Author's response to reviews

Title: A hybrid decision support model to discover informative knowledge in diagnosing acute appendicitis

Authors:

Chang-Sik Son (csson@kmu.ac.kr)
Byoung-Kuk Jang (jangha106@dsmc.or.kr)
Suk-Tae Seo (kenneth78@kmu.ac.kr)
Min-Soo Kim (msk0328@kmu.ac.kr)
Yoon-Nyun Kim (ynkim@dsmc.or.kr)

Version: 2 Date: 12 December 2011

Author's response to reviews: see over
Dear Editor

We thank you for considering our manuscript, titled “A hybrid decision support model to discover informative knowledge in diagnosing acute appendicitis (MS: 8768727696075406)” to the review. Based on the reviewer’s comments and suggestions, we have rewritten some sections of the manuscript. We also copyedited the manuscript to improve the quality of written English. In the following pages, we included the point-by-point response (highlighted in red) to comments of the reviewers. We believe that the reviews have helped us to enrich the main results and contributions in the revised version of the manuscript.

Thank you very kindly for your consideration.

Sincerely,

Yoon-Nyun Kim, MD, PhD (Corresponding author)

Dept. of Internal Medicine, School of Medicine, Keimyung University, Rep. of Korea

E-mail: ynkim@dsmc.or.kr
Reply to editor and reviewers

Editor comments:

Experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the Helsinki Declaration, and any experimental research on animals must follow internationally recognized guidelines. A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate.

Response: We have added in the ‘Methods’ section as follows:

After obtaining the Institutional Review Board (IRB) approval (no. 11–275) from Keimyung University Dongsan Hospital, we retrospectively collected the medical records of all patients attending the emergency medical center complaining mainly of acute abdominal pain between July 2006 and June 2007.
Reviewer: Stamatis Efthathiu

<Major Compulsory Revisions>

1. The authors should apply the decision support model yielded from this study to an independent prospective database of patients with suspected appendicitis in order to validate its accuracy. Unless this is done on a blinded basis the message conveyed is fundamentally weak.

Response: The objective of this study was to design a simple and reliable hybrid decision support model based on statistical analysis and a decision tree algorithm to provide high diagnostic accuracy in patients with suspected acute appendicitis and discover significant diagnostic knowledge among rules derived from the retrospectively collected data. To describe an unbiased estimate of the generalization error between the decision support models, 10-fold cross validation experiments were performed. Although this study should be confirmed by an independent prospective database of patients with suspected appendicitis to validate its feasibility, it is also necessary to be utilized by further studies.

2. The selection of cut-off points for dichotomizing continuous variables in multivariate analysis needs to be described and justified more clearly. This is crucial for the validity of the major study results.

Response: This is a valid point; however, the appropriate selection of cut-off points for dichotomizing continuous variables is very difficult, the criteria for which is dependent on the characteristics of collected data retrospectively. For this issue, we have utilized C5.0 decision tree algorithm based on entropy measure. The validity of the criteria is provided in the revised manuscript (Figure 2 and Table 4 & Figure 3 and Table 5).

3. From a clinical point of view, it remains unclear if the coexistence of all five variables included in the three rules (NEU > 73% and U-glucose negative and chief complaint RLQ pain and EOS > 0.3% and lipase # 32) provides superior diagnostic accuracy. This could be more relevant for everyday, bedside practice than the three rules separately.

Response: Please refer to the ‘Results’ section in the revised manuscript.

4. No mention is made regarding missing values and how they were handled in the analysis (for example listwise or pairwise deletion).
Response: We have included in the ‘Methods’ section as follows:

Only complete medical records with no missing clinical parameters were included, i.e., age, gender, chief complaints, and clinical laboratory findings, such as urinalysis, common blood cell and differential counts, serum electrolytes, routine admission, etc.

5. Has the normality of data distribution been examined (i.e., by the Kolmogorov-Smirnov test)? In many previous similar studies, the values of such predictors have been log-transformed, inasmuch as they showed non-normal distribution.

Response: We have included in the ‘Statistical analysis and decision tree model’ subsection as follows:

Univariate correlations between clinical or laboratory features were evaluated using the Chi-square test or Fisher’s exact test, which are appropriate for categorical data, and using the Student t-test or Mann-Whitney U-test with continuous variables, after checking for normality using the Kolmogorov-Smirnov test. We have confirmed that several variables showed non-normal distribution ($p < 0.05$). The corrected $p$ values are highlighted in the revised manuscript (Tables 1 and 2). Also, we have re-performed the previous experiments to consider newly added variables such as sodium ($p < 0.001$), glucose ($p < 0.01$), and activated partial thromboplastin time ($p < 0.01$). The results are corrected in the Results section of the revised manuscript.

6. Has the multivariate model been assessed for multicollinearity? (the latter is a common problem in simultaneous analyses of various risk factors). This checking could be done, for example, by calculating the variance inflation factor or the condition number.

Response: We have included in the ‘Results’ section as follows:

These variables were tested by linear regression analysis to evaluate multicollinearity among the predictors. The data did not violate the multicollinearity assumption. The tolerance of each independent variable was greater than 0.616. The variance inflation factor (VIF) values of the variables ranged from 1.005 to 1.624.

7. The discussion section could be shortened to be more concise and focused. The paragraph “From the decision support model…normal appendix” (pages 11-12) should be moved to the Results section.

Response: Please refer to the revised manuscript.
8. Were any CRP measurements done?

**Response:** As described in the above, we have used complete medical records with no missing clinical parameters. The consideration will be addressed in future research.

9. Are there any imaging data available for these patients (for example abdomen U/S, which is routinely performed in most patients with acute abdominal pain)?

**Response:** In most patients with acute abdominal pain, USG or CT was performed. Generally, patients presenting to the ED with acute abdominal pain have a disease that may be diagnosed by CT. Imaging studies are cost effective if a definitive diagnosis can be made and observation in a hospital can be avoided. More importantly, imaging studies of patients with an uncertain diagnosis may reduce the rate of perforation, and thus reduce morbidity, mortality, and postoperative hospital stays. However, the routine use of USG or CT in the diagnosis of acute appendicitis in all patients is not well established. If the diagnosis is apparent from the history, physical examination, and laboratory studies, taking the patient directly to surgery without imaging is justified. The results of several studies show no significant change in misdiagnosis of the acute appendicitis after widespread implementation of USG or CT. Thus, the use of imaging studies to diagnose the acute appendicitis is discussed separately. The considerations will provide directions for fruitful further research.

10. The English used as well as the application of abbreviations are generally poor, thus downgrading the verbal support of scientific arguments in this paper.

**Response:** We copyedited the manuscript to improve the written English.

**Level of interest:** An article whose findings are important to those with closely related research interests.

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:** I declare that I have no competing interests
Reviewer: Kohei Akazawa

<Minor Essential Revisions>

1. The authors used Chi-square and Student's t-test to extract the significant clinical factors. However, some strict assumptions should be needed to apply these statistical methods to real clinical datasets. Actually, some continuous variables have considerable variances between AA and Non-AA groups, and Student's t-test could not be used for the variables.

Response: We have included in the ‘Statistical analysis and decision tree model’ subsection as follows:

Univariate correlations between clinical or laboratory features were evaluated using the Chi-square test or Fisher’s exact test, which are appropriate for categorical data, and using the Student t-test or Mann-Whitney U-test with continuous variables, after checking for normality using the Kolmogorov-Smirnov test. We have confirmed that several variables showed non-normal distribution ($p < 0.05$). The corrected $p$ values are highlighted in the revised manuscript (Tables 1 and 2). Also, we have re-performed the previous experiments to consider newly added variables such as sodium ($p < 0.001$), glucose ($p < 0.01$), and activated partial thromboplastin time ($p < 0.01$). The results are corrected in the Results section of the revised manuscript.

2. I think that the enter and removal criteria of 0.01 and 0.05 (these are not p-values. “p=” should be deleted) in multiple logistic regression analysis are too strict. This setting leads to extracting only few factors, with comparison to chi-square and Student's t-test. I am interested in the results of accuracy, sensitivity, specificity, and ROC analysis, when the authors perform the same analysis for the enter and remove criteria of 0.05 and 0.06.

Response: We have included in the ‘Results’ section as follows:

In the revised manuscript, independent risk factors from the full dataset were identified using Wald forward logistic regression, when defining an entry and removal criteria of 0.01 and 0.05, or 0.05 and 0.10, respectively. Regardless of the criteria used, the independent risk factors provided the same results using the two logistic models. Six variables in the final logistic regression that were independently associated with acute appendicitis: complaints, urine glucose, white blood cell, neutrophils, total bilirubin, and lipase (Table 3). The ACC, SENS, SPEC, PPV, and NPV, were 79.8%, 76.3%, 82.8%, 79.5%, and 80.0%, respectively. The AUC of the models was 79.5% (95% CI, 74.7–83.8), indicating fair discriminatory power. After the application of C5.0 decision tree algorithm, five
of six variables were selected and their importance was defined in the following order: neutrophils, complaints, total bilirubin, urine glucose, and lipase. The ACC, SENS, SPEC, PPV, NPV, and AUC measures were 82.5%, 74.3%, 89.7%, 86.3%, 80.0%, and 82.0% (95% CI, 77.4–86.0). In 10-fold cross validation experiments, the difference of clinical factors selected before / after the application of C5.0 decision tree algorithm is given in Tables 6 and 7. The experimental results showed that the decision support model based on multivariate analysis using loose criteria was better than that using strict criteria, especially the AUC measure, although the discriminatory power between the two models was not statistically significant ($p = 0.400; 95\% \ CI, -2.0–5.02$).

3. The expressions of "p=0.000" are incorrect. The authors should describe "p<0.001".

Response: Please refer to the revised manuscript.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.
Reviewer: Christian Ohmann

<Major Compulsory Revisions>

1. Important for diagnostic accuracy is the definition of the population under study. The authors have included patients admitted to an emergency medical center with chief complaint of acute abdominal pain. This includes a variety of diagnoses including acute appendicitis. The other diagnoses are summarized under the heading "normal appendix", which is a very inhomogeneous group of patients. The discriminatory ability of decision support depends strongly (more than the methods used) on the definition of the comparative group. There is no information, how this group is composed and whether there was any selection of patients.

Response: We have changed in the ‘Methods’ section as follows:

Patients diagnosed with complaints other than appendicitis were excluded, such as acute cholecystitis or diverticulitis, appendectomy incidental to another surgical procedure, previous use of antibiotics for chronic appendicitis, and appendectomy for chronic abdominal pain. The eligibility for study group (n = 152) was defined according to the International Classification of Diseases-10 (ICD-10) codes: K35.0 (acute appendicitis with generalized peritonitis), K35.1 (acute appendicitis with peritoneal abscess), and K35.9 (acute appendicitis without generalized peritonitis). Discharged patients (n = 174) admitted to the emergency medical center who complained mainly of acute abdominal pain were defined as the control group. All data collected were reconfirmed by gastroenterologists.

2. No definition is given for acute appendicitis, normal appendix and chief complaints. How was "right lower quadrant pain" defined? What is a normal appendix (normal clinical findings) when the patient is not operated? Is histology necessary for the diagnosis of acute appendicitis? Studies in the literature have attempted to standardize definitions.

Response: Please refer to the above response for acute appendicitis (study group) and normal appendix (control group). The right lower quadrant of the human abdomen, often abbreviated as RLQ, is used to refer to a portion of the abdomen that allows doctors to localize pain and tenderness, scars, lumps and other items of interest. The RLQ extends from the median plan (i.e., a midsagittal plan which bisects the body vertically through the navel, dividing the body exactly in left and right side) to the right of the patient, and from the umbilical plane to the right inguinal ligament. In this study, we spilt the abdomen areas into eight distinct regions (RUQ, RLQ, LUQ, LLQ, abdominal with 4 abdominal quadrants, periumbilical area, lower abdominal with RLQ and LLQ, and upper abdominal with RUQ and LUQ) based on 4 abdominal quadrants [14] such as right upper quadrant (RUQ), right
lower quadrant (RLQ), left upper quadrant (LUQ), left lower quadrant (LLQ).

3. This is a retrospective study, which should have missing data. This is important because missing data are worsening the diagnostic prediction. Unfortunately, there is no information about missing data in the study.

Response: We have included in the ‘Methods’ section as follows:

Only complete medical records with no missing clinical parameters were included, i.e., age, gender, chief complaints, and clinical laboratory findings, such as urinalysis, common blood cell and differential counts, serum electrolytes, routine admission, etc.

<Minor Essential Revisions>

4. In the study cross-validation was performed adequately to estimate diagnostic accuracy. Figure 2 shows the final results for the full data set. It would be nice to know whether training on the different 90% samples resulted in different parameters selected in the multivariate model and with respect to the order of selection of parameters in the application of C5.0 decision tree algorithm.

Response: We have included in the ‘Results’ section as follows:

The six measures were compared using a 10-fold cross validation to assess the generalization ability of these decision support models. The differences in the clinical factors selected before and after the application of the C5.0 decision tree algorithm are shown in Tables 6 and 7.

5. The conclusions drawn are far too far-reaching. As long as the decision support-model is not independently evaluated, no recommendation for clinical use should be given. It is known from de Dombal and other studies that external evaluation may reduce diagnostic accuracy considerably.

Response: Please refer to the ‘Results’ section in the revised manuscript.

6. The C5.0 decision tree algorithm is quite simple. These and other similar approaches have not produced good results in the literature. The authors should explain why this is different in their study. Is it due to selection of patients, aggregation of other diagnoses in one group, inclusion of new parameters, better data quality, etc.

Response: Please refer to the ‘Results’ section in the revised manuscript.
Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests: ‘I declare that I have no competing interests’

Thank you again for your interest in our manuscript.

Sincerely,

Yoon-Nyun Kim, MD, PhD (Corresponding author)

Dept. of Internal Medicine, School of Medicine, Keimyung University, Rep. of Korea

E-mail: ynkim@dsmc.or.kr