving the scheme of the original score, adds several substitute predictive factors that show a good statistical significance. Also, this Improved Alvarado score could be useful for statistical purposes by providing a more precise indexing of the disease. For instance, it could be used, as a clinical indicator, in the International Classification of Disease at a fifth digit level.

**Explanation of the Improved Alvarado score (MANTRELS) (Table 2)**

(Only one substitution factor should be allowed for each diagnostic factor).

Migration can be substituted with hyperesthesia on the right lower quadrant (Sherren’s triangle) and the Massouh sign: Swishing two fingertips starting on the xiphoid sternum down toward the left and right iliac fossa to elicit hyperesthesia due to peritoneal irritation. A positive sign is a grimace of the patient upon a right-sided (and not left) sweep. Also, lightly touching the patient with the stethoscope creates uncomfortable sensation on the affected area. These two tests can replace the migration symptom in children who cannot communicate well.

Anorexia can be substituted with acetonuria or with the Hamburger test in which the physician inquires if the patient would like to consume his favorite food. If the patient wants to eat, the clinician should consider other diagnoses than appendicitis.

Nausea/Vomiting: Tenderness in the RLQ can be substituted with tenderness on percussion with a fist in the right lumbar area and also with the K-sign that is present in retrocecal and paracolic appendicitis. This sign is present in patients of Indian ethnicity particularly from the Kashmir region and coexist with the psoas sign. Tenderness in the RLQ can be substituted also with the obturator sign that is present when the ruptured appendix is adherent to the fascia covering the obturator internus muscle.

Rebound pain: (Blumberg sign) can be replaced with the Rovsign sign, cough test, inspiration or expiration test, heel drop test, hop test, pain on walking, pain with jolts or road bumps.

Elevation of temperature (oral temperature of 37.3 °C or more, but not fever).

Leukocytosis (> 10,000 cells/mm³).

Shift to the left: stabs or bands (> 5%) or segmented neutrophiles (> 75%).

Headache, body aches and shaking chills that could be present with dengue fever [28,29,30], typhoid [31] or paratyphoid fever, malaria and other infectious diseases. (These symptoms give 2 negative points to the score).

**Table 2: Improved Alvarado score (MANTRELS).**

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration</td>
<td>1 point</td>
</tr>
<tr>
<td>Hyperesthesia in the RLQ</td>
<td></td>
</tr>
<tr>
<td>Massouh sign</td>
<td></td>
</tr>
<tr>
<td>Anorexia</td>
<td>1 point</td>
</tr>
<tr>
<td>Acetonuria</td>
<td></td>
</tr>
<tr>
<td>Hamburger test</td>
<td></td>
</tr>
<tr>
<td>Nausea/Vomiting</td>
<td>1 point</td>
</tr>
<tr>
<td>Tenderness in the RLQ</td>
<td>2 points</td>
</tr>
<tr>
<td>Tenderness on percussion in right lumbar area</td>
<td></td>
</tr>
<tr>
<td>K-sign, psoas sign, obturator sign</td>
<td></td>
</tr>
<tr>
<td>Rebound pain in the RLQ</td>
<td>1 point</td>
</tr>
<tr>
<td>Rovsing sign, cough test, inspiration or expiration test, heel drop test, hop test, pain on walking, pain with jolts or road bumps</td>
<td></td>
</tr>
<tr>
<td>Elevation of temperature (37.3 °C or more)</td>
<td>1 point</td>
</tr>
<tr>
<td>Leukocytosis (&gt; 10,000 cells/mm³)</td>
<td>2 points</td>
</tr>
<tr>
<td>Shift to the left: stabs or bands (&gt; 5%)</td>
<td>1 point</td>
</tr>
<tr>
<td>Segmented neutrophiles (&gt; 75%)</td>
<td></td>
</tr>
<tr>
<td>Headache, body aches, chills</td>
<td>-2 points</td>
</tr>
<tr>
<td>Total</td>
<td>10 points</td>
</tr>
</tbody>
</table>

**Interpretation of the improved Alvarado score (MANTRELS)**

According to Chandio, et al. [32] the Alvarado score was found accurate with a p-value of 0.0001 depending on the score points as follows.

1-4 Unlike (Accuracy 85%)

5-6 Possible (Accuracy 60%)

7-8 Probable (Accuracy 83%)

9-10 Very probable (Accuracy 100%)

In the majority of the studies, a score of 1-4 rule-out the diagnosis of acute appendicitis, and a score of 7 or more confirms the diagnosis. With a score of 5-6 the patient can be observed and may need additional studies.

**Conflict of Interest**

This is a Review article with no Conflict of interests.

**References**


