Clinical Decision-Support Tool for Acute Appendicitis

Stéphane Meystre, MD, MS.

Department of Medical Informatics, University of Utah, Salt Lake City, Utah, USA.

Abstract

The diagnosis of acute appendicitis is still an issue, even with the availability of medical imaging techniques like ultrasonography or computed tomography, and after the publication and evaluation of many clinical scores. The main part of the decision still uses the clinical experience, often limited since physicians taking care of emergency department patients are usually residents or junior attending physicians. Clinical scores rarely comply with all required standards of quality, and are always used separately. In this project, a new approach was tried, developing a decision-support tool with an expert system development application, and using three clinical scores chosen for their qualities.

Introduction

Acute appendicitis diagnosis issue: Although the treatment of acute appendicitis is simple and straightforward, its diagnosis remains a challenge, and the negative appendectomy rate in large series range from 15% to 33%. Furthermore, in the patients with either atypical history or equivocal physical examination findings, particularly in women of childbearing age, the negative appendectomy rate has been as high as 45%. With an annual rate of 250'000 cases in the US and an incidence of 86 every 100'000 people worldwide [1,2], acute appendicitis is a common acute abdomen condition in the emergency department [3].

The treatment is surgical: appendectomy (removal of the appendix). It is one of the most frequently performed surgical procedures in the United States and the most common surgical emergency of the abdomen [4].

Acute appendicitis is a common condition in children, accounting for one-third of all pediatric hospital admissions with acute abdominal pain. This difficult diagnosis has to consider and weigh the morbidity of unnecessary appendectomies against the potential morbidity and mortality of neglected acute appendicitis. The pediatric population, especially children below 5 years, is often subject to higher missed or erroneous diagnosis, due to the lack of specific symptoms and signs. Women are even more at risk, because of the large gynecological differential diagnosis. It was shown that the specificity of the Alvarado score was of 92% with men, 90% with children and 67% only with women [5].

How to improve appendicitis diagnosis? Attempts to increase diagnostic accuracy have included computer-aided diagnosis, imaging by ultrasound scanning and computed tomography, and laparoscopy. Ultrasonography was shown to be helpful in combination of clinical evaluation and score(s) eventually [6,7]. In these studies, GCUS (Graded Compression UltraSonography) was used when the clinical evaluation with the score of Alvarado didn't have clear results. It was also well mentioned that the accuracy of ultrasonography is dependent on the experience of the ultrasonographer. This type of imaging was also included in a score, reaching excellent sensitivity and specificity [8]. The score included 5 clinical items and the ultrasonography. The sensitivity was 82%, and the specificity 96%. The use of computed tomography has also increased these last years, with variable results, as described in Lee et al. [3]. They showed that computed tomography and ultrasonography didn't improve the diagnosis, and even caused delays in the treatment, causing avoidable morbidity. Some computer-based systems were developed to try helping clinicians in this difficult task. A group in Norway developed one of these systems to evaluate the physician's probability of acute appendicitis estimate [9]. The system they developed performed like the

Author's address:
Dr. med. Stéphane Meystre
Department of Medical Informatics
University of Utah School of Medicine
30 North 1900 East – Room AB193
Salt Lake City, UT 84132-2913, USA
Phone: 001 801 5871414
E-mail:
stephane.meystre@hsc.utah.edu

experienced surgeons in diagnostic accuracy, but avoided the usual 10% overestimate of initial probability made by surgeons. The area under the ROC (Receiver Operating Characteristic) curve was 0.81, comparable to some clinical scores like Alvarado, Arnbjørnsson and Teicher (area under the ROC 0.81, 0.83 and 0.85 when initially tested). But these scores performed much worse when tested in different groups (0.60-0.66). This system could therefore be used as a decision-support tool [9].

Various clinical scoring systems have been devised to aid diagnosis (Macklin, Radcliffe et al. 1997). They will be discussed below. Despite the availability of new diagnostic means like ultrasonography, computed tomography or laparoscopy, the frequency of misdiagnosis leading to unnecessary appendectomy has not changed, nor has the frequency of perforation decreased [4]. This study, based on the CHARS (Washington State Comprehensive Hospital Abstract Reporting System) and using ICD-9 codes (International Classification of Diseases), showed that the rate of ruptured appendix and misdiagnosis remained stable for the last ten years. Incidental appendectomies decreased and nonincidental appendectomies remained stable.

Clinical scores: As of yet, at least eleven clinical scores have been published and tested to help acute appendicitis diagnosis. The first was published by *Van Way* [10], in 1982. The score was applied to patients with appendectomy, and based on only four history items: gender, duration of pain, nausea and anorexia. A cutoff value was fixed, to separate patients to operate from patients to monitor. Evaluations of the score had insufficient results to recommend it for clinical use [11].

In 1983, *Teicher* [12] published a score also made for appendectomy patients, but with more items: 4 from the history (gender, age, duration of pain and presence of urinary or gynecological symptoms), 2 from the examination (right lower quadrant muscle spasm, rectal mass right side), and one investigation (leucocytes). The cutoff value separated patients for operation or monitoring. It was evaluated and recommended for further large scale testing [11].

Arnbjørnsson published another score in 1985[13], based on many items from the history and the examination (duration of symptoms, location of initial pain, migration of the pain, previous attacks of similar symptoms, anorexia or nausea, vomiting, diarrhea, gender, fever, abdominal tenderness, rebound tenderness, rigidity or guarding, abdominal mass, rectal tenderness and rectal mass). The cutoff value also separated patients for operation or observation. Its evaluation showed some good results, but not good enough to recommend it [11].

The next was published by *Alvarado* [14] in 1986. Using Bayesian analysis, he assigned a numerical value to each of eight signs, symptoms and laboratory values. He recommended immediate operation for a score higher than 6 over 10, observation for scores of 5 to 6, and discharge with outpatient follow-up for scores lower than 5. A rate of initial negative appendectomy of 7%, and a missed appendicitis rate of 14% were reported then. A prospective study in the UK [15] was performed in 1996, concluding that this score couldn't be recommended to use in pediatric surgical practice, because of significantly inferior results obtained with the score compared to their current clinical practice. Another evaluation concluded that it was the only score that fulfilled all standardized criteria. These criteria are based on recommendations about acceptable rates for removing normal and perforated appendices [17,11].

Fenyö published a score in 1987 [16], based on a very long list of items from the history and examination, used separately for men and women (19 items for each gender). A cutoff value allowed separating patients for monitoring or operation only. Evaluated by Ohmann et al. [11], the results showed that only two of the standard criteria were fulfilled.

In 1988, *Lindberg* [18] published a score that also fulfilled two of the four standard criteria when evaluated by Ohmann et al. [11]. It originally contained 10 items from the history, examination and investigation, and separated patients in three groups: operation, observation and exclusion. In 2000, a *modified version* was proposed [19], considering the temperature instead of the aggravation of pain by coughing and modifying the values and score points for each item. It was retrospectively tested on 197 pediatric patients (2 to 17 years old) with excellent results: only 2% of initial negative appendicitis rate (unnecessary operations), but 8% would have been discharged wrongly (missed appendicitis rate). They also showed that the sensitivity was much lower for girls (76%) than boys (91%), but the specificity was similar (87-88%).

The next score was proposed by *Izbicki* in 1990 [20], based on 7 items (gender, leucocytes, guarding, rebound pain, migration of pain, duration of pain and type of pain). The cutoff value only separated the patients for operation or observation, and it performed badly when tested by Ohmann et al. [11]: initial negative appendectomy rate of 49%!

Then, *de Dombal* published a score in 1991 [21], to separate acute appendicitis from non-specific abdominal pain, and used 7 items (migration of pain, aggravation with movements or coughing, simultaneously nausea, anorexia and vomiting, skin color, tenderness in right lower quadrant, rebound and guarding, and rectal tenderness). The score allowed separating patients to exclude, observe or operate. Tested by Ohmann et al. [11], it performed very badly with a potential perforation rate of 82%!

In 1992, *Christian* published a simpler score [22], only based on 5 items with 1 or 0 points (abdominal pain, vomiting, tenderness, low grade fever and polymorphonuclear leucocytosis). It separated patients to operate from those to observe, and didn't perform well when tested [11].

The last score published was developed by *Ohmann* after his disappointing evaluation of all the previous scores, in 1995 [23]. He based it on 8 items (tenderness in right lower quadrant, rebound pain, no urinary symptoms, continuous pain, leucocytosis, age, migration of pain and guarding), with cutoff values to decide whether to discharge, observe or operate patients. Evaluated in 1999, it performed very well with a missed appendicitis rate of only 0.9% and an area under the ROC curve of 0.901 [24]. Another evaluation showed a missed appendicitis rate of 1.1% (falling to 0% when combined with the clinical evaluation!) [25].

Methods

Terms definitions: Many terms are used to evaluate the diagnosis and treatment of acute appendicitis. They have been well defined, and should be used with precision to allow comparison of different methods of diagnosis and treatment.

Initial negative appendectomy rate:

$$\frac{\text{Patients without acute app. assigned to op.}}{\text{Patients in op. group}}$$

(means the proportion of patients without acute appendicitis assigned to the operation group). Should be ≤ 15%.

Potential perforation rate:

$$\frac{\text{Patients with acute app. not assigned to op.}}{\text{Patients with acute app.}}$$

(means proportion of patients with acute appendicitis not assigned to the operation group). Should be ≤ 35%.

Initial missed perforation rate:

$$\frac{\text{Patients with perforated app. not assigned to op.}}{\text{Patients with perforated app.}}$$

(means proportion of patients with perforated appendicitis not assigned to the operation group). Should be ≤ 15%.

Missed appendicitis rate:

$$\frac{\text{Patients with acute app. assigned to exclusion group}}{\text{Patients with acute app.}}$$

(means proportion of patients with acute appendicitis assigned to the exclusion group). Should be ≤ 5%[11].

The knowledge base: The goal of the diagnosis of acute appendicitis is to avoid unnecessary operations (lowest initial negative appendectomy rate as possible, at least less than 15%), to avoid missed appendicitis (lowest missed appendicitis rate as possible, at least lower than 5%), and after these to have low potential perforation rates and low initial missed perforation rates. Being able to operate only real acute appendicitis lowers the morbidity associated with the surgery and reduces the costs. Avoiding keeping patients without appendicitis in observation and discharging only them also diminishes the costs. To approach these goals, three scores were chosen. The score of Ohmann was selected for its low missed appendicitis rate, to decide which patients to discharge. The modified score of Lindberg was selected for its low initial negative appendectomy rate, to select the patients to operate. All other patients should be observed. The score of Alvarado was also included because of its consideration as a standard to help acute appendicitis diagnosis.

In summary, the chosen scores and the rates when they were tested [11, 24, 25, 19]:

Scores:	Lindberg modified	Lindberg modified	Alvarado
Initial negative appendectomy rate	2%	14,3-15,6%	5,3-29%
Potential perforation rate	28,7%	37-38,5%	19-76%
Initial missed perforation rate	not available	40-40,9%	14-65%

Table 1: Chosen clinical scores

The expert system application: CORVID™ [26]. This Java™-based application was

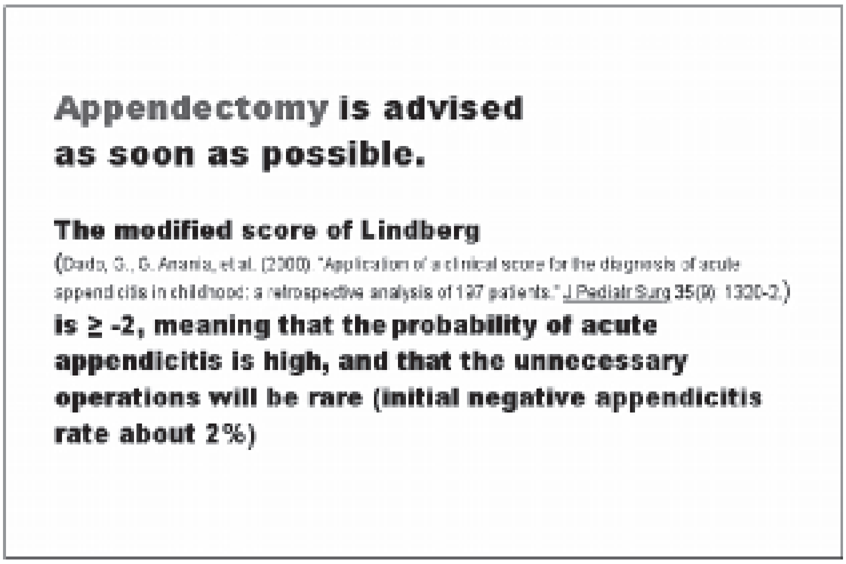

Figure 1: Screenshot of the decision-support tool

chosen for its ability to derive applets or Java applications from the system created. Developed by Exsys® Inc., this application allows the development of on-line expert systems for the Web. It has three important features: object-oriented structure, Logic Blocks and practical Java delivery. The object-oriented structure is based on Microsoft's Visual Basic model, making it easy to build systems using methods and properties of Variables, while not requiring the developers to change the way they think and describe their decision-making steps and logic. Logic Blocks can be any combination of rules and trees that have a related function, anything from an entire knowledge base to a single rule. This allows the logic to be organized into blocks that behave as objects. They can be run via forward or backward chaining, and even associated with a spreadsheet file. CORVID applications are delivered by a small applet (about 100 Kbytes) that allows robust interface design options, and is fast for users to download. If special system features are required, the applet can communicate with other applets on the page. For data only available on a server, CORVID can access CGI (Common Gateway Interface), ASP (Active Server Pages) and JSP™ (JavaServer Pages™) pages to perform server-side calculations [26]. In our situation, the knowledge base was created with the three scores described above, introduced into CORVID as Variables and Logic Blocks. Each item tested and its answers were introduced as Variables (Static Lists) and other Variables (Numeric) were created to assign values to the scores. The three total values of the scores were also created as Variables (Numeric). Logic Blocks were then created to ask the user the questions needed and select the right answers with the associated score values. All the Logic Blocks (16) are controlled by a Command Block that runs each Logic Block in a forward chaining manner, asking the user to enter all answers. The scores are then calculated and one of the three possible attitudes with the patient (discharge, observation or operation) is displayed. If the score of Ohmann is lower than 6, the discharge with outpatient follow-up is proposed. (Fig. 1)

If the modified score of Lindberg is higher than -3, the operation is proposed. For all other cases, inpatient observation is advised. For information, the results of all three scores are finally displayed, and the details of each score are shown.

Discussion

Because the potential perforation rate and the initial missed perforation rate have not been included in the knowledge base, the possible results of this tool have to be considered cautiously. One could imagine that a score could be developed to have the lowest possible initial negative appendectomy rate, thus having almost no false positives, but a lot of false negatives. It means that a lot of not operated patients would have an acute appendicitis, and the potential perforation rate and initial missed perforation rate would be very high. But, as told before in this paper, the reasonable goals for these two rates are 35% and 15%, respectively. The summary table (table 1) shows that the modified score of Lindberg, chosen to select the patients to operate, has a potential perforation rate of 28.7%, which is below these standards. Even if the initial missed perforation rate is unknown because non-operated patients with perforation are not cited, we can say that the potential issue of observing a lot of patients that should be operated can be minimized.

All the scores studied use about the same clinical signs or symptoms and laboratory values, changing the selected ones and their weighs. The resulting scores have different selectivities and specificities; some perform quite well, and some not so well. But maybe some other elements from the clinical evaluation could be considered, like the family history. A prospective study [27] indicated that heredity (first-degree relatives) is a significant factor in pediatric patients who have appendicitis, showing that children operated on for appendicitis were twice as likely to have a positive family history than those with right lower quadrant pain, and practically 3 times as likely to have a positive family history than those without abdominal pain.

Conclusion

The stability of the missed appendicitis and perforated ones over the last years [4], even with the increasing availability of CT, US and laparoscopy, shows that some new help is needed to improve the diagnosis and treatment of acute appendicitis. To help clinicians in their difficult task, a lot of different approaches have been considered and evaluated, but none had given results good enough to be considered a real standard and give strong evidence to clinicians. The attempt to combine clinical scores is new, and the results could be promising, but it has not been evaluated yet. Standardized performance criteria should be applied to ensure objective judgment, and whenever possible, a prospective study should be done - for example, by comparing a baseline phase without scoring and a subsequent test phase with scoring [11].

Acknowledgments

The author has no relationships with Exsys®, Inc.

References

1 Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. Am J Epidemiol 1990;132(5):910-25.

2 Korner H, Sondenaa K, Soreide JA, Andersen E, Nysted A, Lende TH, et al. Incidence of acute nonperforated and perforated appendicitis: age-specific and sex-specific analysis. World J Surg 1997;21(3):313-7.

3 Lee SL, Walsh AJ, Ho HS. Computed tomography and ultrasonography do not improve and may delay the diagnosis and treatment of acute appendicitis. Arch Surg 2001;136(5):556-62.

4 Flum DR, Morris A, Koepsell T, Dellinger EP. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. Jama 2001;286(14):1748-53.

5 Malik AA, Wani NA. Continuing diagnostic challenge of acute appendicitis: evaluation through modified Alvarado score. Aust N Z J Surg 1998;68(7):504-5.

6 Davidson PM, Douglas CD, Hosking CS. Graded compression ultrasonography in the assessment of the “tough decision” acute abdomen in childhood. Pediatr Surg Int 1999;15(1):32-5.

7 Douglas CD, Macpherson NE, Davidson PM, Gani JS. Randomised controlled trial of ultrasonography in diagnosis of acute appendicitis, incorporating the Alvarado score. Bmj 2000;321(7266):919-22.

8 Galindo Gallego M, Fadrique B, Nieto MA, Calleja S, Fernandez-Acenero MJ, Ais G, et al. Evaluation of ultrasonography and clinical diagnostic scoring in suspected appendicitis. Br J Surg 1998;85(1):37-40.

9 Hallan S, Asberg A, Edna TH. Estimating the probability of acute appendicitis using clinical criteria of a structured record sheet: the physician against the computer. Eur J Surg 1997;163(6):427-32.

10 Van Way CW, 3rd, Murphy JR, Dunn EL, Elerding SC. A feasibility study of computer aided diagnosis in appendicitis. Surg Gynecol Obstet 1982;155(5):685-8.

11 Ohmann C, Yang Q, Franke C. Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. Eur J Surg 1995;161(4):273-81.

12 Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. Ann Surg 1983;198(6):753-9.

13 Arnbjørnsson E. Scoring system for computer-aided diagnosis of acute appendicitis. The value of prospective versus retrospective studies. Ann Chir Gynaecol 1985;74(4):159-66.

14 Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 1986;15(5):557-64.

15 Macklin CP, Radcliffe GS, Merei JM, Stringer MD. A prospective evaluation of the modified Alvarado score for acute appendicitis in children. Ann R Coll Surg Engl 1997;79(3):203-5.

16 Fenyö G. Routine use of a scoring system for decision-making in suspected acute appendicitis in adults. Acta Chir Scand 1987;153(9):545-51.

17 Hoffmann J, Rasmussen OO. Aids in the diagnosis of acute appendicitis. Br J Surg 1989;76(8):774-9.

18 Lindberg G, Feny G. Algorithmic diagnosis of appendicitis using Bayes' theorem and logistic regression. In: Press OU, editor. Bayesian Statistics 3; 1988. p. 665-8.

19 Dado G, Anania G, Baccarani U, Marcotti E, Donini A, Risaliti A, et al. Application of a clinical score for the diagnosis of acute appendicitis in childhood: a retrospective analysis of 197 patients. J Pediatr Surg 2000;35(9):1320-2.

20 Izbicki JR, Wilker DK, Mandelkow HK, Muller K, Siebeck M, Geissler K, et al. [Retro- and prospective studies on the value of clinical and laboratory chemical data in acute appendicitis]. Chirurg 1990;61(12):887-93; discussion 893-4.

21 De Dombal FT. The diagnosis of acute abdominal pain with computer assistance: worldwide perspective. Ann Chir 1991;45(4):273-7.

22 Christian F, Christian GP. A simple scoring system to reduce the negative appendicectomy rate. Ann R Coll Surg Engl 1992;74(4):281-5.

23 Ohmann C, Franke C, Yang Q, Margulies M, Chan M, van Elk PJ, et al. [Diagnostic score for acute appendicitis]. Chirurg 1995;66(2):135-41.

24 Zielke A, Sitter H, Rampp TA, Schafer E, Hasse C, Lorenz W, et al. [Validation of a diagnostic scoring system (Ohmann score) in acute appendicitis]. Chirurg 1999;70(7):777-83; discussion 784.

25 Zielke A, Sitter H, Rampp T, Bohrer T, Rothmund M. Clinical decision-making, ultrasonography, and scores for evaluation of suspected acute appendicitis. World J Surg 2001;25(5):578-84.

26 Exsys® Inc. CORVID™, http://www.exsys.com

27 Gauderer MW, Crane MM, Green JA, DeCou JM, Abrams RS. Acute appendicitis in children: the importance of family history. J Pediatr Surg 2001;36(8):1214-7.